

Arak M. Mathai · Serge B. Provost · Hans J. Haubold

Multivariate Statistical Analysis in the Real and Complex Domains

OPEN ACCESS

 Springer

Multivariate Statistical Analysis in the Real and Complex Domains

Arak M. Mathai • Serge B. Provost • Hans J. Haubold

Multivariate Statistical Analysis in the Real and Complex Domains

 Springer

Arak M. Mathai
Mathematics and Statistics
McGill University
Montreal, Canada

Serge B. Provost
Statistical and Actuarial Sciences
The University of Western Ontario
London, ON, Canada

Hans J. Haubold
Vienna International Centre
UN Office for Outer Space Affairs
Vienna, Austria



ISBN 978-3-030-95863-3 ISBN 978-3-030-95864-0 (eBook)
<https://doi.org/10.1007/978-3-030-95864-0>

© The Editor(s) (if applicable) and The Author(s) 2022, corrected publication 2022. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Multivariate statistical analysis often proves to be a challenging subject for students. The difficulty arises in part from the reliance on several types of symbols such as subscripts, superscripts, bars, tildes, bold-face characters, lower- and uppercase Roman and Greek letters, and so on. However, resorting to such notations is necessary in order to refer to the various quantities involved such as scalars and matrices either in the real or complex domain. When the first author began to teach courses in advanced mathematical statistics and multivariate analysis at McGill University, Canada, and other academic institutions around the world, he was seeking means of making the study of multivariate analysis more accessible and enjoyable. He determined that the subject could be made simpler by treating mathematical and random variables alike, thus avoiding the distinct notation that is generally utilized to represent random and non-random quantities. Accordingly, all scalar variables, whether mathematical or random, are denoted by lowercase letters and all vector/matrix variables are denoted by capital letters, with vectors and matrices being identically denoted since vectors can be viewed as matrices having a single row or column. As well, variables belonging to the complex domain are readily identified as such by placing a tilde over the corresponding lowercase and capital letters. Moreover, he noticed that numerous formulas expressed in terms of summations, subscripts, and superscripts could be more efficiently represented by appealing to matrix methods. He further observed that the study of multivariate analysis could be simplified by initially delivering a few lectures on Jacobians of matrix transformations and elementary special functions of matrix argument, and by subsequently deriving the statistical density functions as special cases of these elementary functions as is done for instance in the present book for the real and complex matrix-variate gamma and beta density functions. Basic notes in these directions were prepared and utilized by the first author for his lectures over the past decades. The second and third authors then joined him and added their contributions to flesh out this material to full-fledged book form. Many of the notable features that distinguish this monograph from other books on the subject are listed next.

Special Features

- 1.** As the title of the book suggests, its most distinctive feature is its development of a parallel theory of multivariate analysis in the complex domain side by side with the corresponding treatment of the real cases. Various quantities involving complex random variables such as Hermitian forms are widely used in many areas of applications such as light scattering, quantum physics, and communication theory, to name a few. A wide readership is expected as, to our knowledge, this is the first book in the area that systematically combines in a single source the real results and their complex counterparts. Students will be able to better grasp the results that are holding in the complex field by relating them to those existing in the real field.
- 2.** In order to avoid resorting to an excessive number of symbols to denote scalar, vector, and matrix variables in the real and complex domains, the following consistent notations are employed throughout the book: All real scalar variables, whether mathematical or random, are denoted by lowercase letters and all real vector/matrix variables are denoted by capital letters, a tilde being placed on the corresponding variables in the complex domain.
- 3.** Mathematical variables and random variables are treated the same way and denoted by the same type of letters in order to avoid the double notation often utilized to represent random and mathematical variables as well as the potentially resulting confusion. If probabilities are to be attached to every value that a variable takes, then mathematical variables can be construed as degenerate random variables. This simplified notation will enable students from mathematics, physics, and other disciplines to easily understand the subject matter without being perplexed. Although statistics students may initially find this notation somewhat unsettling, the adjustment ought to prove rapid.
- 4.** Matrix methods are utilized throughout the book so as to limit the number of summations, subscripts, superscripts, and so on. This makes the representations of the various results simpler and elegant.
- 5.** A connection is established between statistical distribution theory of scalar, vector, and matrix variables in the real and complex domains and fractional calculus. This should foster further growth in both of these fields, which may borrow results and techniques from each other.
- 6.** Connections of concepts encountered in multivariate analysis to concepts occurring in geometrical probabilities are pointed out so that each area can be enriched by further work in the other one. Geometrical probability problems of random lengths, random areas, and random volumes in the complex domain may not have been developed yet. They may now be tackled by making use of the results presented in this book.

7. Classroom lecture style is employed as this book's writing style so that the reader has the impression of listening to a lecture upon reading the material.
8. The central concepts and major results are followed by illustrative worked examples so that students may easily comprehend the meaning and significance of the stated results. Additional problems are provided as exercises for the students to work out so that the remaining questions they still may have can be clarified.
9. Throughout the book, the majority of the derivations of known or original results are innovative and rather straightforward as they rest on simple applications of results from matrix algebra, vector/matrix derivatives, and elementary special functions.
10. Useful results on vector/matrix differential operators are included in the mathematical preliminaries for the real case, and the corresponding operators in the complex domain are developed in Chap. 3. They are utilized to derive maximum likelihood estimators of vector/matrix parameters in the real and complex domains in a more straightforward manner than is otherwise the case with the usual lengthy procedures. The vector/matrix differential operators in the complex domain may actually be new whereas their counterparts, the real, case may be found in Mathai (1997) [see Chapter 1, reference list].
11. The simplified and consistent notation of dX is used to denote the wedge product of the differentials of all functionally independent real scalar variables in X , whether X is a scalar, a vector, or a square or rectangular matrix, with $d\tilde{X}$ being utilized for the corresponding wedge product of differentials in the complex domain.
12. Equation numbering is done sequentially chapter/section-wise; for example, (3.5.4) indicates the fourth equation appearing in Sect. 5 of Chap. 3. To make the numbering scheme more concise and descriptive, the section titles, lemmas, theorems, exercises, and equations pertaining to the complex domain will be identified by appending the letter 'a' to the respective section numbers such as (3.5a.4). The notation (i), (ii), . . . , is employed for neighboring equation numbers related to a given derivation.
13. References to the previous materials or equation numbers as well as references to subsequent results appearing in the book are kept a minimum. In order to enhance readability, the main notations utilized in each chapter are repeated at the beginning of each one of them. As well, the reader may notice certain redundancies in the statements. These are intentional and meant to make the material easier to follow.
14. Due to the presence of numerous parameters, students generally find the subject of factor analysis quite difficult to grasp and apply effectively. Their understanding of the topic should be significantly enhanced by the explicit derivations that are provided, which incidentally are believed to be original.

15. Only the basic material in each topic is covered. The subject matter is clearly discussed and several worked examples are provided so that the students can acquire a clear understanding of this primary material. Only the materials used in each chapter are given as reference—mostly the authors' own works. Additional reading materials are listed at the very end of the book. After acquainting themselves with the introductory material presented in each chapter, the readers ought to be capable of mastering more advanced related topics on their own.

Multivariate analysis encompasses a vast array of topics. Even if the very primary materials pertaining to most of these topics were included in a basic book such as the present one, the length of the resulting monograph would be excessive. Hence, certain topics had to be chosen in order to produce a manuscript of a manageable size. The selection of the topics to be included or excluded is authors' own choice, and it is by no means claimed that those included in the book are the most important ones or that those being omitted are not relevant. Certain pertinent topics, such as confidence regions, multiple confidence intervals, multivariate scaling, tests based on arbitrary statistics, and logistic and ridge regressions, are omitted so as to limit the size of the book. For instance, only some likelihood ratio statistics or λ -criteria based tests on normal populations are treated in Chap. 6 on tests of hypotheses, whereas the authors could have discussed various tests of hypotheses on parameters associated with the exponential, multinomial, or other populations, as they also have worked on such problems. As well, since results related to elliptically contoured distributions including the spherically symmetric case might be of somewhat limited interest, this topic is not pursued further subsequently to its introduction in Chap. 3. Nevertheless, standard applications such as principal component analysis, canonical correlation analysis, factor analysis, classification problems, multivariate analysis of variance, profile analysis, growth curves, cluster analysis, and correspondence analysis are properly covered.

Tables of percentage points are provided for the normal, chisquare, Student- t , and F distributions as well as for the null distributions of the statistics for testing the independence and for testing the equality of the diagonal elements given that the population covariance matrix is diagonal, as they are frequently required in applied areas. Numerical tables for other relevant tests encountered in multivariate analysis are readily available in the literature.

This work may be used as a reference book or as a textbook for a full course on multivariate analysis. Potential readership includes mathematicians, statisticians, physicists, engineers, as well as researchers and graduate students in related fields. Chapters 1–8 or sections thereof could be covered in a one- to two-semester course on mathematical statistics or multivariate analysis, while a full course on applied multivariate analysis might

focus on Chaps. 9–15. Readers with little interest in complex analysis may omit the sections whose numbers are followed by an ‘a’ without any loss of continuity. With this book and its numerous new derivations, those who are already familiar with multivariate analysis in the real domain will have an opportunity to further their knowledge of the subject and to delve into the complex counterparts of the results.

The authors wish to thank the following former students of the Centre for Mathematical and Statistical Sciences, India, for making use of a preliminary draft of portions of the book for their courses and communicating their comments: Dr. T. Princy, Cochin University of Science and Technology, Kochi, Kerala, India; Dr. Nicy Sebastian, St. Thomas College, Calicut University, Thrissur, Kerala, India; and Dr. Dilip Kumar, Kerala University, Trivandrum, India. The authors also wish to express their thanks to Dr. C. Satheesh Kumar, Professor of Statistics, University of Kerala, and Dr. Joby K. Jose, Professor of Statistics, Kannur University, for their pertinent comments on the second drafts of the chapters. The authors have no conflict of interest to declare. The second author would like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

Montreal, ON, Canada
London, ON, Canada
Vienna, Austria
July 1, 2022

Arak M. Mathai
Serge B. Provost
Hans J. Haubold

Contents

Preface/Special features	v
Contents	xi
List of Symbols	xxv
1 Mathematical Preliminaries	1
1.1 Introduction	1
1.2 Determinants	12
1.2.1 Inverses by row operations or elementary operations	24
1.3 Determinants of Partitioned Matrices	28
1.4 Eigenvalues and Eigenvectors	31
1.5 Definiteness of Matrices, Quadratic and Hermitian Forms	35
1.5.1 Singular value decomposition	38
1.6 Wedge Product of Differentials and Jacobians	39
1.7 Differential Operators	50
1.7.1 Some basic applications of the vector differential operator	52
References	56
2 The Univariate Gaussian and Related Distributions	57
2.1 Introduction	57
2.1a The Complex Scalar Gaussian Variable	58
2.1.1 Linear functions of Gaussian variables in the real domain	60
2.1a.1 Linear functions in the complex domain	61
2.1.2 The chisquare distribution in the real domain	62
2.1a.2 The chisquare distribution in the complex domain	64
2.1.3 The type-2 beta and F distributions in the real domain	66
2.1a.3 The type-2 beta and F distributions in the complex domain	69

2.1.4	Power transformation of type-1 and type-2 beta random variables	71
2.1.5	Exponentiation of real scalar type-1 and type-2 beta variables	72
2.1.6	The Student- t distribution in the real domain	73
2.1a.4	The Student- t distribution in the complex domain	74
2.1.7	The Cauchy distribution in the real domain	75
2.2	Quadratic Forms, Chisquaredness and Independence in the Real Domain	78
2.2a	Hermitian Forms, Chisquaredness and Independence in the Complex Domain	82
2.2.1	Extensions of the results in the real domain	84
2.2a.1	Extensions of the results in the complex domain	86
2.3	Simple Random Samples from Real Populations and Sampling Distributions	88
2.3a	Simple Random Samples from a Complex Gaussian Population	94
2.3.1	Noncentral chisquare having n degrees of freedom in the real domain	97
2.3a.1	Noncentral chisquare having n degrees of freedom in the complex domain	101
2.4	Distributions of Products and Ratios and Connection to Fractional Calculus	104
2.5	General Structures	109
2.5.1	Product of real scalar gamma variables	109
2.5.2	Product of real scalar type-1 beta variables	111
2.5.3	Product of real scalar type-2 beta variables	111
2.5.4	General products and ratios	112
2.5.5	The H-function	112
2.6	A Collection of Random Variables	115
2.6.1	Chebyshev's inequality	115
2.7	Parameter Estimation: Point Estimation	120
2.7.1	The method of moments and the method of maximum likelihood	120
2.7.2	Bayes' estimates	123
2.7.3	Interval estimation	124
	References	126
3	The Multivariate Gaussian and Related Distributions	129
3.1	Introduction	129
3.1a	The Multivariate Gaussian Density in the Complex Domain	131

3.2	The Multivariate Normal or Gaussian Distribution, Real Case	135
	3.2.1 The moment generating function in the real case	141
3.2a	The Moment Generating Function in the Complex Case	146
	3.2a.1 Moments from the moment generating function	148
	3.2a.2 Linear functions	151
3.3	Marginal and Conditional Densities, Real Case	154
3.3a	Conditional and Marginal Densities in the Complex Case	160
3.4	Chisquaredness and Independence of Quadratic Forms in the Real Case	166
	3.4.1 Independence of quadratic forms	169
3.4a	Chisquaredness and Independence in the Complex Gaussian Case	172
	3.4a.1 Independence of Hermitian forms	175
3.5	Samples from a p -variate Real Gaussian Population	178
3.5a	Simple Random Sample from a p -variate Complex Gaussian Population	181
	3.5.1 Some simplifications of the sample matrix in the real Gaussian case	182
	3.5.2 Linear functions of the sample vectors	184
	3.5a.1 Some simplifications of the sample matrix in the complex Gaussian case	186
	3.5a.2 Linear functions of the sample vectors in the complex domain	186
	3.5.3 Maximum likelihood estimators of the p -variate real Gaussian distribution	187
	3.5a.3 MLE's in the complex p -variate Gaussian case	191
	3.5a.4 Matrix derivatives in the complex domain	191
	3.5.4 Properties of maximum likelihood estimators	196
	3.5.5 Some limiting properties in the p -variate case	204
3.6	Elliptically Contoured Distribution, Real Case	206
	3.6.1 Some properties of elliptically contoured distributions	206
	3.6.2 The density of $u = r^2$	208
	3.6.3 Mean value vector and covariance matrix	209

	3.6.4	Marginal and conditional distributions	210
	3.6.5	The characteristic function of an elliptically contoured distribution	212
	References		214
4	The Matrix-Variate Gaussian Distribution		217
4.1	Introduction		217
4.2	Real Matrix-variate and Multivariate Gaussian Distributions		218
4.2a	The Matrix-variate Gaussian Density, Complex Case		221
	4.2.1	Some properties of a real matrix-variate Gaussian density	224
	4.2a.1	Some properties of a complex matrix-variate Gaussian density	225
	4.2.2	Additional properties in the real and complex cases	226
	4.2.3	Some special cases	229
4.3	Moment Generating Function and Characteristic Function, Real Case		232
4.3a	Moment Generating and Characteristic Functions, Complex Case		235
	4.3.1	Distribution of the exponent, real case	238
	4.3a.1	Distribution of the exponent, complex case	239
	4.3.2	Linear functions in the real case	240
	4.3a.2	Linear functions in the complex case	246
	4.3.3	Partitioning of the parameter matrix	252
	4.3.4	Distributions of quadratic and bilinear forms	253
4.4	Marginal Densities in the Real Matrix-variate Gaussian Case		255
4.4a	Marginal Densities in the Complex Matrix-variate Gaussian Case		261
4.5	Conditional Densities in the Real Matrix-variate Gaussian Case		265
4.5a	Conditional Densities in the Matrix-variate Complex Gaussian Case		270
	4.5.1	Re-examination of the case $q = 1$	274
4.6	Sampling from a Real Matrix-variate Gaussian Density		276
	4.6.1	The distribution of the sample sum of products matrix, real case	278
	4.6.2	Linear functions of sample vectors	279
	4.6.3	The general real matrix-variate case	280
4.6a	The General Complex Matrix-variate Case		284
4.7	The Singular Matrix-variate Gaussian Distribution		286
	References		287

5	Matrix-Variate Gamma and Beta Distributions	289
5.1	Introduction	289
5.1a	The Complex Matrix-variate Gamma	291
5.2	The Real Matrix-variate Gamma Density	292
	5.2.1 The mgf of the real matrix-variate gamma distribution	295
5.2a	The Matrix-variate Gamma Function and Density, Complex Case	298
	5.2a.1 The mgf of the complex matrix-variate gamma distribution ..	300
5.3	Matrix-variate Type-1 Beta and Type-2 Beta Densities, Real Case	302
	5.3.1 Some properties of real matrix-variate type-1 and type-2 beta densities	305
5.3a	Matrix-variate Type-1 and Type-2 Beta Densities, Complex Case ...	305
	5.3.2 Explicit evaluation of type-1 matrix-variate beta integrals, real case	306
	5.3a.1 Evaluation of matrix-variate type-1 beta integrals, complex case	311
	5.3.3 General partitions, real case	314
	5.3.4 Methods avoiding integration over the Stiefel manifold	315
	5.3.5 Arbitrary moments of the determinants, real gamma and beta matrices	317
	5.3a.2 Arbitrary moments of the determinants in the complex case .	324
5.4	The Densities of Some General Structures	327
	5.4.1 The G-function	329
	5.4.2 Some special cases of the G-function	331
	5.4.3 The H-function	332
	5.4.4 Some special cases of the H-function	333
5.5, 5.5a	The Wishart Density	335
	5.5.1 Explicit evaluations of the matrix-variate gamma integral, real case	336
	5.5a.1 Evaluation of matrix-variate gamma integrals in the complex case	338
	5.5.2 Triangularization of the Wishart matrix in the real case	340
	5.5a.2 Triangularization of the Wishart matrix in the complex domain	342

5.5.3	Samples from a p -variate Gaussian population and the Wishart density	343
5.5a.3	Sample from a complex Gaussian population and the Wishart density	344
5.5.4	Some properties of the Wishart distribution, real case	345
5.5.5	The generalized variance	347
5.5.6	Inverse Wishart distribution	349
5.5.7	Marginal distributions of a Wishart matrix	350
5.5.8	Connections to geometrical probability problems	355
5.6	The Distribution of the Sample Correlation Coefficient	355
5.6.1	The special case $\rho = 0$	357
5.6.2	The multiple and partial correlation coefficients	358
5.6.3	Different derivations of $\rho_{1.(2...p)}$	358
5.6.4	Distributional aspects of the sample multiple correlation coefficient	361
5.6.5	The partial correlation coefficient	364
5.7	Distributions of Products and Ratios of Matrix-variate Random Variables	366
5.7.1	The density of a product of real matrices	367
5.7.2	M-convolution and fractional integral of the second kind	368
5.7.3	A pathway extension of fractional integrals	370
5.7.4	The density of a ratio of real matrices	370
5.7.5	A pathway extension of first kind integrals, real matrix-variate case	373
5.7a	Density of a Product and Integrals of the Second Kind	374
5.7a.1	Density of a product and fractional integral of the second kind, complex case	374
5.7a.2	Density of a ratio and fractional integrals of the first kind, complex case	376
5.8	Densities Involving Several Matrix-variate Random Variables, Real Case	378
5.8.1	The type-1 Dirichlet density, real scalar case	378
5.8.2	The type-2 Dirichlet density, real scalar case	380
5.8.3	Some properties of Dirichlet densities in the real scalar case ..	381
5.8.4	Some generalizations of the Dirichlet models	382
5.8.5	A pseudo Dirichlet model	383
5.8.6	The type-1 Dirichlet model in real matrix-variate case	384

	5.8.7	The type-2 Dirichlet model in the real matrix-variate case . . .	386
	5.8.8	A pseudo Dirichlet model	388
5.8a		Dirichlet Models in the Complex Domain	389
	5.8a.1	A type-2 Dirichlet model in the complex domain	390
		References	391
6		Hypothesis Testing and Null Distributions	395
6.1		Introduction	395
6.2		Testing $H_o : \mu = \mu_0$ (Given) When Σ is Known, the Real $N_p(\mu, \Sigma)$ Case	397
	6.2.1	Paired variables and linear functions	399
	6.2.2	Independent Gaussian populations	400
6.2a		Testing $H_o : \mu = \mu_o$ (given) When Σ is Known, Complex Gaussian Case	403
	6.2.3	Test involving a subvector of a mean value vector when Σ is known	405
	6.2.4	Testing $\mu_1 = \cdots = \mu_p$, with Σ known, real Gaussian case . .	408
	6.2.5	Likelihood ratio criterion for testing $H_o : \mu_1 = \cdots = \mu_p$, Σ known	411
6.3		Testing $H_o : \mu = \mu_o$ (given) When Σ is Unknown, Real Gaussian Case	414
	6.3.1	The distribution of the test statistic	416
	6.3.2	Paired values or linear functions when Σ is unknown	419
	6.3.3	Independent Gaussian populations	421
	6.3.4	Testing $\mu_1 = \cdots = \mu_p$ when Σ is unknown in the real Gaussian case	426
	6.3.5	Likelihood ratio test for testing $H_o : \mu_1 = \cdots = \mu_p$ when Σ is unknown	429
6.4		Testing Hypotheses on the Population Covariance Matrix	431
	6.4.1	Arbitrary moments of λ	433
	6.4.2	The asymptotic distribution of $-2 \ln \lambda$ when testing $H_o : \Sigma = \Sigma_o$	434
	6.4.3	Tests on Wilks' concept of generalized variance	436
6.5		The Sphericity Test or Testing if $H_o : \Sigma = \sigma^2 I$, Given a $N_p(\mu, \Sigma)$ Sample	436
6.6		Testing the Hypothesis that the Covariance Matrix is Diagonal	445
6.7		Equality of Diagonal Elements, Given that Σ is Diagonal, Real Case	449
6.8		Hypothesis that the Covariance Matrix is Block Diagonal, Real Case	453

	6.8.1	Special case: $k = 2$	457
	6.8.2	General case: $k = 2$	457
6.9		Hypothesis that the Mean Value and Covariance Matrix are Given	460
6.10		Testing Hypotheses on Linear Regression Models or Linear Hypotheses	463
	6.10.1	A simple linear model	467
	6.10.2	Hypotheses on individual parameters	470
6.11		Problem Involving Two or More Independent Gaussian Populations	473
	6.11.1	Equal but unknown covariance matrices	474
6.12		Equality of Covariance Matrices in Independent Gaussian Populations	475
	6.12.1	Asymptotic behavior	478
6.13		Testing the Hypothesis that k Independent p -variate Real Gaussian Populations are Identical and Multivariate Analysis of Variance	479
	6.13.1	Conditional and marginal hypotheses	481
	6.13.2	Arbitrary moments of λ_1	482
	6.13.3	The asymptotic distribution of $-2 \ln \lambda_1$	484
	References	489
7		Rectangular Matrix-Variate Distributions	493
	7.1	Introduction	493
	7.2	Rectangular Matrix-Variate Gamma Density, Real Case	494
	7.2.1	Extension of the gamma density to the real rectangular matrix-variate case	495
	7.2.2	Multivariate gamma and Maxwell-Boltzmann densities, real case	499
	7.2.3	Some properties of the rectangular matrix-variate gamma density	502
	7.2.4	Connection to the volume of a random parallelotope	503
	7.2.5	Pathway to real matrix-variate gamma and Maxwell-Boltzmann densities	503
	7.2.6	Multivariate gamma and Maxwell-Boltzmann densities, pathway model	507
	7.2.7	Concluding remarks	509
7.2a		Complex Matrix-Variate Gamma and Maxwell-Boltzmann Densities	511

7.2a.1	Extension of the Matrix-Variate Gamma Density in the Complex Domain.....	512
7.2a.2	The multivariate gamma density in the complex matrix-variate case	513
7.2a.3	Arbitrary moments, complex case	514
7.2a.4	A pathway extension in the complex case.....	514
7.2a.5	The Maxwell-Boltzmann and Raleigh cases in the complex domain	515
7.3	Real Rectangular Matrix-Variate Type-1 and Type-2 Beta Densities .	517
7.3.1	Arbitrary moments	519
7.3.2	Special case: $p = 1$	520
7.3a	Rectangular Matrix-Variate Type-1 Beta Density, Complex Case	521
7.3a.1	Different versions of the type-1 beta density, the complex case	521
7.3a.2	Multivariate type-1 beta density, the complex case	522
7.3a.3	Arbitrary moments in the complex case	523
7.4	The Real Rectangular Matrix-Variate Type-2 Beta Density	525
7.4.1	The real type-2 beta density in the multivariate case	526
7.4.2	Moments in the real rectangular matrix-variate type-2 beta density	526
7.4.3	A pathway extension in the real case	528
7.4a	Complex Rectangular Matrix-Variate Type-2 Beta Density	529
7.4a.1	Multivariate complex type-2 beta density	530
7.4a.2	Arbitrary moments in the complex type-2 beta density	532
7.4a.3	A pathway version of the complex rectangular matrix-variate type-1 beta density	532
7.5,7.5a	Ratios Involving Rectangular Matrix-Variate Random Variables	534
7.5.1	Multivariate F, Student- t and Cauchy densities	537
7.5a.1	A complex multivariate Student- t having ν degrees of freedom.....	538
7.5a.2	A complex multivariate Cauchy density	538
7.6	Rectangular Matrix-Variate Dirichlet Density, Real Case	538
7.6.1	Certain properties, real rectangular matrix-variate type-1 Dirichlet density	540
7.6.2	A multivariate version of the real matrix-variate type-1 Dirichlet density	542
7.7	Generalizations of the Real Rectangular Dirichlet Models.....	544
References	547

8	The Distributions of Eigenvalues and Eigenvectors	549
8.1	Introduction	549
	8.1.1 Matrix-variate gamma and beta densities, real case	549
8.1a	Matrix-variate Gamma and Beta Densities, Complex Case	552
	8.1.2 Real Wishart matrices	552
8.2	Some Eigenvalues and Eigenvectors, Real Case	553
8.2a	The Distributions of Eigenvalues in the Complex Case	558
	8.2.1 Eigenvalues of matrix-variate gamma and Wishart matrices, real case	560
	8.2a.1 Eigenvalues of complex matrix-variate gamma and Wishart matrices	563
	8.2.2 An alternative procedure in the real case	566
	8.2.3 The joint density of the eigenvectors in the real case	568
	8.2a.2 An alternative procedure in the complex case	572
8.3	The Singular Real Case	576
	8.3.1 Singular Wishart and matrix-variate gamma distributions, real case	580
	8.3.2 A direct evaluation as an eigenvalue problem	584
8.3a	The Singular Complex Case	587
	8.3a.1 Singular gamma or singular Gaussian distribution, complex case	588
	8.3a.2 A direct method of evaluation in the complex case	591
8.4	The Case of One Wishart or Gamma Matrix in the Real Domain	592
8.4a	The Case of One Wishart or Gamma Matrix, Complex Domain	594
	References	595
9	Principal Component Analysis	597
9.1	Introduction	597
9.2	Principal Components	598
9.3	Issues to Be Mindful of when Constructing Principal Components ..	604
9.4	The Vector of Principal Components	607
	9.4.1 Principal components viewed from differing perspectives	610
9.5	Sample Principal Components	611
	9.5.1 Estimation and evaluation of the principal components	615
	9.5.2 L_1 and L_2 norm principal components	618
9.6	Distributional Aspects of Eigenvalues and Eigenvectors	620

9.6.1	The distributions of the largest and smallest eigenvalues	623
9.6.2	Simplification of the factor $\prod_{i < j} (\lambda_i - \lambda_j)$	624
9.6.3	The distributions of the eigenvalues	626
9.6.4	Density of the smallest eigenvalue λ_p in the real matrix-variate gamma case	628
9.6.5	Density of the largest eigenvalue λ_1 in the real matrix-variate gamma case	630
9.6.6	Density of the largest eigenvalue λ_1 in the general real case	633
9.6.7	Density of the smallest eigenvalue λ_p in the general real case	635
	References	638
10	Canonical Correlation Analysis	641
10.1	Introduction	641
	10.1.1 An invariance property	645
10.2	Pairs of Canonical Variables	646
10.3	Estimation of the Canonical Correlations and Canonical Variables	652
	10.3.1 An iterative procedure	656
10.4	The Sampling Distribution of the Canonical Correlation Matrix	663
	10.4.1 The joint density of the eigenvalues and eigenvectors	665
	10.4.2 Testing whether the population canonical correlations equal zero	668
10.5	The General Sampling Distribution of the Canonical Correlation Matrix	669
	10.5.1 The sampling distribution of the multiple correlation coefficient	673
	References	676
11	Factor Analysis	679
11.1	Introduction	679
11.2	Linear Models from Different Disciplines	680
	11.2.1 A linear regression model	680
	11.2.2 A design of experiment model	681
11.3	A General Linear Model for Factor Analysis	682
	11.3.1 The identification problem	685

	11.3.2	Scaling or units of measurement	687
11.4		Maximum Likelihood Estimators of the Parameters	688
	11.4.1	Maximum likelihood estimators under identification conditions	690
	11.4.2	Simplifications of $ \Sigma $ and $\text{tr}(\Sigma^{-1}S)$	691
	11.4.3	Special case $\Psi = \sigma^2 I_p$	692
	11.4.4	Maximum value of the exponent	696
11.5		General Case	699
	11.5.1	The exponent in the likelihood function	702
11.6		Tests of Hypotheses	704
	11.6.1	Asymptotic distribution of the likelihood ratio statistic	706
	11.6.2	How to decide on the number r of main factors?	706
		References	709
12		Classification Problems	711
	12.1	Introduction	711
	12.2	Probabilities of Classification	713
	12.3	Two Populations with Known Distributions	714
	12.3.1	Best procedure	716
	12.4	Linear Discriminant Function	727
	12.5	Classification When the Population Parameters are Unknown	730
	12.5.1	Some asymptotic results	732
	12.5.2	Another method	733
	12.5.3	A new sample from π_1 or π_2	736
	12.6	Maximum Likelihood Method of Classification	738
	12.7	Classification Involving k Populations	744
	12.7.1	Classification when the populations are real Gaussian	745
	12.7.2	Some distributional aspects	748
	12.7.3	Classification when the population parameters are unknown	753
	12.8	The Maximum Likelihood Method when the Population Covariances Are Equal	754
	12.9	Maximum Likelihood Method and Unequal Covariance Matrices	755
13		Multivariate Analysis of Variation	759
	13.1	Introduction	759
	13.2	Multivariate Case of One-Way Classification Data Analysis	763
	13.2.1	Some properties	766

13.3	The Likelihood Ratio Criterion	767
13.3.1	Arbitrary moments of the likelihood ratio criterion	769
13.3.2	Structural representation of the likelihood ratio criterion	770
13.3.3	Some special cases	771
13.3.4	Asymptotic distribution of the λ -criterion	785
13.3.5	MANOVA and testing the equality of population mean values	786
13.3.6	When H_0 is rejected	788
13.4	MANOVA for Two-Way Classification Data	791
13.4.1	The model in a two-way classification	792
13.4.2	Estimation of parameters in a two-way classification	792
13.5	Multivariate Extension of the Two-Way Layout	796
13.5.1	Likelihood ratio test for multivariate two-way layout	798
13.5.2	Asymptotic distribution of λ in the MANOVA two-way layout	800
13.5.3	Exact densities of w in some special cases	801
	References	812
14	Profile Analysis and Growth Curves	813
14.1	Introduction	813
14.1.1	Profiles	813
14.2	Parallel Profiles	814
14.3	Coincident Profiles	818
14.3.1	Conditional hypothesis for coincident profiles	819
14.4	Level Profiles	820
14.4.1	Level profile, two independent Gaussian populations	821
14.5	Generalizations to k Populations	824
14.6	Growth Curves or Repeated Measures Designs	824
14.6.1	Growth curve model	826
14.6.2	Maximum likelihood estimation	828
14.6.3	Some special cases	833
14.6.4	Asymptotic distribution of the likelihood ratio criterion λ ...	835
14.6.5	A general structure	842
	References	844

15 Cluster Analysis and Correspondence Analysis	845
15.1 Introduction	845
15.1.1 Clusters	845
15.1.2 Distance measures	846
15.2 Different Methods of Clustering	849
15.2.1 Optimization or partitioning	849
15.3 Hierarchical Methods of Clustering	853
15.3.1 Single linkage or nearest neighbor method	854
15.3.2 Average linking as a modified distance measure	858
15.3.3 The centroid method	858
15.3.4 The median method	859
15.3.5 The residual sum of products method	859
15.3.6 Other criteria for partitioning or optimization	860
15.4 Correspondence Analysis	864
15.4.1 Two-way contingency table	865
15.4.2 Some general computations	867
15.5 Various Representations of Pearson's χ^2 Statistic	869
15.5.1 Testing the hypothesis of no association in a two-way contingency table	871
15.6 Plot of Row and Column Profiles	872
15.7 Correspondence Analysis in a Multi-way Contingency Table	878
Correction to: Multivariate Statistical Analysis in the Real and Complex Domains	C1
Tables of Percentage Points	887
Some Additional Reading Materials	907
Author Index	915
Subject Index	919

List of Symbols

$A = (a_{ij})$: a matrix, Sect. 1.1	2
$\text{diag}(d_1, \dots, d_p)$: a diagonal matrix, Sect. 1.1	4
$U \cdot V$: dot product of the vectors U and V , Sect. 1.2	6
$I = I_m$: an identity matrix, Sect. 1.1	8
A' : transpose of the matrix A , Sect. 1.1	8
$\text{tr}(A)$: trace of the square matrix A , Sect. 1.1	8
$\ X\ $: norm or length of the vector X , Sect. 1.1	9
$ A $: $\det(A)$ determinant of the square matrix A , Sect. 1.2	14
A^{-1} : regular inverse of the matrix A , Sect. 1.2	23
$\text{Cof}(A)$: cofactor matrix of A , Sect. 1.2	23
$i = \sqrt{-1}$: basic complex number, Sect. 1.4	34
A^* : complex conjugate transpose of A , Sect. 1.4	34
$A > O, \geq O, < O, \leq O$: definiteness of matrices, Sect. 1.5	36
$X'AX > 0, \geq 0, < 0, \leq 0$: definiteness of quadratic forms, Sect. 1.5	36
$X^*AX > 0, \geq 0, < 0, \leq 0$: definiteness of Hermitian forms, Sect. 1.5	36
dX : wedge product of differentials, Sect. 1.6	41
J : Jacobian, Sect. 1.6	41
\tilde{x}, \tilde{X} : scalar, vector/matrix variables in the complex domain, Sect. 1.6	42
$ \det(A) $: absolute value of the determinant of A , Sect. 1.6	43
$\frac{\partial}{\partial \tilde{X}}$: a real vector differential operator, Sect. 1.7	50
$N_p(\mu, \Sigma)$: p -variate real Gaussian density, $p = 1$, Sect. 2.1	57
$E[\cdot]$: expected value of (\cdot) , Sect. 2.1	58
$\text{Var}(\cdot)$: variance of (\cdot) , Sect. 2.1	58
$\text{Cov}(\cdot, \cdot)$: covariance, Sect. 2.1	58
$\tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$: p -variate complex Gaussian density, $p = 1$, Sect. 2.1	58
$\Re(\cdot)$: real part of (\cdot) , Sect. 2.1	59

$M_u(t)$: moment generating function (mgf), Sect. 2.1	59
$\text{Cov}(V)$: covariance matrix in the vector V , Sect. 2.1	60
$\Gamma(\alpha)$: gamma function, Sect. 2.1	63
χ_m^2 : real chisquare with m degrees of freedom, Sect. 2.1	63
$\tilde{\chi}_m^2$: complex chisquare with m degrees of freedom, Sect. 2.1a	65
$F_{m,n}$: real F-variable with m and n degrees of freedom, Sect. 2.1	67
$\tilde{F}_{m,n}$: complex F-variable with m and n degrees of freedom, Sect. 2.1a	70
$\chi_m^2(\lambda)$: real noncentral chisquare, Sect. 2.3	97
$(a)_m$: Pochhammer symbol, Sect. 2.3	98
${}_0F_1(\cdot)$: Bessel series, hypergeometric series, Sect. 2.3	99
$\tilde{\chi}_m^2(\lambda)$: complex noncentral chisquare, Sect. 2.3a	101
$G_{p,q}^{m,n}(\cdot)$: G-function, Sect. 2.5	110
$H_{p,q}^{m,n}(\cdot)$: H-function, Sect. 2.5	113
MLE: maximum likelihood estimator, Sect. 2.7	121
$x y$: x given y , conditional statement, Sect. 2.7	123
$Pr\{a \leq x \leq b\}$: a probability statement, Sect. 2.7	124
$\frac{\partial}{\partial \tilde{X}}$: complex vector differential operator, Sect. 3.2a	148
$\frac{\partial}{\partial Y}$: real matrix differential operator, Sect. 3.5	188
$\frac{\partial}{\partial \tilde{B}}$: complex matrix differential operator, Sect. 3.5a	192
$A^{\frac{1}{2}}$: positive definite square root, Sect. 4.2	218
$f_{p,q}(X)$: real matrix-variate density, Sect. 4.2	219
$\tilde{f}_{p,q}(\tilde{X})$: complex matrix-variate density, Sect. 4.2a	221
$\Gamma_p(\alpha)$: real matrix-variate gamma function, Sect. 4.2	224
O_p : full orthogonal group, Sect. 4.2	224
$\tilde{\Gamma}_p(\alpha)$: complex matrix-variate gamma function, Sect. 4.2a	225
\tilde{O}_p : full unitary group, Sect. 4.2a	225
$N_{p,q}(M, A, B)$: real matrix-variate Gaussian distribution, Sect. 4.5	266
$\tilde{N}_{p,q}(\tilde{M}, A, B)$: complex matrix-variate Gaussian distribution, Sect. 4.5a	270
$*T$: adjusted parameter matrix, Sect. 5.2	296
$L_f(*T)$: Laplace transform, matrix-variate case, Sect. 5.2	296
$M_X(*T)$: moment generating function, matrix-variate case, Sect. 5.2	296
$B_p(\alpha, \beta)$: real matrix-variate beta function, Sect. 5.3	302
$\tilde{B}_p(\alpha, \beta)$: complex matrix-variate beta function, Sect. 5.3a	305
${}_2F_1(\cdot)$: Gauss' hypergeometric function, Sect. 5.3	323
$J_\nu(\cdot)$: Bessel function, Sect. 5.4	333
$E_{\alpha,\beta}^\gamma(\cdot)$: Mittag-Leffler function, Sect. 5.4	334
$K_\rho^\gamma(\cdot)$: Krätzel function, Sect. 5.4	334

$W_p(m, \Sigma)$: real Wishart density, Sect. 5.5	336
$\tilde{W}_p(m, \Sigma)$: complex Wishart density, Sect. 5.5a	336
\mathbf{X} : sample matrix, Sect. 5.5	343
$\bar{\mathbf{X}}$: matrix of sample means, Sect. 5.5	343
S : sample sum of products matrix, Sect. 5.5	343
r, ρ : sample and population correlation coefficients, Sect. 5.6	356
$\rho_{1.(2\dots p)}$: multiple correlation coefficient, Sect. 5.6	358
$r_{1.(2\dots p)}$: sample multiple correlation coefficient, Sect. 5.6	358
$\rho_{12.(3\dots p)}, r_{12.(3\dots p)}$: partial correlations, Sect. 5.6	365
$K_{2,U,\gamma}^{-\alpha} f$: real Erdélyi-Kober fractional integral, second kind, Sect. 5.7	368
$K_{1,U,\gamma}^{-\alpha} f$: real Erdélyi-Kober fractional integral, first kind, Sect. 5.7	371
$\tilde{K}_{2,U,\gamma}^{-\alpha} \tilde{f}$: complex Erdélyi-Kober fractional integral, second kind, Sect. 5.7a	375
L : likelihood function, Sect. 6.1	396
H_o : null hypothesis, Sect. 6.1	396
H_1 : alternate hypothesis, Sect. 6.1	396
$\chi_{m,\alpha}^2$: chisquare tail value, Sect. 6.2	398
$F_{m,n,\alpha}$: F-tail value, Sect. 6.3	417
$t_{m,\alpha}$: Student-t tail value, Sect. 6.10	470
\sum_K : summation over $K = (k_1, \dots, k_p)$, Sect. 9.6	625
\sum_{r_1, \dots, r_p} : summation over r_1, \dots, r_p , Sect. 9.6	626
$C(i j)$: cost function, Sect. 12.1	712
$Pr\{i j, A\}$: probability of classification, Sect. 12.2	713
MANOVA: multivariate analysis of variance, Sect. 13.1	759
$x_i, x_j, x_{..}$: summation with respect to subscripts, Sect. 13.1	761
$d(X_i, X_j)$: distance measure, Sect. 15.1.2	846

Chapter 1

Mathematical Preliminaries



1.1. Introduction

It is assumed that the reader has had adequate exposure to basic concepts in Probability, Statistics, Calculus and Linear Algebra. This chapter provides a brief review of the results that will be needed in the remainder of this book. No detailed discussion of these topics will be attempted. For essential materials in these areas, the reader is, for instance, referred to Mathai and Haubold (2017a, 2017b). Some properties of vectors, matrices, determinants, Jacobians and wedge product of differentials to be utilized later on, are included in the present chapter. For the sake of completeness, we initially provide some elementary definitions. First, the concepts of vectors, matrices and determinants are introduced.

Consider the consumption profile of a family in terms of the quantities of certain food items consumed every week. The following table gives this family's consumption profile for three weeks:

Table 1.1: Consumption profile

	Rice	Lentils	Carrot	Beans
Week 1	2.00	0.50	1.00	2.00
Week 2	1.50	0.50	0.75	1.50
Week 3	2.00	0.50	0.50	1.25

All the numbers appearing in this table are in kilograms (kg). In Week 1 the family consumed 2 kg of rice, 0.5 kg of lentils, 1 kg of carrots and 2 kg of beans. Looking at the consumption over three weeks, we have an arrangement of 12 numbers into 3 rows and 4 columns. If this consumption profile is expressed in symbols, we have the following representation:

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} 2.00 & 0.50 & 1.00 & 2.00 \\ 1.50 & 0.50 & 0.75 & 1.50 \\ 2.00 & 0.50 & 0.50 & 1.25 \end{bmatrix}$$

where, for example, $a_{11} = 2.00$, $a_{13} = 1.00$, $a_{22} = 0.50$, $a_{23} = 0.75$, $a_{32} = 0.50$, $a_{34} = 1.25$.

Definition 1.1.1. A matrix An arrangement of mn items into m rows and n columns is called an m by n (written as $m \times n$) matrix.

Accordingly, the above consumption profile matrix is 3×4 (3 by 4), that is, it has 3 rows and 4 columns. The standard notation consists in enclosing the mn items within round () or square [] brackets as in the above representation. The above 3×4 matrix is represented in different ways as A , (a_{ij}) and items enclosed by square brackets. The mn items in the $m \times n$ matrix are called *elements of the matrix*. Then, in the above matrix A , a_{ij} = the i -th row, j -th column element or the (i,j) -th element. In the above illustration, $i = 1, 2, 3$ (3 rows) and $j = 1, 2, 3, 4$ (4 columns). A general $m \times n$ matrix A can be written as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}. \quad (1.1.1)$$

The elements are separated by spaces in order to avoid any confusion. Should there be any possibility of confusion, then the elements will be separated by commas. Note that the plural of “matrix” is “matrices”. Observe that the position of each element in Table 1.1 has a meaning. The elements cannot be permuted as rearranged elements will give different matrices. In other words, *two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal if and only if $a_{ij} = b_{ij}$ for all i and j , that is, they must be element-wise equal.*

In Table 1.1, the first row, which is also a 1×4 matrix, represents this family’s first week’s consumption. The fourth column represents the consumption of beans over the three weeks’ period. Thus, each row and each column in an $m \times n$ matrix has a meaning and represents different aspects. In Eq. (1.1.1), all rows are $1 \times n$ matrices and all columns are $m \times 1$ matrices. A $1 \times n$ matrix is called a *row vector* and an $m \times 1$ matrix is called a *column vector*. For example, in Table 1.1, there are 3 row vectors and 4 column vectors. If the row vectors are denoted by R_1, R_2, R_3 and the column vectors by C_1, C_2, C_3, C_4 , then we have

$$R_1 = [2.00 \ 0.50 \ 1.00 \ 2.00], R_2 = [1.50 \ 0.50 \ 0.75 \ 1.50], R_3 = [2.00 \ 0.50 \ 0.50 \ 1.25]$$

and

$$C_1 = \begin{bmatrix} 2.00 \\ 1.50 \\ 2.00 \end{bmatrix}, C_2 = \begin{bmatrix} 0.50 \\ 0.50 \\ 0.50 \end{bmatrix}, C_3 = \begin{bmatrix} 1.00 \\ 0.75 \\ 0.50 \end{bmatrix}, C_4 = \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix}.$$

If the total consumption in Week 1 and Week 2 is needed, it is obtained by adding the row vectors element-wise:

$$R_1 + R_2 = [2.00 + 1.50 \quad 0.50 + 0.50 \quad 1.00 + 0.75 \quad 2.00 + 1.50] = [3.50 \quad 1.00 \quad 1.75 \quad 3.50].$$

We will define the addition of two matrices in the same fashion as in the above illustration. For the addition to hold, both matrices must be of the same order $m \times n$. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Then the sum, denoted by $A + B$, is defined as

$$A + B = (a_{ij} + b_{ij})$$

or equivalently as the matrix obtained by adding the corresponding elements. For example,

$$C_1 + C_3 = \begin{bmatrix} 2.00 \\ 1.50 \\ 2.00 \end{bmatrix} + \begin{bmatrix} 1.00 \\ 0.75 \\ 0.50 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 2.25 \\ 2.50 \end{bmatrix}.$$

Repeating the addition, we have

$$C_1 + C_3 + C_4 = (C_1 + C_3) + \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 3.00 \\ 2.25 \\ 2.50 \end{bmatrix} + \begin{bmatrix} 2.00 \\ 1.50 \\ 1.25 \end{bmatrix} = \begin{bmatrix} 5.00 \\ 3.75 \\ 3.75 \end{bmatrix}.$$

In general, if $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, $D = (d_{ij})$ are $m \times n$ matrices, then $A + B + C + D = (a_{ij} + b_{ij} + c_{ij} + d_{ij})$, that is, it is the matrix obtained by adding the corresponding elements.

Suppose that in Table 1.1, we wish to express the elements in terms of grams instead of kilograms; then, each and every element therein must be multiplied by 1000. Thus, if A is the matrix corresponding to Table 1.1 and B is the matrix in terms of grams, we have

$$A = \begin{bmatrix} 2.00 & 0.50 & 1.00 & 2.00 \\ 1.50 & 0.50 & 0.75 & 1.50 \\ 2.00 & 0.50 & 0.50 & 1.25 \end{bmatrix},$$

$$B = \begin{bmatrix} 1000 \times 2.00 & 1000 \times 0.50 & 1000 \times 1.00 & 1000 \times 2.00 \\ 1000 \times 1.50 & 1000 \times 0.50 & 1000 \times 0.75 & 1000 \times 1.50 \\ 1000 \times 2.00 & 1000 \times 0.50 & 1000 \times 0.50 & 1000 \times 1.25 \end{bmatrix}.$$

We may write this symbolically as $B = 1000 \times A = 1000 A$. Note that 1000 is a 1×1 matrix or a scalar quantity. Any 1×1 matrix is called a scalar quantity. Then we may define *scalar multiplication* of a matrix A by the scalar quantity c as $c A$ or $A = (a_{ij}) \Rightarrow c A = (c a_{ij})$ or it is obtained by multiplying each and every element of A by the scalar quantity c . As a convention, c is written on the left of A as $c A$ and not as $A c$. Then, if $c = -1$, then $c A = (-1)A = -A$ and $A + (-1)A = A - A = O$ where the capital O denotes a matrix whose elements are all equal to zero. A general $m \times n$ matrix wherein every element is zero is referred to as a *null matrix* and it is written as O (not zero). We may also note that if A, B, C are $m \times n$ matrices, then $A + (B + C) = (A + B) + C$. Moreover, $A + O = O + A = A$. If $m = n$, in which case the number of rows is equal to the number of columns, the resulting matrix is referred to as a *square matrix* because it is a square arrangement of elements; otherwise the matrix is called a *rectangular matrix*. Some special cases of square matrices are the following: For an $n \times n$ matrix or a *square matrix of order n* , suppose that $a_{ij} = 0$ for all $i \neq j$ (that is, all non-diagonal elements are zeros; here “diagonal” means the diagonal going from top left to bottom right) and if there is at least one nonzero diagonal element, then such a matrix is called a *diagonal matrix* and it is usually written as $\text{diag}(d_1, \dots, d_n)$ where d_1, \dots, d_n are the diagonal elements. Here are some examples of 3×3 diagonal matrices:

$$D_1 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 7 \end{bmatrix}, D_2 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, D_3 = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}, a \neq 0.$$

If in D_3 , $a = 1$ so that all the diagonal elements are unities, the resulting matrix is called an *identity matrix* and a diagonal matrix whose diagonal elements are all equal to some number a that is not equal to 0 or 1, is referred to as a *scalar matrix*. A square non-null matrix $A = (a_{ij})$ that contains at least one nonzero element below its leading diagonal and whose elements above the leading diagonal are all equal to zero, that is, $a_{ij} = 0$ for all $i < j$, is called a *lower triangular matrix*. Some examples of 2×2 lower triangular matrices are the following:

$$T_1 = \begin{bmatrix} 5 & 0 \\ 2 & 1 \end{bmatrix}, T_2 = \begin{bmatrix} 3 & 0 \\ 1 & 0 \end{bmatrix}, T_3 = \begin{bmatrix} 0 & 0 \\ -3 & 0 \end{bmatrix}.$$

If, in a square non-null matrix, all elements below the leading diagonal are zeros and there is at least one nonzero element above the leading diagonal, then such a square matrix is referred to as an *upper triangular matrix*. Here are some examples:

$$T_1 = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 5 \end{bmatrix}, T_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}, T_3 = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Multiplication of Matrices Once again, consider Table 1.1. Suppose that by consuming 1 kg of rice, the family is getting 700 g (where g represents grams) of starch, 2 g protein and 1 g fat; that by eating 1 kg of lentils, the family is getting 200 g of starch, 100 g of protein and 100 g of fat; that by consuming 1 kg of carrots, the family is getting 100 g of starch, 200 g of protein and 150 g of fat; and that by eating 1 kg of beans, the family is getting 50 g of starch, 100 g of protein and 200 g of fat, respectively. Then the starch-protein-fat matrix, denoted by B , is the following where the rows correspond to rice, lentil, carrots and beans, respectively:

$$B = \begin{bmatrix} 700 & 2 & 1 \\ 200 & 100 & 100 \\ 100 & 200 & 150 \\ 50 & 100 & 200 \end{bmatrix}.$$

Let B_1, B_2, B_3 be the columns of B . Then, the first column B_1 of B represents the starch intake per kg of rice, lentil, carrots and beans respectively. Similarly, the second column B_2 represents the protein intake per kg and the third column B_3 represents the fat intake, that is,

$$B_1 = \begin{bmatrix} 700 \\ 200 \\ 100 \\ 50 \end{bmatrix}, B_2 = \begin{bmatrix} 2 \\ 100 \\ 200 \\ 100 \end{bmatrix}, B_3 = \begin{bmatrix} 1 \\ 100 \\ 150 \\ 200 \end{bmatrix}.$$

Let the rows of the matrix A in Table 1.1 be denoted by A_1, A_2 and A_3 , respectively, so that

$$A_1 = [2.00 \ 0.50 \ 1.00 \ 2.00], A_2 = [1.50 \ 0.50 \ 0.75 \ 1.50], A_3 = [2.00 \ 0.50 \ 0.50 \ 1.25].$$

Then, the total intake of starch by the family in Week 1 is available from

$$2.00 \times 700 + 0.50 \times 200 + 1.00 \times 100 + 2.00 \times 50 = 1700g.$$

This is the sum of the element-wise products of A_1 with B_1 . We will denote this by $A_1 \cdot B_1$ (A_1 dot B_1). The total intake of protein by the family in Week 1 is determined as follows:

$$A_1 \cdot B_2 = 2.00 \times 2 + 0.50 \times 100 + 1.00 \times 200 + 2.00 \times 100 = 454 \text{ g}$$

and the total intake of fat in Week 1 is given by

$$A_1 \cdot B_3 = 2.00 \times 1 + 0.50 \times 100 + 1.00 \times 150 + 2.00 \times 200 = 602 \text{ g.}$$

Thus, the dot product of A_1 with B_1, B_2, B_3 provides the intake of starch, protein and fat in Week 1. Similarly, the dot product of A_2 with B_1, B_2, B_3 gives the intake of starch, protein and fat in Week 2. Thus, the configuration of starch, protein and fat intake over the three weeks is

$$AB = \begin{bmatrix} A_1 \cdot B_1 & A_1 \cdot B_2 & A_1 \cdot B_3 \\ A_2 \cdot B_1 & A_2 \cdot B_2 & A_2 \cdot B_3 \\ A_3 \cdot B_1 & A_3 \cdot B_2 & A_3 \cdot B_3 \end{bmatrix}.$$

A matrix having one column and m rows is an $m \times 1$ matrix that is referred to as *a column vector of m elements or a column vector of order m* . A matrix having one row and n column is a $1 \times n$ matrix called *a row vector of n components or a row vector of order n* . Let A be a row or column vector of order n , which consist of n elements or components. Let the elements comprising A be denoted by a_1, \dots, a_n . Let B be a row or column vector of order n consisting of the elements b_1, \dots, b_n . Then, the dot product of A and B , denoted by $A \cdot B = B \cdot A$ is defined as $A \cdot B = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$ so that or the corresponding elements of A and B are multiplied and added up. Let A be an $m \times n$ matrix whose m rows are written as A_1, \dots, A_m . Let B be another $n \times r$ matrix whose r columns are written as B_1, \dots, B_r . Note that the number of columns of A is equal to the number of rows of B , which in this case is n . When the number of columns of A is equal to the number of rows of B , the product AB is defined and equal to

$$AB = \begin{bmatrix} A_1 \cdot B_1 & A_1 \cdot B_2 & \dots & A_1 \cdot B_r \\ A_2 \cdot B_1 & A_2 \cdot B_2 & \dots & A_2 \cdot B_r \\ \vdots & \vdots & \ddots & \vdots \\ A_m \cdot B_1 & A_m \cdot B_2 & \dots & A_m \cdot B_r \end{bmatrix} \text{ with } A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}, B = [B_1 \ B_2 \ \dots \ B_r],$$

the resulting matrix AB being of order $m \times r$. When AB is defined, BA need not be defined. However, if $r = m$, then BA is also defined, otherwise not. In other words, if $A = (a_{ij})$ is $m \times n$ and if $B = (b_{ij})$ is $n \times r$ and if $C = (c_{ij}) = AB$, then $c_{ij} = A_i \cdot B_j$ where A_i is the i -th row of A and B_j is the j -th column of B or $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$ for all i and j . For example,

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \end{bmatrix}, B = \begin{bmatrix} 2 & -2 \\ 3 & 2 \\ 1 & 0 \end{bmatrix} \Rightarrow$$

$$AB = \begin{bmatrix} (1)(2) + (-1)(3) + (0)(1) & (1)(-2) + (-1)(2) + (0)(0) \\ (2)(2) + (3)(3) + (5)(1) & 2(-2) + 3(2) + (5)(0) \end{bmatrix} = \begin{bmatrix} -1 & -4 \\ 18 & 2 \end{bmatrix}.$$

Note that, in this case, BA is defined and equal to

$$\begin{aligned} BA &= \begin{bmatrix} 2 & -2 \\ 3 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \end{bmatrix} \\ &= \begin{bmatrix} (2)(1) + (-2)(2) & (2)(-1) + (-2)(3) & (2)(0) + (-2)(5) \\ (3)(1) + (2)(2) & (3)(-1) + (2)(3) & (3)(0) + (2)(5) \\ (1)(1) + (0)(2) & (1)(-1) + (0)(3) & (1)(0) + (0)(5) \end{bmatrix} \\ &= \begin{bmatrix} -2 & -8 & -10 \\ 7 & 3 & 10 \\ 1 & -1 & 0 \end{bmatrix}. \end{aligned}$$

As another example, let

$$A = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad B = [2 \ 3 \ 5] \Rightarrow AB = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} [2 \ 3 \ 5] = \begin{bmatrix} 2 & 3 & 5 \\ -2 & -3 & -5 \\ 4 & 6 & 10 \end{bmatrix}$$

which is 3×3 , whereas BA is 1×1 :

$$BA = [2 \ 3 \ 5] \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = 9.$$

As yet another example, let

$$A = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 0 \end{bmatrix}.$$

Note that here both A and B are lower triangular matrices. The products AB and BA are defined since both A and B are 3×3 matrices. For instance,

$$\begin{aligned} AB &= \begin{bmatrix} (2)(1) + (0)(0) + (0)(1) & (2)(0) + (0)(2) + (0)(-1) & (2)(0) + (0)(0) + (0)(0) \\ (-1)(1) + (1)(0) + (0)(1) & (-1)(0) + (1)(2) + (0)(-1) & (-1)(0) + (1)(0) + (0)(0) \\ (1)(1) + (1)(0) + (1)(1) & (1)(0) + (1)(2) + (1)(-1) & (1)(0) + (1)(0) + (1)(0) \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 2 & 1 & 0 \end{bmatrix}. \end{aligned}$$

Observe that since A and B are lower triangular, AB is also lower triangular. Here are some general properties of products of matrices: When the product of the matrices is defined, in which case we say that they are conformable for multiplication,

- (1): the product of two lower triangular matrices is lower triangular;
- (2): the product of two upper triangular matrices is upper triangular;
- (3): the product of two diagonal matrices is diagonal;
- (4): if A is $m \times n$, $IA = A$ where $I = I_m$ is an identity matrix, and $A I_n = A$;
- (5): $OA = O$ whenever OA is defined, and $AO = O$ whenever AO is defined.

Transpose of a Matrix A matrix whose rows are the corresponding columns of A or, equivalently, a matrix whose columns are the corresponding rows of A is called the transpose of A denoted as A' (A prime). For example,

$$A_1 = [1 \ 2 \ -1] \Rightarrow A'_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow A'_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix};$$

$$A_3 = \begin{bmatrix} 1 & 3 \\ 3 & 7 \end{bmatrix} \Rightarrow A'_3 = \begin{bmatrix} 1 & 3 \\ 3 & 7 \end{bmatrix} = A_3; \quad A_4 = \begin{bmatrix} 0 & 5 \\ -5 & 0 \end{bmatrix} \Rightarrow A'_4 = \begin{bmatrix} 0 & -5 \\ 5 & 0 \end{bmatrix} = -A_4.$$

Observe that A_2 is lower triangular and A'_2 is upper triangular, that $A'_3 = A_3$, and that $A'_4 = -A_4$. Note that if A is $m \times n$, then A' is $n \times m$. If A is 1×1 then A' is the same scalar (1×1) quantity. If A is a square matrix and $A' = A$, then A is called a *symmetric matrix*. If B is a square matrix and $B' = -B$, then B is called a *skew symmetric matrix*. Within a skew symmetric matrix $B = (b_{ij})$, a diagonal element must satisfy the equation $b'_{jj} = -b_{jj}$, which necessitates that $b_{jj} = 0$, whether B be real or complex. Here are some properties of the transpose: The transpose of a lower triangular matrix is upper triangular; the transpose of an upper triangular matrix is lower triangular; the transpose of a diagonal matrix is diagonal; the transpose of an $m \times n$ null matrix is an $n \times m$ null matrix;

$$(A')' = A; \quad (AB)' = B'A'; \quad (A_1 A_2 \cdots A_k)' = A'_k \cdots A'_2 A'_1; \quad (A + B)' = A' + B'$$

whenever AB , $A + B$, and $A_1 A_2 \cdots A_k$ are defined.

Trace of a Square Matrix The trace is defined only for square matrices. Let $A = (a_{ij})$ be an $n \times n$ matrix whose leading diagonal elements are $a_{11}, a_{22}, \dots, a_{nn}$; then the trace of A , denoted by $\text{tr}(A)$ is defined as $\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$, that is, the sum of the elements comprising the leading diagonal. The following properties can directly be deduced from the definition. Whenever AB and BA are defined, $\text{tr}(AB) = \text{tr}(BA)$ where AB need not be equal to BA . If A is $m \times n$ and B is $n \times m$, then AB is $m \times m$ whereas

BA is $n \times n$; however, the traces are equal, that is, $\text{tr}(AB) = \text{tr}(BA)$, which implies that $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$.

Length of a Vector Let V be a $n \times 1$ real column vector or a $1 \times n$ real row vector then V can be represented as a point in n -dimensional Euclidean space when the elements are real numbers. Consider a 2-dimensional vector (or 2-vector) with the elements $(1, 2)$. Then, this vector corresponds to the point depicted in Fig. 1.1.

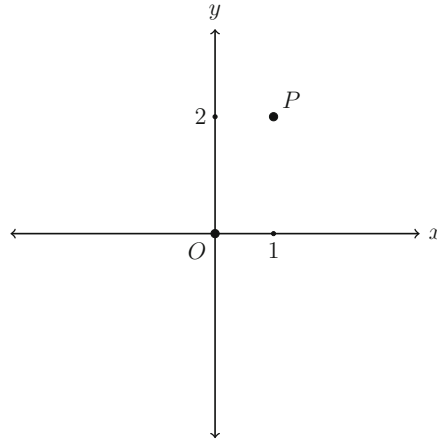


Figure 1.1 The point $P = (1, 2)$ in the plane

Let O be the origin and P be the point. Then, the length of the resulting vector is the Euclidean distance between O and P , that is, $+\sqrt{(1)^2 + (2)^2} = +\sqrt{5}$. Let $U = (u_1, \dots, u_n)$ be a real n -vector, either written as a row or a column. Then the length of U , denoted by $\|U\|$ is defined as follows:

$$\|U\| = +\sqrt{u_1^2 + \dots + u_n^2}$$

whenever the elements u_1, \dots, u_n are real. If u_1, \dots, u_n are complex numbers then $\|U\| = \sqrt{|u_1|^2 + \dots + |u_n|^2}$ where $|u_j|$ denotes the absolute value or modulus of u_j . If $u_j = a_j + ib_j$, with $i = \sqrt{-1}$ and a_j, b_j real, then $|u_j| = +\sqrt{a_j^2 + b_j^2}$. If the length of a vector is unity, that vector is called a *unit vector*. For example, $e_1 = (1, 0, \dots, 0)$, $e_2 =$

$(0, 1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ are all unit vectors. As well, $V_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $V_2 = (\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}})$ are unit vectors. If two n -vectors, U_1 and U_2 , are such that $U_1 \cdot U_2 = 0$, that is, their dot product is zero, then the two vectors are said to be *orthogonal to each other*. For example, if $U_1 = (1, 1)$ and $U_2 = (1, -1)$, then $U_1 \cdot U_2 = 0$ and U_1 and U_2 are orthogonal to each other; similarly, if $U_1 = (1, 1, 1)$ and $U_2 = (1, -2, 1)$, then $U_1 \cdot U_2 = 0$ and U_1 and U_2 are orthogonal to each other. If U_1, \dots, U_k are k vectors, each of order n , all being either row vectors or column vectors, and if $U_i \cdot U_j = 0$ for all $i \neq j$, that is, all distinct vectors are orthogonal to each other, then we say that U_1, \dots, U_k forms an *orthogonal system of vectors*. In addition, if the length of each vector is unity, $\|U_j\| = 1, j = 1, \dots, k$, then we say that U_1, \dots, U_k is an *orthonormal system of vectors*. If a matrix A is real and its rows and its columns form an orthonormal system, then A is called an *orthonormal matrix*. In this case, $AA' = I_n$ and $A'A = I_n$; accordingly, any square matrix A of real elements such that $AA' = I_n$ and $A'A = I_n$ is referred to as an *orthonormal matrix*. If only one equation holds, that is, B is a real matrix such that either $BB' = I, B'B \neq I$ or $B'B = I, BB' \neq I$, then B is called a *semiorthonormal matrix*. For example, consider the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}; \text{ then } AA' = I_2, A'A = I_2,$$

and A is an orthonormal matrix. As well,

$$A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \Rightarrow AA' = I_3, A'A = I_3,$$

and A here is orthonormal. However,

$$B = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow BB' = I_2, B'B \neq I$$

so that B is semiorthonormal. On deleting some rows from an orthonormal matrix, we obtain a semiorthonormal matrix such that $BB' = I$ and $B'B \neq I$. Similarly, if we delete some of the columns, we end up with a semiorthonormal matrix such that $B'B = I$ and $BB' \neq I$.

Linear Independence of Vectors Consider the vectors $U_1 = (1, 1, 1), U_2 = (1, -2, 1), U_3 = (3, 0, 3)$. Then, we can easily see that $U_3 = 2U_1 + U_2 = 2(1, 1, 1) + (1, -2, 1) = (3, 0, 3)$ or $U_3 - 2U_1 - U_2 = O$ (a null vector). In this case, one of the vectors can be written

as a linear function of the others. Let $V_1 = (1, 1, 1)$, $V_2 = (1, 0, -1)$, $V_3 = (1, -2, 1)$. Can any one of these be written as a linear function of others? If that were possible, then there would exist a linear function of V_1, V_2, V_3 that is equal to is a null vector. Let us consider the equation $a_1 V_1 + a_2 V_2 + a_3 V_3 = (0, 0, 0)$ where a_1, a_2, a_3 are scalars where at least one of them is nonzero. Note that $a_1 = 0, a_2 = 0$ and $a_3 = 0$ will always satisfy the above equation. Thus, our question is whether $a_1 = 0, a_2 = 0, a_3 = 0$ is the only solution.

$$\begin{aligned} a_1 V_1 + a_2 V_2 + a_3 V_3 = O &\Rightarrow a_1(1, 1, 1) + a_2(1, 0, -1) + a_3(1, -2, 1) = (0, 0, 0) \\ &\Rightarrow a_1 + a_2 + a_3 = 0 \text{ (i); } a_1 - 2a_3 = 0 \text{ (ii); } a_1 - a_2 + a_3 = 0. \end{aligned} \quad (iii)$$

From (ii), $a_1 = 2a_3$. Then, from (iii), $3a_3 - a_2 = 0 \Rightarrow a_2 = 3a_3$; then from (i), $2a_3 + 3a_3 + a_3 = 0$ or $6a_3 = 0$ or $a_3 = 0$. Thus, $a_2 = 0, a_1 = 0$ and there is no nonzero a_1 or a_2 or a_3 satisfying the equation and hence V_1, V_2, V_3 cannot be linearly dependent; so, they are linearly independent. Hence, we have the following definition: Let U_1, \dots, U_k be k vectors, each of order n , all being either row vectors or column vectors, so that addition and linear functions are defined. Let a_1, \dots, a_k be scalar quantities. Consider the equation

$$a_1 U_1 + a_2 U_2 + \dots + a_k U_k = O \text{ (a null vector)}. \quad (iv)$$

If $a_1 = 0, a_2 = 0, \dots, a_k = 0$ is the only solution to (iv), then U_1, \dots, U_k are linearly independent, otherwise they are linearly dependent. If they are linearly dependent, then at least one of the vectors can be expressed as a linear function of others. The following properties can be established from the definition: Let U_1, \dots, U_k be n -vectors, $k \leq n$.

- (1) If U_1, \dots, U_k are mutually orthogonal, then they are linearly independent, that is, if $U_i \cdot U_j = 0$, for all $i \neq j$, then U_1, \dots, U_k are linearly independent;
- (2) There cannot be more than n mutually orthogonal n -vectors;
- (3) There cannot be more than n linearly independent n -vectors.

Rank of a Matrix The maximum number of linearly independent row vectors of a $m \times n$ matrix is called the *row rank of the matrix*; the maximum number of linearly independent column vectors is called the *column rank of the matrix*. It can be shown that the row rank of any matrix is equal to its column rank, and this common rank is called the *rank of the matrix*. If r is the rank of a $m \times n$ matrix, then $r \leq m$ and $r \leq n$. If $m \leq n$ and the rank is m or if $n \leq m$ and the rank is n , then the matrix is called a *full rank matrix*. A square matrix of full rank is called a *nonsingular matrix*. When the rank of an $n \times n$ matrix is $r < n$, this matrix is referred to as a *singular matrix*. Singularity is defined only for square matrices. The following properties clearly hold:

- (1) *A diagonal matrix with at least one zero diagonal element is singular or a diagonal matrix with all nonzero diagonal elements is nonsingular;*
- (2) *A triangular matrix (upper or lower) with at least one zero diagonal element is singular or a triangular matrix with all diagonal elements nonzero is nonsingular;*
- (3) *A square matrix containing at least one null row vector or at least one null column vector is singular;*
- (4) *Linear independence or dependence in a collection of vectors of the same order and category (either all are row vectors or all are column vectors) is not altered by multiplying any of the vectors by a nonzero scalar;*
- (5) *Linear independence or dependence in a collection of vectors of the same order and category is not altered by adding any vector of the set to any other vector in the same set;*
- (6) *Linear independence or dependence in a collection of vectors of the same order and category is not altered by adding a linear combination of vectors from the same set to any other vector in the same set;*
- (7) *If a collection of vectors of the same order and category is a linearly dependent system, then at least one of the vectors can be made null by the operations of scalar multiplication and addition.*

Note: We have defined “vectors” as an ordered set of items such as an ordered set of numbers. One can also give a general definition of a vector as an element in a set S which is closed under the operations of scalar multiplication and addition (these operations are to be defined on S), that is, letting S be a set of items, if $V_1 \in S$ and $V_2 \in S$, then $cV_1 \in S$ and $V_1 + V_2 \in S$ for all scalar c and for all V_1 and V_2 , that is, if V_1 is an element in S , then cV_1 is also an element in S and if V_1 and V_2 are in S , then $V_1 + V_2$ is also in S , where operations cV_1 and $V_1 + V_2$ are to be properly defined. One can impose additional conditions on S . However, for our discussion, the notion of vectors as ordered set of items will be sufficient.

1.2. Determinants

Determinants are defined only for square matrices. They are certain scalar functions of the elements of the square matrix under consideration. We will motivate this particular function by means of an example that will also prove useful in other areas. Consider two 2-vectors, either both row vectors or both column vectors. Let

$U = OP$ and $V = OQ$ be the two vectors as shown in Fig. 1.2. If the vectors are separated by a nonzero angle θ then one can create the parallelogram $OPSQ$ with these two vectors as shown in Fig. 1.2.

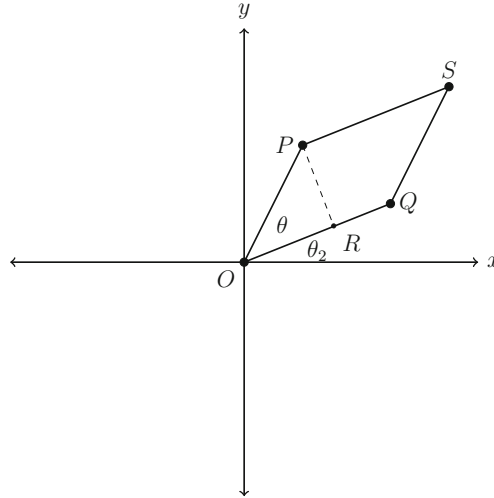


Figure 1.2 Parallelogram generated from two vectors

The area of the parallelogram is twice the area of the triangle OPQ . If the perpendicular from P to OQ is PR , then the area of the triangle is $\frac{1}{2}PR \times OQ$ or the area of the parallelogram $OPSQ$ is $PR \times \|V\|$ where PR is $OP \times \sin \theta = \|U\| \times \sin \theta$. Therefore the area is $\|U\| \|V\| \sin \theta$ or the area, denoted by v is

$$v = \|U\| \|V\| \sqrt{(1 - \cos^2 \theta)}.$$

If θ_1 is the angle U makes with the x -axis and θ_2 , the angle V makes with the x -axis, then if U and V are as depicted in Fig. 1.2, then $\theta = \theta_1 - \theta_2$. It follows that

$$\cos \theta = \cos(\theta_1 - \theta_2) = \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 = \frac{U \cdot V}{\|U\| \|V\|},$$

as can be seen from Fig. 1.2. In this case,

$$v = \|U\| \|V\| \sqrt{1 - \left(\frac{U \cdot V}{\|U\| \|V\|}\right)^2} = \sqrt{(\|U\|)^2 (\|V\|)^2 - (U \cdot V)^2} \tag{1.2.1}$$

and

$$v^2 = (\|U\|)^2(\|V\|)^2 - (U \cdot V)^2. \quad (1.2.2)$$

This can be written in a more convenient way. Letting $X = \begin{pmatrix} U \\ V \end{pmatrix}$,

$$XX' = \begin{pmatrix} U \\ V \end{pmatrix} (U' \ V') = \begin{bmatrix} UU' & UV' \\ VU' & VV' \end{bmatrix} = \begin{bmatrix} U \cdot U & U \cdot V \\ V \cdot U & V \cdot V \end{bmatrix}. \quad (1.2.3)$$

On comparing (1.2.2) and (1.2.3), we note that (1.2.2) is available from (1.2.3) by taking a scalar function of the following type. Consider a matrix

$$C = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \text{ then (1.2.2) is available by taking } ad - bc$$

where a, b, c, d are scalar quantities. A scalar function of this type is the determinant of the matrix C .

A general result can be deduced from the above procedure: If U and V are n -vectors and if θ is the angle between them, then

$$\cos \theta = \frac{U \cdot V}{\|U\| \|V\|}$$

or the dot product of U and V divided by the product of their lengths when $\theta \neq 0$, and the numerator is equal to the denominator when $\theta = 2n\pi$, $n = 0, 1, 2, \dots$. We now provide a formal definition of the determinant of a square matrix.

Definition 1.2.1. The Determinant of a Square Matrix Let $A = (a_{ij})$ be a $n \times n$ matrix whose rows (columns) are denoted by $\alpha_1, \dots, \alpha_n$. For example, if α_i is the i -th row vector, then

$$\alpha_i = (a_{i1} \ a_{i2} \ \dots \ a_{in}).$$

The determinant of A will be denoted by $|A|$ or $\det(A)$ when A is real or complex and the absolute value of the determinant of A will be denoted by $|\det(A)|$ when A is in the complex domain. Then, $|A|$ will be a function of $\alpha_1, \dots, \alpha_n$, written as

$$|A| = \det(A) = f(\alpha_1, \dots, \alpha_i, \dots, \alpha_j, \dots, \alpha_n),$$

which will be defined by the following four axioms (postulates or assumptions): (this definition also holds if the elements of the matrix are in the complex domain)

$$(1) \quad f(\alpha_1, \dots, c\alpha_i, \dots, \alpha_n) = cf(\alpha_1, \dots, \alpha_i, \dots, \alpha_n),$$

which is equivalent to saying that if any row (column) is multiplied by a scalar quantity c (including zero), then the whole determinant is multiplied by c ;

$$(2) \quad f(\alpha_1, \dots, \alpha_i, \dots, \alpha_i + \alpha_j, \dots, \alpha_n) = f(\alpha_1, \dots, \alpha_i, \dots, \alpha_j, \dots, \alpha_n),$$

which is equivalent to saying that if any row (column) is added to any other row (column), then the value of the determinant remains the same;

$$(3) \quad f(\alpha_1, \dots, \gamma_i + \delta_i, \dots, \alpha_n) = f(\alpha_1, \dots, \gamma_i, \dots, \alpha_n) + f(\alpha_1, \dots, \delta_i, \dots, \alpha_n),$$

which is equivalent to saying that if any row (column), say the i -th row (column) is written as a sum of two vectors, $\alpha_i = \gamma_i + \delta_i$ then the determinant becomes the sum of two determinants such that γ_i appears at the position of α_i in the first one and δ_i appears at the position of α_i in the second one;

$$(4) \quad f(e_1, \dots, e_n) = 1$$

where e_1, \dots, e_n are the basic unit vectors as previously defined; this axiom states that the determinant of an identity matrix is 1.

Let us consider some corollaries resulting from Axioms (1) to (4). On combining Axioms (1) and (2), we have that the value of a determinant remains unchanged if a linear function of any number of rows (columns) is added to any other row (column). As well, the following results are direct consequences of the axioms.

(i): *The determinant of a diagonal matrix is the product of the diagonal elements* [which can be established by repeated applications of Axiom (1)];

(ii): *If any diagonal element in a diagonal matrix is zero, then the determinant is zero, and thereby the corresponding matrix is singular; if none of the diagonal elements of a diagonal matrix is equal to zero, then the matrix is nonsingular.*

(iii): *If any row (column) of a matrix is null, then the determinant is zero or the matrix is singular* [Axiom (1)];

(iv): *If any row (column) is a linear function of other rows (columns), then the determinant is zero* [By Axioms (1) and (2), we can reduce that row (column) to a null vector]. Thus, the determinant of a singular matrix is zero or if the row (column) vectors form a linearly dependent system, then the determinant is zero.

By using Axioms (1) and (2), we can reduce a triangular matrix to a diagonal form when evaluating its determinant. For this purpose we shall use the following standard notation: " $c(i) + (j) \Rightarrow$ " means " c times the i -th row is added to the j -th row which

results in the following:” Let us consider a simple example. Consider a triangular matrix and its determinant. Evaluate the determinant of the following matrix:

$$T = \begin{bmatrix} 2 & 1 & 5 \\ 0 & 3 & 4 \\ 0 & 0 & -4 \end{bmatrix}.$$

It is an upper triangular matrix. We take out -4 from the third row by using Axiom (1). Then,

$$|T| = -4 \begin{vmatrix} 2 & 1 & 5 \\ 0 & 3 & 4 \\ 0 & 0 & 1 \end{vmatrix}.$$

Now, add (-4) times the third row to the second row and (-5) times the third row to the first row. This in symbols is “ $-4(3) + (2)$, $-5(3) + (1) \Rightarrow$ ”. The net result is that the elements 5 and 4 in the last column are eliminated without affecting the other elements, so that

$$|T| = -4 \begin{vmatrix} 2 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Now take out 3 from the second row and then use the second row to eliminate 1 in the first row. After taking out 3 from the second row, the operation is “ $-1(2) + (1) \Rightarrow$ ”. The result is the following:

$$|T| = (-4)(3) \begin{vmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Now, take out 2 from the first row, then by Axiom (4) the determinant of the resulting identity matrix is 1, and hence $|T|$ is nothing but the product of the diagonal elements. Thus, we have the following result:

(v): *The determinant of a triangular matrix (upper or lower) is the product of its diagonal elements; accordingly, if any diagonal element in a triangular matrix is zero, then the determinant is zero and the matrix is singular.* For a triangular matrix to be nonsingular, all its diagonal elements must be non-zeros.

The following result follows directly from Axioms (1) and (2). The proof is given in symbols.

(vi): *If any two rows (columns) are interchanged (this means one transposition), then the resulting determinant is multiplied by -1 or every transposition brings in -1 outside that determinant as a multiple.* If an odd number of transpositions are done, then the whole

determinant is multiplied by -1 , and for even number of transpositions, the multiplicative factor is $+1$ or no change in the determinant. An outline of the proof follows:

$$\begin{aligned}
 |A| &= f(\alpha_1, \dots, \alpha_i, \dots, \alpha_j, \dots, \alpha_n) \\
 &= f(\alpha_1, \dots, \alpha_i, \dots, \alpha_i + \alpha_j, \dots, \alpha_n) \quad [\text{Axiom (2)}] \\
 &= -f(\alpha_1, \dots, \alpha_i, \dots, -\alpha_i - \alpha_j, \dots, \alpha_n) \quad [\text{Axiom (1)}] \\
 &= -f(\alpha_1, \dots, -\alpha_j, \dots, -\alpha_i - \alpha_j, \dots, \alpha_n) \quad [\text{Axiom (2)}] \\
 &= f(\alpha_1, \dots, \alpha_j, \dots, -\alpha_i - \alpha_j, \dots, \alpha_n) \quad [\text{Axiom (1)}] \\
 &= f(\alpha_1, \dots, \alpha_j, \dots, -\alpha_i, \dots, \alpha_n) \quad [\text{Axiom (2)}] \\
 &= -f(\alpha_1, \dots, \alpha_j, \dots, \alpha_i, \dots, \alpha_n) \quad [\text{Axiom (1)}].
 \end{aligned}$$

Now, note that the i -th and j -th rows (columns) are interchanged and the result is that the determinant is multiplied by -1 .

With the above six basic properties, we are in a position to evaluate most of the determinants.

Example 1.2.1. Evaluate the determinant of the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 2 & -1 & 1 & 0 \\ 3 & 0 & 1 & 4 \end{bmatrix}.$$

Solution 1.2.1. Since, this is a triangular matrix, its determinant will be product of its diagonal elements. Proceeding step by step, take out 2 from the first row by using Axiom (1). Then $-1(1) + (2)$, $-2(1) + (3)$, $-3(1) + (3) \Rightarrow$. The result of these operations is the following:

$$|A| = 2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 4 \end{vmatrix}.$$

Now, take out 5 from the second row so that $1(2) + (3) \Rightarrow$, the result being the following:

$$|A| = (2)(5) \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 4 \end{vmatrix}$$

The diagonal element in the third row is 1 and there is nothing to be taken out. Now $-1(3) + (4) \Rightarrow$ and then, after having taken out 4 from the fourth row, the result is

$$|A| = (2)(5)(1)(4) \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

Now, by Axiom (4) the determinant of the remaining identity matrix is 1. Therefore, the final solution is $|A| = (2)(5)(1)(4) = 40$.

Example 1.2.2. Evaluate the determinant of the following matrix:

$$A = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 2 & 1 & -1 & 0 \\ 5 & 2 & 1 & 3 \end{bmatrix}.$$

Solution 1.2.2. Since the first row, first column element is a convenient number 1 we start operating with the first row. Otherwise, we bring a convenient number to the (1, 1)-th position by interchanges of rows and columns (with each interchange the determinant is to be multiplied by (-1)). Our aim will be to reduce the matrix to a triangular form so that the determinant is the product of the diagonal elements. By using the first row let us wipe out the elements in the first column. The operations are $-2(1) + (3)$, $-5(1) + (4) \Rightarrow$. Then

$$|A| = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 2 & 1 & -1 & 0 \\ 5 & 2 & 1 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & -3 & -9 & -2 \\ 0 & -8 & -19 & -2 \end{vmatrix}.$$

Now, by using the second row we want to wipe out the elements below the diagonal in the second column. But the first number is 3. One element in the third row can be wiped out by simply adding $1(2) + (3) \Rightarrow$. This brings the following:

$$|A| = \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & -7 & -1 \\ 0 & -8 & -19 & -2 \end{vmatrix}.$$

If we take out 3 from the second row then it will bring in fractions. We will avoid fractions by multiplying the second row by 8 and the fourth row by 3. In order preserve the value,

we keep $\frac{1}{(8)(3)}$ outside. Then, we add the second row to the fourth row or $(2) + (4) \Rightarrow$. The result of these operations is the following:

$$|A| = \frac{1}{(8)(3)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -7 & -1 \\ 0 & -24 & -57 & -6 \end{vmatrix} = \frac{1}{(8)(3)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -7 & -1 \\ 0 & 0 & -41 & 2 \end{vmatrix}.$$

Now, multiply the third row by 41 and fourth row by 7 and then add $-1(3) + (4) \Rightarrow$. The result is the following:

$$|A| = \frac{1}{(8)(3)(7)(41)} \begin{vmatrix} 1 & 2 & 4 & 1 \\ 0 & 24 & 16 & 8 \\ 0 & 0 & -287 & -41 \\ 0 & 0 & 0 & 55 \end{vmatrix}.$$

Now, take the product of the diagonal elements. Then

$$|A| = \frac{(1)(24)(-287)(55)}{(8)(3)(7)(41)} = -55.$$

Observe that we did not have to repeat the 4×4 determinant each time. After wiping out the first column elements, we could have expressed the determinant as follows because only the elements in the second row and second column onward would then have mattered. That is,

$$|A| = (1) \begin{vmatrix} 3 & 2 & 1 \\ 0 & -7 & -1 \\ -8 & -19 & -2 \end{vmatrix}.$$

Similarly, after wiping out the second column elements, we could have written the resulting determinant as

$$|A| = \frac{(1)(24)}{(8)(3)} \begin{vmatrix} -7 & -1 \\ -41 & 2 \end{vmatrix},$$

and so on.

Example 1.2.3. Evaluate the determinant of a 2×2 general matrix.

Solution 1.2.3. A general 2×2 determinant can be opened up by using Axiom (3), that is,

$$\begin{aligned} |A| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad [\text{Axiom (3)}] \\ &= a_{11} \begin{vmatrix} 1 & 0 \\ a_{21} & a_{22} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 \\ a_{21} & a_{22} \end{vmatrix} \quad [\text{Axiom (1)}]. \end{aligned}$$

If any of a_{11} or a_{12} is zero, then the corresponding determinant is zero. In the second determinant on the right, interchange the second and first columns, which will bring a minus sign outside the determinant. That is,

$$|A| = a_{11} \begin{vmatrix} 1 & 0 \\ a_{21} & a_{22} \end{vmatrix} - a_{12} \begin{vmatrix} 1 & 0 \\ a_{22} & a_{21} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The last step is done by using the property that the determinant of a triangular matrix is the product of the diagonal elements. We can also evaluate the determinant by using a number of different procedures. Taking out a_{11} if $a_{11} \neq 0$,

$$|A| = a_{11} \begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} \\ a_{21} & a_{22} \end{vmatrix}.$$

Now, perform the operation $-a_{21}(1) + (2)$ or $-a_{21}$ times the first row is added to the second row. Then,

$$|A| = a_{11} \begin{vmatrix} 1 & \frac{a_{12}}{a_{11}} \\ 0 & a_{22} - \frac{a_{12}a_{21}}{a_{11}} \end{vmatrix}.$$

Now, expanding by using a property of triangular matrices, we have

$$|A| = a_{11}(1) \left[a_{22} - \frac{a_{12}a_{21}}{a_{11}} \right] = a_{11}a_{22} - a_{12}a_{21}. \quad (1.2.4)$$

Consider a general 3×3 determinant evaluated by using Axiom (3) first.

$$\begin{aligned} |A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11} \begin{vmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 0 & 0 & 1 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + a_{12} \begin{vmatrix} 0 & 1 & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 0 & 0 & 1 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}. \end{aligned}$$

The first step consists in opening up the first row by making use of Axiom (3). Then, eliminate the elements in rows 2 and 3 within the column headed by 1. The next step is to bring the columns whose first element is 1 to the first column position by transpositions. The first matrix on the right-hand side is already in this format. One transposition is needed in the second matrix and two are required in the third matrix. After completing the transpositions, the next step consists in opening up each determinant along their second row and observing that the resulting matrices are lower triangular or can be made so after transposing their last two columns. The final result is then obtained. The last two steps are executed below:

$$\begin{aligned}
 |A| &= a_{11} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{21} & a_{23} \\ 0 & a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} 1 & 0 & 0 \\ 0 & a_{21} & a_{22} \\ 0 & a_{31} & a_{32} \end{vmatrix} \\
 &= a_{11}[a_{22}a_{33} - a_{23}a_{32}] - a_{12}[a_{21}a_{33} - a_{23}a_{31}] + a_{13}[a_{21}a_{32} - a_{22}a_{31}] \\
 &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.
 \end{aligned} \tag{1.2.5}$$

A few observations are in order. Once 1 is brought to the first row first column position in every matrix and the remaining elements in this first column are eliminated, one can delete the first row and first column and take the determinant of the remaining submatrix because only those elements will enter into the remaining operations involving opening up the second and successive rows by making use of Axiom (3). Hence, we could have written

$$|A| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

This step is also called *the cofactor expansion of the matrix*. In a general matrix $A = (a_{ij})$, the cofactor of the element a_{ij} is equal to $(-1)^{i+j} M_{ij}$ where M_{ij} is the minor of a_{ij} . This minor is obtained by deleting the i -th row and j -th column and then taking the determinant of the remaining elements. The second item to be noted from (1.2.5) is that, in the final expression for $|A|$, each term has one and only one element from each row and each column of A . Some elements have plus signs in front of them and others have minus signs. For each term, write the first subscript in the natural order 1, 2, 3 for the 3×3 case and in the general $n \times n$ case, write the first subscripts in the natural order 1, 2, \dots , n . Now, examine the second subscripts. Let the number of transpositions needed to bring the second subscripts into the natural order 1, 2, \dots , n be ρ . Then, that term is multiplied by $(-1)^\rho$ so that an even number of transpositions produces a plus sign and an odd number of transpositions brings a minus sign, or equivalently if ρ is even, the coefficient is plus 1 and if ρ is odd, the coefficient is -1 . This also enables us to open up a general determinant. This will be considered after pointing out one more property for

a 3×3 case. The final representation in the 3×3 case in (1.2.5) can also be written up by using the following mechanical procedure. Write all elements in the matrix A in the natural order. Then, augment this arrangement with the first two columns. This yields the following format:

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} . \end{array}$$

Now take the products of the elements along the diagonals going from the top left to the bottom right. These are the elements with the plus sign. Take the products of the elements in the second diagonals or the diagonals going from the bottom left to the top right. These are the elements with minus sign. As a result, $|A|$ is as follows:

$$\begin{aligned} |A| &= [a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}] \\ &\quad - [a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}]. \end{aligned}$$

This mechanical procedure applies only in the 3×3 case. The general expansion is the following:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum_{i_1} \dots \sum_{i_n} (-1)^{\rho(i_1, \dots, i_n)} a_{1i_1} a_{2i_2} \dots a_{ni_n} \quad (1.2.6)$$

where $\rho(i_1, \dots, i_n)$ is the number of transpositions needed to bring the second subscripts i_1, \dots, i_n into the natural order $1, 2, \dots, n$.

The cofactor expansion of a general matrix is obtained as follows: Suppose that we open up a $n \times n$ determinant A along the i -th row using Axiom (3). Then, after taking out $a_{i1}, a_{i2}, \dots, a_{in}$, we obtain n determinants where, in the first determinant, 1 occupies the $(i, 1)$ -th position, in the second one, 1 is at the $(i, 2)$ -th position and so on so that, in the j -th determinant, 1 occupies the (i, j) -th position. Given that i -th row, we can now eliminate all the elements in the columns corresponding to the remaining 1. We now bring this 1 into the first row first column position by transpositions in each determinant. The number of transpositions needed to bring this 1 from the j -th position in the i -th row to the first position in the i -th row, is $j - 1$. Then, to bring that 1 to the first row first column position, another $i - 1$ transpositions are required, so that the total number of transpositions needed is $(i - 1) + (j - 1) = i + j - 2$. Hence, the multiplicative factor is $(-1)^{i+j-2} = (-1)^{i+j}$, and the expansion is as follows:

$$\begin{aligned} |A| &= (-1)^{i+1} a_{i1} M_{i1} + (-1)^{i+2} a_{i2} M_{i2} + \dots + (-1)^{i+n} a_{in} M_{in} \\ &= a_{i1} C_{i1} + a_{i2} C_{i2} + \dots + a_{in} C_{in} \end{aligned} \quad (1.2.7)$$

where $C_{ij} = (-1)^{i+j} M_{ij}$, C_{ij} is the cofactor of a_{ij} and M_{ij} is the minor of a_{ij} , the minor being obtained by taking the determinant of the remaining elements after deleting the i -th row and j -th column of A . Moreover, if we expand along a certain row (column) and the cofactors of some other row (column), then the result will be zero. That is,

$$0 = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}, \text{ for all } i \neq j. \quad (1.2.8)$$

Inverse of a Matrix Regular inverses exist only for square matrices that are nonsingular. The standard notation for a regular inverse of a matrix A is A^{-1} . It is defined as $AA^{-1} = I_n$ and $A^{-1}A = I_n$. The following properties can be deduced from the definition. First, we note that $AA^{-1} = A^0 = I = A^{-1}A$. When A and B are $n \times n$ nonsingular matrices, then $(AB)^{-1} = B^{-1}A^{-1}$, which can be established by pre- or post-multiplying the right-hand side by AB . Accordingly, with $A^m = A \times A \times \cdots \times A$, $A^{-m} = A^{-1} \times \cdots \times A^{-1} = (A^m)^{-1}$, $m = 1, 2, \dots$, and when A_1, \dots, A_k are $n \times n$ nonsingular matrices, $(A_1A_2 \cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1} \cdots A_2^{-1}A_1^{-1}$. We can also obtain a formula for the inverse of a nonsingular matrix A in terms of cofactors. Assuming that A^{-1} exist and letting $\text{Cof}(A) = (C_{ij})$ be the matrix of cofactors of A , that is, if $A = (a_{ij})$ and if C_{ij} is the cofactor of a_{ij} then $\text{Cof}(A) = (C_{ij})$. It follows from (1.2.7) and (1.2.8) that

$$A^{-1} = \frac{1}{|A|}(\text{Cof}(A))' = \frac{1}{|A|} \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}', \quad (1.2.9)$$

that is, the transpose of the cofactor matrix divided by the determinant of A . What about $A^{\frac{1}{2}}$? For a scalar quantity a , we have the definition that if b exists such that $b \times b = a$, then b is a square root of a . Consider the following 2×2 matrices:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, B_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, B_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I_2^2 = I_2, B_j^2 = I_2, j = 1, 2, 3, 4.$$

Thus, if we use the definition $B^2 = A$ and claim that B is the square root of A , there are several candidates for B ; this means that, in general, the square root of a matrix cannot be uniquely determined. However, if we restrict ourselves to the class of positive definite matrices, then a square root can be uniquely defined. The definiteness of matrices will be considered later.

1.2.1. Inverses by row operations or elementary operations

Basic elementary matrices are of two types. Let us call them the E -type and the F -type. An elementary matrix of the E -type is obtained by taking an identity matrix and multiplying any row (column) by a nonzero scalar. For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_2 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

where E_1, E_2, E_3 are elementary matrices of the E -type obtained from the identity matrix I_3 . If we pre-multiply an arbitrary matrix A with an elementary matrix of the E -type, then the same effect will be observed on the rows of the arbitrary matrix A . For example, consider a 3×3 matrix $A = (a_{ij})$. Then, for example,

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ -2a_{21} & -2a_{22} & -2a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Thus, the same effect applies to the rows, that is, the second row is multiplied by (-2) . Observe that E -type elementary matrices are always nonsingular and so, their regular inverses exist. For instance,

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_1 E_1^{-1} = I_3 = E_1^{-1} E_1; E_2^{-1} = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_2 E_2^{-1} = I_3.$$

Observe that post-multiplication of an arbitrary matrix by an E -type elementary matrix will have the same effect on the columns of the arbitrary matrix. For example, $A E_1$ will have the same effect on the columns of A , that is, the second column of A is multiplied by -2 ; $A E_2$ will result in the first column of A being multiplied by 5 , and so on. The F -type elementary matrix is created by adding any particular row of an identity matrix to another one of its rows. For example, consider a 3×3 identity matrix I_3 and let

$$F_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, F_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, F_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

where F_1 is obtained by adding the first row to the second row of I_3 ; F_2 is obtained by adding the first row to the third row of I_3 ; and F_3 is obtained by adding the second row of I_3 to the third row. As well, F -type elementary matrices are nonsingular, and for instance,

$$F_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad F_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

where $F_1 F_1^{-1} = I_3$, $F_2^{-1} F_2 = I_3$ and $F_3^{-1} F_3 = I_3$. If we pre-multiply an arbitrary matrix A by an F -type elementary matrix, then the same effect will be observed on the rows of A . For example,

$$F_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + a_{11} & a_{22} + a_{12} & a_{23} + a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Thus, the same effect applies to the rows, namely, the first row is added to the second row in A (as F_1 was obtained by adding the first row of I_3 to the second row of I_3). The reader may verify that $F_2 A$ has the effect of the first row being added to the third row and $F_3 A$ will have the effect of the second row being added to the third row. By combining E - and F -type elementary matrices, we end up with a G -type matrix wherein a multiple of any particular row of an identity matrix is added to another one of its rows. For example, letting

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad G_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix},$$

it is seen that G_1 is obtained by adding 5 times the first row to the second row in I_3 , and G_2 is obtained by adding -2 times the first row to the third row in I_3 . Pre-multiplication of an arbitrary matrix A by G_1 , that is, $G_1 A$, will have the effect that 5 times the first row of A will be added to its second row. Similarly, G_2 will have the effect that -2 times the first row of A will be added to its third row. Being product of E - and F -type elementary matrices, G -type matrices are also nonsingular. We also have the result that if A, B, C are $n \times n$ matrices and $B = C$, then $AB = AC$ as long as A is nonsingular. In general, if A_1, \dots, A_k are $n \times n$ nonsingular matrices, we have

$$\begin{aligned} B = C &\Rightarrow A_k A_{k-1} \cdots A_2 A_1 B = A_k A_{k-1} \cdots A_2 A_1 C; \\ &\Rightarrow A_1 A_2 B = A_1 (A_2 B) = (A_1 A_2) B = (A_1 A_2) C = A_1 (A_2 C). \end{aligned}$$

We will evaluate the inverse of a nonsingular square matrix by making use of elementary matrices. The procedure will also verify whether a regular inverse exists for a given matrix. If a regular inverse for a square matrix A exists, then $AA^{-1} = I$. We can pre- or post-multiply A by elementary matrices. For example,

$$\begin{aligned} AA^{-1} = I &\Rightarrow E_k F_r \cdots E_1 F_1 AA^{-1} = E_k F_r \cdots E_1 F_1 I \\ &\Rightarrow (E_k \cdots F_1 A) A^{-1} = (E_k \cdots F_1). \end{aligned}$$

Thus, if the operations $E_k \cdots F_1$ on A reduced A to an identity matrix, then A^{-1} is $E_k \cdots F_1$. If an inconsistency has occurred during the process, we can conclude that there is no inverse for A . Hence, our aim in performing our elementary operations on the left of A is to reduce it to an identity matrix, in which case the product of the elementary matrices on the right-hand side of the last equation will produce the inverse of A .

Example 1.2.4. Evaluate A^{-1} if it exists, where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & -1 & 1 \end{bmatrix}.$$

Solution 1.2.4. If A^{-1} exists then $AA^{-1} = I$ which means

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & -1 & 1 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is our starting equation. Only the configuration of the elements matters. The matrix notations and the symbol A^{-1} can be disregarded. Hence, we consider only the configuration of the numbers of the matrix A on the left and the numbers in the identity matrix on the right. Then we pre-multiply A and pre-multiply the identity matrix by only making use of elementary matrices. In the first set of steps, our aim consists in reducing every element in the first column of A to zeros, except the first one, by only using the first row. For each elementary operation on A , the same elementary operation is done on the identity matrix also. Now, utilizing the second row of the resulting A , we reduce all the elements in the second column of A to zeros except the second one and continue in this manner until all the elements in the last columns except the last one are reduced to zeros by making use of the last row, thus reducing A to an identity matrix, provided of course that A is nonsingular. In our example, the elements in the first column can be made equal to zeros

by applying the following operations. We will employ the following standard notation: $a(i) + (j) \Rightarrow$ meaning a times the i -th row is added to the j -th row, giving the result. Consider $(1) + (2)$; $-2(1) + (3)$; $-1(1) + (4) \Rightarrow$ (that is, the first row is added to the second row; and then -2 times the first row is added to the third row; then -1 times the first row is added to the fourth row), (for each elementary operation on A we do the same operation on the identity matrix also) the net result being

$$\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & -2 & 0 & -1 & 0 & 0 & 1 \end{array} \Leftrightarrow$$

Now, start with the second row of the resulting A and the resulting identity matrix and try to eliminate all the other elements in the second column of the resulting A . This can be achieved by performing the following operations: $(2) + (3)$; $-1(2) + (1) \Rightarrow$

$$\begin{array}{cccc|cccc} 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & -2 & 0 & -1 & 0 & 0 & 1 \end{array} \Leftrightarrow$$

Now, start with the third row and eliminate all other elements in the third column. This can be achieved by the following operations. Writing the row used in the operations (the third one in this case) within the first set of parentheses for each operation, we have $2(3) + (4)$; $-2(3) + (2)$; $(3) + (1) \Rightarrow$

$$\begin{array}{cccc|cccc} 1 & 0 & 0 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 3 & -1 & -2 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 & -3 & 2 & 2 & 1 \end{array} \Leftrightarrow$$

Divide the 4th row by 2 and then perform the following operations: $\frac{1}{2}(4)$; $-1(4) + (3)$; $(4) + (2)$; $-1(4) + (1) \Rightarrow$

$$\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & -\frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{array} \Leftrightarrow$$

Thus,

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{bmatrix}.$$

This result should be verified to ensure that it is free of computational errors. Since

$$AA^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -1 & 0 & -\frac{1}{2} \\ \frac{3}{2} & 0 & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ -\frac{3}{2} & 1 & 1 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

the result is indeed correct.

Example 1.2.5. Evaluate A^{-1} if it exists where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix}.$$

Solution 1.2.5. If A^{-1} exists, then $AA^{-1} = I_3$. Write

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 0 & 2 \end{bmatrix} A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Starting with the first row, eliminate all other elements in the first column with the following operations: $-1(1) + (2)$; $-2(1) + (3) \Rightarrow$

$$\begin{array}{cccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -2 & 0 & \Leftrightarrow & -1 & 1 & 0 \\ 0 & -2 & 0 & & -2 & 0 & 1 \end{array}$$

The second and third rows on the left side being identical, the left-hand side matrix is singular, which means that A is singular. Thus, the inverse of A does not exist in this case.

1.3. Determinants of Partitioned Matrices

Consider a matrix A written in the following format:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ where } A_{11}, A_{12}, A_{21}, A_{22} \text{ are submatrices.}$$

For example,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A_{11} = [a_{11}], \quad A_{12} = [a_{12} \ a_{13}], \quad A_{21} = \begin{bmatrix} a_{21} \\ a_{31} \end{bmatrix}$$

and $A_{22} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$. The above is a 2×2 partitioning or a partitioning into two sub-matrices by two sub-matrices. But a 2×2 partitioning is not unique. We may also consider

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_{22} = [a_{33}], \quad A_{12} = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}, \quad A_{21} = [a_{31} \ a_{32}],$$

which is another 2×2 partitioning of A . We can also have a 1×2 or 2×1 partitioning into sub-matrices. We may observe one interesting property. Consider a block diagonal matrix. Let

$$A = \begin{bmatrix} A_{11} & O \\ O & A_{22} \end{bmatrix} \Rightarrow |A| = \begin{vmatrix} A_{11} & O \\ O & A_{22} \end{vmatrix},$$

where A_{11} is $r \times r$, A_{22} is $s \times s$, $r + s = n$ and O indicates a null matrix. Observe that when we evaluate the determinant, all the operations on the first r rows will produce the determinant of A_{11} as a coefficient, without affecting A_{22} , leaving an $r \times r$ identity matrix in the place of A_{11} . Similarly, all the operations on the last s rows will produce the determinant of A_{22} as a coefficient, leaving an $s \times s$ identity matrix in place of A_{22} . In other words, for a diagonal block matrix whose diagonal blocks are A_{11} and A_{22} ,

$$|A| = |A_{11}| \times |A_{22}|. \quad (1.3.1)$$

Given a triangular block matrix, be it lower or upper triangular, then its determinant is also the product of the determinants of the diagonal blocks. For example, consider

$$A = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix} \Rightarrow |A| = \begin{vmatrix} A_{11} & A_{12} \\ O & A_{22} \end{vmatrix}.$$

By using A_{22} , we can eliminate A_{12} without affecting A_{11} and hence, we can reduce the matrix of the determinant to a diagonal block form without affecting the value of the determinant. Accordingly, the determinant of an upper or lower triangular block matrix whose diagonal blocks are A_{11} and A_{22} , is

$$|A| = |A_{11}| |A_{22}|. \quad (1.3.2)$$

Partitioning is done to accommodate further operations such as matrix multiplication. Let A and B be two matrices whose product AB is defined. Suppose that we consider a

2×2 partitioning of A and B into sub-matrices; if the multiplication is performed treating the sub-matrices as if they were scalar quantities, the following format is obtained:

$$\begin{aligned} AB &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}. \end{aligned}$$

If all the products of sub-matrices on the right-hand side are defined, then we say that A and B are conformably partitioned for the product AB . Let A be a $n \times n$ matrix whose determinant is defined. Let us consider the 2×2 partitioning

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ where } A_{11} \text{ is } r \times r, A_{22} \text{ is } s \times s, r + s = n.$$

Then, A_{12} is $r \times s$ and A_{21} is $s \times r$. In this case, the first row block is $[A_{11} \ A_{12}]$ and the second row block is $[A_{21} \ A_{22}]$. When evaluating a determinant, we can add linear functions of rows to any other row or linear functions of rows to other blocks of rows without affecting the value of the determinant. What sort of a linear function of the first row block could be added to the second row block so that a null matrix O appears in the position of A_{21} ? It is $-A_{21}A_{11}^{-1}$ times the first row block. Then, we have

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & A_{12} \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{vmatrix}.$$

This is a triangular block matrix and hence its determinant is the product of the determinants of the diagonal blocks. That is,

$$|A| = |A_{11}| |A_{22} - A_{21}A_{11}^{-1}A_{12}|, |A_{11}| \neq 0.$$

From symmetry, it follows that

$$|A| = |A_{22}| |A_{11} - A_{12}A_{22}^{-1}A_{21}|, |A_{22}| \neq 0. \quad (1.3.3)$$

Let us now examine the inverses of partitioned matrices. Let A and A^{-1} be conformably partitioned for the product AA^{-1} . Consider a 2×2 partitioning of both A and A^{-1} . Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

where A_{11} and A^{11} are $r \times r$ and A_{22} and A^{22} are $s \times s$ with $r + s = n$, A is $n \times n$ and nonsingular. $AA^{-1} = I$ gives the following:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & I_s \end{bmatrix}.$$

That is,

$$A_{11}A^{11} + A_{12}A^{21} = I_r \quad (i)$$

$$A_{11}A^{12} + A_{12}A^{22} = O \quad (ii)$$

$$A_{21}A^{11} + A_{22}A^{21} = O \quad (iii)$$

$$A_{21}A^{12} + A_{22}A^{22} = I_s. \quad (iv)$$

From (ii), $A^{12} = -A_{11}^{-1}A_{12}A^{22}$. Substituting in (iv),

$$A_{21}[-A_{11}^{-1}A_{12}A^{22}] + A_{22}A^{22} = I_s \Rightarrow [A_{22} - A_{21}A_{11}^{-1}A_{12}]A^{22} = I_s.$$

That is,

$$A^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, |A_{11}| \neq 0, \quad (1.3.4)$$

and, from symmetry, it follows that

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, |A_{22}| \neq 0 \quad (1.3.5)$$

$$A_{11} = (A^{11} - A^{12}(A^{22})^{-1}A^{21})^{-1}, |A^{22}| \neq 0 \quad (1.3.6)$$

$$A_{22} = (A^{22} - A^{21}(A^{11})^{-1}A^{12})^{-1}, |A^{11}| \neq 0. \quad (1.3.7)$$

The rectangular components A_{12} , A_{21} , A^{12} , A^{21} can also be evaluated in terms of the sub-matrices by making use of Eqs. (i)–(iv).

1.4. Eigenvalues and Eigenvectors

Let A be $n \times n$ matrix, X be an $n \times 1$ vector, and λ be a scalar quantity. Consider the equation

$$AX = \lambda X \Rightarrow (A - \lambda I)X = O.$$

Observe that $X = O$ is always a solution. If this equation has a non-null vector X as a solution, then the determinant of the coefficient matrix must be zero because this matrix must be singular. If the matrix $(A - \lambda I)$ were nonsingular, its inverse $(A - \lambda I)^{-1}$ would exist and then, on pre-multiplying $(A - \lambda I)X = O$ by $(A - \lambda I)^{-1}$, we would have $X = O$, which is inadmissible since $X \neq O$. That is,

$$|A - \lambda I| = 0, \lambda \text{ being a scalar quantity.} \quad (1.4.1)$$

Since the matrix A is $n \times n$, equation (1.4.1) has n roots, which will be denoted by $\lambda_1, \dots, \lambda_n$. Then

$$|A - \lambda I| = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda), \quad AX_j = \lambda_j X_j.$$

Then, $\lambda_1, \dots, \lambda_n$ are called *the eigenvalues of A and $X_j \neq O$, an eigenvector corresponding to the eigenvalue λ_j .*

Example 1.4.1. Compute the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

Solution 1.4.1. Consider the equation

$$\begin{aligned} |A - \lambda I| = 0 &\Rightarrow \left| \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0 \Rightarrow \\ \begin{vmatrix} 1 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = 0 &\Rightarrow (1 - \lambda)(2 - \lambda) - 1 = 0 \Rightarrow \lambda^2 - 3\lambda + 1 = 0 \Rightarrow \\ \lambda = \frac{3 \pm \sqrt{9 - 4}}{2} &\Rightarrow \lambda_1 = \frac{3}{2} + \frac{\sqrt{5}}{2}, \quad \lambda_2 = \frac{3}{2} - \frac{\sqrt{5}}{2}. \end{aligned}$$

An eigenvector X_1 corresponding to $\lambda_1 = \frac{3}{2} + \frac{\sqrt{5}}{2}$ is given by $AX_1 = \lambda_1 X_1$ or $(A - \lambda_1 I)X_1 = O$. That is,

$$\begin{aligned} \begin{bmatrix} 1 - (\frac{3}{2} + \frac{\sqrt{5}}{2}) & 1 \\ 1 & 2 - (\frac{3}{2} + \frac{\sqrt{5}}{2}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \\ \left(-\frac{1}{2} - \frac{\sqrt{5}}{2}\right)x_1 + x_2 &= 0, & (i) \\ x_1 + \left(\frac{1}{2} - \frac{\sqrt{5}}{2}\right)x_2 &= 0. & (ii) \end{aligned}$$

Since $A - \lambda_1 I$ is singular, both (i) and (ii) must give the same solution. Letting $x_2 = 1$ in (ii), $x_1 = -\frac{1}{2} + \frac{\sqrt{5}}{2}$. Thus, one solution X_1 is

$$X_1 = \begin{bmatrix} -\frac{1}{2} + \frac{\sqrt{5}}{2} \\ 1 \end{bmatrix}.$$

Any nonzero constant multiple of X_1 is also a solution to $(A - \lambda_1 I)X_1 = O$. An eigenvector X_2 corresponding to the eigenvalue λ_2 is given by $(A - \lambda_2 I)X_2 = O$. That is,

$$\begin{aligned} \left(-\frac{1}{2} + \frac{\sqrt{5}}{2}\right)x_1 + x_2 &= 0, & (iii) \\ x_1 + \left(\frac{1}{2} + \frac{\sqrt{5}}{2}\right)x_2 &= 0. & (iv) \end{aligned}$$

Hence, one solution is

$$X_2 = \begin{bmatrix} -\frac{1}{2} - \frac{\sqrt{5}}{2} \\ 1 \end{bmatrix}.$$

Any nonzero constant multiple of X_2 is also an eigenvector corresponding to λ_2 .

Even if all elements of a matrix A are real, its eigenvalues can be real, positive, negative, zero, irrational or complex. Complex and irrational roots appear in pairs. If $a + ib$, $i = \sqrt{-1}$, and a, b real, is an eigenvalue, then $a - ib$ is also an eigenvalue of the same matrix. The following properties can be deduced from the definition:

- (1): *The eigenvalues of a diagonal matrix are its diagonal elements;*
- (2): *The eigenvalues of a triangular (lower or upper) matrix are its diagonal elements;*
- (3): *If any eigenvalue is zero, then the matrix is singular and its determinant is zero;*
- (4): *If λ is an eigenvalue of A and if A is nonsingular, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} ;*
- (5): *If λ is an eigenvalue of A , then λ^k is an eigenvalue of A^k , $k = 1, 2, \dots$, their associated eigenvector being the same;*
- (7): *The eigenvalues of an identity matrix are unities; however, the converse need not be true;*
- (8): *The eigenvalues of a scalar matrix with diagonal elements a, a, \dots, a are a repeated n times when the order of A is n ; however, the converse need not be true;*
- (9): *The eigenvalues of an orthonormal matrix, $AA' = I$, $A'A = I$, are ± 1 ; however, the converse need not be true;*
- (10): *The eigenvalues of an idempotent matrix, $A = A^2$, are ones and zeros; however, the converse need not be true. The only nonsingular idempotent matrix is the identity matrix;*
- (11): *At least one of the eigenvalues of a nilpotent matrix of order r , that is, $A \neq O, \dots, A^{r-1} \neq O, A^r = O$, is null;*
- (12): *For an $n \times n$ matrix, both A and A' have the same eigenvalues;*
- (13): *The eigenvalues of a symmetric matrix are real;*
- (14): *The eigenvalues of a skew symmetric matrix can only be zeros and purely imaginary numbers;*
- (15): *The determinant of A is the product of its eigenvalues: $|A| = \lambda_1 \lambda_2 \cdots \lambda_n$;*
- (16): *The trace of a square matrix is equal to the sum of its eigenvalues;*
- (17): *If $A = A'$ (symmetric), then the eigenvectors corresponding to distinct eigenvalues are orthogonal;*
- (18): *If $A = A'$ and A is $n \times n$, then there exists a full set of n eigenvectors which are linearly independent, even if some eigenvalues are repeated.*

Result (16) requires some explanation. We have already derived the following two results:

$$|A| = \sum_{i_1} \cdots \sum_{i_n} (-1)^{\rho} a_{1i_1} a_{2i_2} \cdots a_{ni_n} \quad (v)$$

and

$$|A - \lambda I| = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda). \quad (vi)$$

Equation (vi) yields a polynomial of degree n in λ where λ is a variable. When $\lambda = 0$, we have $|A| = \lambda_1 \lambda_2 \cdots \lambda_n$, the product of the eigenvalues. The term containing $(-1)^n \lambda^n$, when writing $|A - \lambda I|$ in the format of equation (v), can only come from the term $(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda)$ (refer to the explicit form for the 3×3 case discussed earlier). Two factors containing λ will be missing in the next term with the highest power of λ . Hence, $(-1)^n \lambda^n$ and $(-1)^{n-1} \lambda^{n-1}$ can only come from the term $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$, as can be seen from the expansion in the 3×3 case discussed in detail earlier. From (v) the coefficient of $(-1)^{n-1} \lambda^{n-1}$ is $a_{11} + a_{22} + \cdots + a_{nn} = \text{tr}(A)$ and from (vi), the coefficient of $(-1)^{n-1} \lambda^{n-1}$ is $\lambda_1 + \cdots + \lambda_n$. Hence $\text{tr}(A) = \lambda_1 + \cdots + \lambda_n = \text{sum of the eigenvalues of } A$. This does not mean that $\lambda_1 = a_{11}, \lambda_2 = a_{22}, \dots, \lambda_n = a_{nn}$, only that the sums will be equal.

Matrices in the Complex Domain When the elements in $A = (a_{ij})$ can also be complex quantities, then a typical element in A will be of the form $a + ib$, $i = \sqrt{-1}$, and a, b real. The complex conjugate of A will be denoted by \bar{A} and the conjugate transpose will be denoted by A^* . Then, for example,

$$A = \begin{bmatrix} 1+i & 2i & 3-i \\ 4i & 5 & 1+i \\ 2-i & i & 3+i \end{bmatrix} \Rightarrow \bar{A} = \begin{bmatrix} 1-i & -2i & 3+i \\ -4i & 5 & 1-i \\ 2+i & -i & 3-i \end{bmatrix}, \quad A^* = \begin{bmatrix} 1-i & -4i & 2+i \\ -2i & 5 & -i \\ 3+i & 1-i & 3-i \end{bmatrix}.$$

Thus, we can also write $A^* = (\bar{A})' = (\bar{A}')'$. When a matrix A is in the complex domain, we may write it as $A = A_1 + iA_2$ where A_1 and A_2 are real matrices. Then $\bar{A} = A_1 - iA_2$ and $A^* = A_1' - iA_2'$. In the above example,

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 5 & 1 \\ 2 & 0 & 3 \end{bmatrix} + i \begin{bmatrix} 1 & 2 & -1 \\ 4 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \Rightarrow$$

$$A_1 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 5 & 1 \\ 2 & 0 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 & -1 \\ 4 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix}.$$

A Hermitian Matrix If $A = A^*$, then A is called a Hermitian matrix. In the representation $A = A_1 + iA_2$, if $A = A^*$, then $A_1 = A_1'$ or A_1 is real symmetric, and $A_2 = -A_2'$ or A_2 is real skew symmetric. Note that when X is an $n \times 1$ vector, X^*X is real. Let

$$X = \begin{bmatrix} 2 \\ 3+i \\ 2i \end{bmatrix} \Rightarrow \bar{X} = \begin{bmatrix} 2 \\ 3-i \\ -2i \end{bmatrix}, \quad X^* = [2 \ 3-i \ -2i],$$

$$X^*X = [2 \ 3-i \ -2i] \begin{bmatrix} 2 \\ 3+i \\ 2i \end{bmatrix} = (2^2 + 0^2) + (3^2 + 1^2) + (0^2 + (-2)^2) = 18.$$

Consider the eigenvalues of a Hermitian matrix A , which are the solutions of $|A - \lambda I| = 0$. As in the real case, λ may be real, positive, negative, zero, irrational or complex. Then, for $X \neq O$,

$$AX = \lambda X \Rightarrow \quad (i)$$

$$X^*A^* = \bar{\lambda}X^* \quad (ii)$$

by taking the conjugate transpose. Since λ is scalar, its conjugate transpose is $\bar{\lambda}$. Pre-multiply (i) by X^* and post-multiply (ii) by X . Then for $X \neq O$, we have

$$X^*AX = \lambda X^*X \quad (iii)$$

$$X^*A^*X = \bar{\lambda}X^*X. \quad (iv)$$

When A is Hermitian, $A = A^*$, and so, the left-hand sides of (iii) and (iv) are the same. On subtracting (iv) from (iii), we have $0 = (\lambda - \bar{\lambda})X^*X$ where X^*X is real and positive, and hence $\lambda - \bar{\lambda} = 0$, which means that the imaginary part is zero or λ is real. If A is skew Hermitian, then we end up with $\lambda + \bar{\lambda} = 0 \Rightarrow \lambda$ is zero or purely imaginary. The above procedure also holds for matrices in the real domain. Thus, in addition to properties (13) and (14), we have the following properties:

(19) *The eigenvalues of a Hermitian matrix are real; however, the converse need not be true;*

(20) *The eigenvalues of a skew Hermitian matrix are zero or purely imaginary; however, the converse need not be true.*

1.5. Definiteness of Matrices, Quadratic and Hermitian Forms

Let X be an $n \times 1$ vector of real scalar variables x_1, \dots, x_n so that $X' = (x_1, \dots, x_n)$. Let $A = (a_{ij})$ be a real $n \times n$ matrix. Then, all the elements of the *quadratic form*, $u = X'AX$, are of degree 2. One can always write A as an equivalent symmetric matrix when A is the matrix in a quadratic form. Hence, without any loss of generality, we may assume $A = A'$ (symmetric) when A appears in a quadratic form $u = X'AX$.

Definiteness of a quadratic form and definiteness of a matrix are only defined for $A = A'$ (symmetric) in the real domain and for $A = A^*$ (Hermitian) in the complex domain. Hence, the basic starting condition is that either $A = A'$ or $A = A^*$. If for *all* non-null X , that is, $X \neq O$, $X'AX > 0$, $A = A'$, then A is said to be a *positive definite matrix* and $X'AX > 0$ is called a *positive definite quadratic form*. If for *all* non-null X , $X^*AX > 0$, $A = A^*$, then A is referred to as a *Hermitian positive definite matrix* and the corresponding Hermitian form $X^*AX > 0$, as *Hermitian positive definite*. Similarly, if for *all* non-null X , $X'AX \geq 0$, $X^*AX \geq 0$, then A is *positive semi-definite* or *Hermitian positive semi-definite*. If for *all* non-null X , $X'AX < 0$, $X^*AX < 0$, then A is *negative definite* and if $X'AX \leq 0$, $X^*AX \leq 0$, then A is *negative semi-definite*. The standard notations being utilized are as follows:

$A > O$ (A and $X'AX > 0$ are *real positive definite*; (O is a capital o and not zero)

$A \geq O$ (A and $X'AX \geq 0$ are *positive semi-definite*)

$A < O$ (A and $X'AX < 0$ are *negative definite*)

$A \leq O$ (A and $X'AX \leq 0$ are *negative semi-definite*).

All other matrices, which do not belong to any of those four categories, are called *indefinite matrices*. That is, for example, A is such that for some X , $X'AX > 0$, and for some other X , $X'AX < 0$, then A is an *indefinite matrix*. The corresponding Hermitian cases are:

$A > O$, $X^*AX > 0$ (Hermitian positive definite)

$A \geq O$, $X^*AX \geq 0$ (Hermitian positive semi-definite)

$A < O$, $X^*AX < 0$ (Hermitian negative definite)

$A \leq O$, $X^*AX \leq 0$ (Hermitian negative semi-definite). (1.5.1)

In all other cases, the matrix A and the Hermitian form X^*AX are *indefinite*. Certain conditions for the definiteness of $A = A'$ or $A = A^*$ are the following:

(1) *All the eigenvalues of A are positive* $\Leftrightarrow A > O$; *all eigenvalues are greater than or equal to zero* $\Leftrightarrow A \geq O$; *all eigenvalues are negative* $\Leftrightarrow A < O$; *all eigenvalues are ≤ 0* $\Leftrightarrow A \leq O$; *all other matrices $A = A'$ or $A = A^*$ for which some eigenvalues are positive and some others are negative are indefinite*.

(2) $A = A'$ or $A = A^*$ and *all the leading minors of A are positive* (leading minors are determinants of the leading sub-matrices, the r -th leading sub-matrix being obtained by deleting all rows and columns from the $(r + 1)$ -th onward), then $A > O$; *if the leading minors are ≥ 0* , then $A \geq O$; *if all the odd order minors are negative and all the even order minors are positive*, then $A < O$; *if the odd order minors are ≤ 0 and the even order*

minors are ≥ 0 , then $A \leq O$; all other matrices are indefinite. If $A \neq A'$ or $A \neq A^*$, then no definiteness can be defined in terms of the eigenvalues or leading minors. Let

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}.$$

Note that A is real symmetric as well as Hermitian. Since $X'AX = 2x_1^2 + 5x_2^2 > 0$ for all real x_1 and x_2 as long as both x_1 and x_2 are not both equal to zero, $A > O$ (positive definite). $X^*AX = 2|x_1|^2 + 5|x_2|^2 = 2[\sqrt{(x_{11}^2 + x_{12}^2)}]^2 + 5[\sqrt{(x_{21}^2 + x_{22}^2)}]^2 > 0$ for all $x_{11}, x_{12}, x_{21}, x_{22}$, as long as all are not simultaneously equal to zero, where $x_1 = x_{11} + ix_{12}$, $x_2 = x_{21} + ix_{22}$ with $x_{11}, x_{12}, x_{21}, x_{22}$ being real and $i = \sqrt{-1}$. Consider

$$B = \begin{bmatrix} -1 & 0 \\ 0 & -4 \end{bmatrix}, C = \begin{bmatrix} 2 & 0 \\ 0 & -6 \end{bmatrix}.$$

Then, $B < O$ and C is indefinite. Consider the following symmetric matrices:

$$A_1 = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} -2 & 1 \\ 1 & -5 \end{bmatrix}.$$

The leading minors of A_1 are $5 > 0$, $\begin{vmatrix} 5 & 2 \\ 2 & 4 \end{vmatrix} = 16 > 0$. The leading minors of A_2 are $2 > 0$, $\begin{vmatrix} 2 & 2 \\ 2 & 1 \end{vmatrix} = -2 < 0$. The leading minors of A_3 are $-2 < 0$, $\begin{vmatrix} -2 & 1 \\ 1 & -5 \end{vmatrix} = 9 > 0$. Hence $A_1 > O$, $A_3 < O$ and A_2 is indefinite. The following results will be useful when reducing a quadratic form or Hermitian form to its canonical form.

(3) For every real $A = A'$ (symmetric), there exists an orthonormal matrix Q , $QQ' = I$, $Q'Q = I$ such that $Q'AQ = \text{diag}(\lambda_1, \dots, \lambda_n)$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the $n \times n$ matrix A and $\text{diag}(\dots)$ denotes a diagonal matrix. In this case, a real quadratic form will reduce to the following linear combination:

$$X'AX = Y'Q'AQY = Y'\text{diag}(\lambda_1, \dots, \lambda_n)Y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2, Y = Q'X. \quad (1.5.2)$$

(4): For every Hermitian matrix $A = A^*$, there exists a unitary matrix U , $U^*U = I$, $UU^* = I$ such that

$$X^*AX = Y^*\text{diag}(\lambda_1, \dots, \lambda_n)Y = \lambda_1 |y_1|^2 + \dots + \lambda_n |y_n|^2, Y = U^*X. \quad (1.5.3)$$

When $A > O$ (real positive definite or Hermitian positive definite) then all the λ_j 's are real and positive. Then, $X'AX$ and X^*AX are strictly positive.

Let A and B be $n \times n$ matrices. If $AB = BA$, in which case we say that A and B commute, then both A and B can be simultaneously reduced to their canonical forms (diagonal forms with the diagonal elements being the eigenvalues) with the same orthonormal or unitary matrix P , $PP' = I$, $P'P = I$ if P is real and $PP^* = I$, $P^*P = I$ if complex, such that $P'AP = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $P'BP = \text{diag}(\mu_1, \dots, \mu_n)$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A and μ_1, \dots, μ_n are the eigenvalues of B . In the complex case, $P^*AP = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $P^*BP = \text{diag}(\mu_1, \dots, \mu_n)$. Observe that the eigenvalues of Hermitian matrices are real.

1.5.1. Singular value decomposition

For an $n \times n$ symmetric matrix $A = A'$, we have stated that there exists an $n \times n$ orthonormal matrix P , $PP' = I_n$, $P'P = I_n$ such that $P'AP = D = \text{diag}(\lambda_1, \dots, \lambda_n)$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A . If a square matrix A is not symmetric, there exists a nonsingular matrix Q such that $Q^{-1}AQ = D = \text{diag}(\lambda_1, \dots, \lambda_n)$ when the rank of A is n . If the rank is less than n , then we may be able to obtain a nonsingular Q such that the above representation holds; however, this is not always possible. If A is a $p \times q$ rectangular matrix for $p \neq q$ or if $p = q$ and $A \neq A'$, then can we find two orthonormal matrices U and V such that $A = U \begin{bmatrix} \Lambda & O \\ O & O \end{bmatrix} V'$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$, $UU' = I_p$, $U'U = I_p$, $VV' = I_q$, $V'V = I_q$ and k is the rank of A . This representation is equivalent to the following:

$$A = U \begin{bmatrix} \Lambda & O \\ O & O \end{bmatrix} V' = U_{(1)} \Lambda V'_{(1)} \quad (i)$$

where $U_{(1)} = [U_1, \dots, U_k]$, $V_{(1)} = [V_1, \dots, V_k]$, U_j being the normalized eigenvector of AA' corresponding to the eigenvalue λ_j^2 and V_j , the normalized eigenvector of $A'A$ corresponding to the eigenvalue λ_j^2 . The representation given in (i) is known as the *singular value decomposition of A* and $\lambda_1 > 0, \dots, \lambda_k > 0$ are called the *singular values of A* . Then, we have

$$AA' = U \begin{bmatrix} \Lambda^2 & O \\ O & O \end{bmatrix} U' = U_{(1)} \Lambda^2 U'_{(1)}, \quad A'A = V \begin{bmatrix} \Lambda^2 & O \\ O & O \end{bmatrix} V' = V_{(1)} \Lambda^2 V'_{(1)}. \quad (ii)$$

Thus, the procedure is the following: If $p \leq q$, compute the nonzero eigenvalues of AA' , otherwise compute the nonzero eigenvalues of $A'A$. Denote them by $\lambda_1^2, \dots, \lambda_k^2$ where k is the rank of A . Construct the following normalized eigenvectors U_1, \dots, U_k from AA' . This

gives $U_{(1)} = [U_1, \dots, U_k]$. Then, by using the same eigenvalues λ_j^2 , $j = 1, \dots, k$, determine the normalized eigenvectors, V_1, \dots, V_k , from $A'A$, and let $V_{(1)} = [V_1, \dots, V_k]$. Let us verify the above statements with the help of an example. Let

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Then,

$$AA' = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad A'A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

The eigenvalues of AA' are $\lambda_1^2 = 3$ and $\lambda_2^2 = 2$. The corresponding normalized eigenvectors of AA' are U_1 and U_2 , where

$$U_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad U_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \text{so that } U_{(1)} = [U_1, U_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now, by using $\lambda_1^2 = 3$ and $\lambda_2^2 = 2$, compute the normalized eigenvectors from $A'A$. They are:

$$V_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad V_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \text{so that } V'_{(1)} = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Then $\Lambda = \text{diag}(\sqrt{3}, \sqrt{2})$. Also,

$$U_{(1)}\Lambda V'_{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \end{bmatrix} = A.$$

This establishes the result. Observe that

$$\begin{aligned} AA' &= [U_{(1)}\Lambda V'_{(1)}][U_{(1)}\Lambda V'_{(1)}]' = U_{(1)}\Lambda^2 U'_{(1)} \\ A'A &= [U_{(1)}\Lambda V'_{(1)}]'[U_{(1)}\Lambda V'_{(1)}] = V_{(1)}\Lambda^2 V'_{(1)} \\ \Lambda^2 &= \text{diag}(\lambda_1^2, \lambda_2^2). \end{aligned}$$

1.6. Wedge Product of Differentials and Jacobians

If $y = f(x)$ is an explicit function of x , where x and y are real scalar variables, then we refer to x as the independent variable and to y as the dependent variable. In the present context, “independent” means that values for x are preassigned and the corresponding values of y are evaluated from the formula $y = f(x)$. The standard notations for small increment in x and the corresponding increment in y are Δx and Δy , respectively. By

convention, $\Delta x > 0$ and Δy can be positive, negative or zero depending upon the function f . If Δx goes to zero, then the limit is zero. However, if Δx goes to zero in the presence of the ratio $\frac{\Delta y}{\Delta x}$, then we have a different situation. Consider the identity

$$\Delta y \equiv \left(\frac{\Delta y}{\Delta x} \right) \Delta x \Rightarrow dy = A dx, \quad A = \frac{dy}{dx}. \quad (1.6.1)$$

This identity can always be written due to our convention $\Delta x > 0$. Consider $\Delta x \rightarrow 0$. If $\frac{\Delta y}{\Delta x}$ attains a limit at some stage as $\Delta x \rightarrow 0$, let us denote it by $A = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$, then the value of Δx at that stage is the differential of x , namely dx , and the corresponding Δy is dy and A is the ratio of the differentials $A = \frac{dy}{dx}$. If x_1, \dots, x_k are independent variables and if $y = f(x_1, \dots, x_k)$, then by convention $\Delta x_1 > 0, \dots, \Delta x_k > 0$. Thus, in light of (1.6.1), we have

$$dy = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_k} dx_k \quad (1.6.2)$$

where $\frac{\partial f}{\partial x_j}$ is the partial derivative of f with respect to x_j or the derivative of f with respect to x_j , keeping all other variables fixed.

Wedge Product of Differentials Let dx and dy be differentials of the real scalar variables x and y . Then the wedge product or skew symmetric product of dx and dy is denoted by $dx \wedge dy$ and is defined as

$$dx \wedge dy = -dy \wedge dx \Rightarrow dx \wedge dx = 0 \text{ and } dy \wedge dy = 0. \quad (1.6.3)$$

This definition indicates that higher order wedge products involving the same differential are equal to zero. Letting

$$y_1 = f_1(x_1, x_2) \text{ and } y_2 = f_2(x_1, x_2),$$

it follows from the basic definitions that

$$dy_1 = \frac{\partial f_1}{\partial x_1} dx_1 + \frac{\partial f_1}{\partial x_2} dx_2 \text{ and } dy_2 = \frac{\partial f_2}{\partial x_1} dx_1 + \frac{\partial f_2}{\partial x_2} dx_2.$$

By taking the wedge product and using the properties specified in (1.6.3), that is, $dx_1 \wedge dx_1 = 0$, $dx_2 \wedge dx_2 = 0$, $dx_2 \wedge dx_1 = -dx_1 \wedge dx_2$, we have

$$\begin{aligned} dy_1 \wedge dy_2 &= \left[\frac{\partial f_1}{\partial x_1} \frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2} \frac{\partial f_2}{\partial x_1} \right] dx_1 \wedge dx_2 \\ &= \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{vmatrix} dx_1 \wedge dx_2. \end{aligned}$$

In the general case we have the following corresponding result:

$$dy_1 \wedge \dots \wedge dy_k = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \dots & \frac{\partial f_k}{\partial x_k} \end{vmatrix} dx_1 \wedge \dots \wedge dx_k \quad (1.6.4)$$

$$dY = J dX \Rightarrow dX = \frac{1}{J} dY \text{ if } J \neq 0$$

where $dX = dx_1 \wedge \dots \wedge dx_k$, $dY = dy_1 \wedge \dots \wedge dy_k$ and $J = |(\frac{\partial f_i}{\partial x_j})|$ = the determinant of the matrix of partial derivatives where the (i, j) -th element is the partial derivative of f_i with respect to x_j . In dX and dY , the individual real scalar variables can be taken in any order to start with. However, for each interchange of variables, the result is to be multiplied by -1 .

Example 1.6.1. Consider the transformation $x_1 = r \cos^2 \theta$, $x_2 = r \sin^2 \theta$, $0 \leq r < \infty$, $0 \leq \theta \leq \frac{\pi}{2}$, $x_1 \geq 0$, $x_2 \geq 0$. Determine the relationship between $dx_1 \wedge dx_2$ and $dr \wedge d\theta$.

Solution 1.6.1. Taking partial derivatives, we have

$$\frac{\partial x_1}{\partial r} = \cos^2 \theta, \quad \frac{\partial x_1}{\partial \theta} = -2r \cos \theta \sin \theta,$$

$$\frac{\partial x_2}{\partial r} = \sin^2 \theta, \quad \frac{\partial x_2}{\partial \theta} = 2r \cos \theta \sin \theta.$$

Then, the determinant of the matrix of partial derivatives is given by

$$\begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos^2 \theta & -2r \cos \theta \sin \theta \\ \sin^2 \theta & 2r \cos \theta \sin \theta \end{vmatrix} = 2r \cos \theta \sin \theta$$

since $\cos^2 \theta + \sin^2 \theta = 1$. Hence,

$$dx_1 \wedge dx_2 = 2r \cos \theta \sin \theta dr \wedge d\theta, \quad J = 2r \cos \theta \sin \theta.$$

We may also establish this result by direct evaluation.

$$dx_1 = \frac{\partial x_1}{\partial r} dr + \frac{\partial x_1}{\partial \theta} d\theta = \cos^2 \theta dr - 2r \cos \theta \sin \theta d\theta,$$

$$dx_2 = \frac{\partial x_2}{\partial r} dr + \frac{\partial x_2}{\partial \theta} d\theta = \sin^2 \theta dr + 2r \cos \theta \sin \theta d\theta,$$

$$\begin{aligned}
dx_1 \wedge dx_2 &= \cos^2 \theta \sin^2 \theta dr \wedge dr + \cos^2 \theta (2r \cos \theta \sin \theta) dr \wedge d\theta \\
&\quad - \sin^2 \theta (2r \cos \theta \sin \theta) d\theta \wedge dr - (2r \cos \theta \sin \theta)^2 d\theta \wedge d\theta \\
&= 2r \cos \theta \sin \theta [\cos^2 \theta dr \wedge d\theta - \sin^2 \theta d\theta \wedge dr], \quad [dr \wedge dr = 0, d\theta \wedge d\theta = 0] \\
&= 2r \cos \theta \sin \theta [\cos^2 \theta + \sin^2 \theta] dr \wedge d\theta, \quad [d\theta \wedge dr = -dr \wedge d\theta] \\
&= 2r \cos \theta \sin \theta dr \wedge d\theta.
\end{aligned}$$

Linear Transformation Consider the linear transformation $Y = AX$ where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}.$$

Then, $\frac{\partial y_i}{\partial x_j} = a_{ij} \Rightarrow \left(\frac{\partial y_i}{\partial x_j}\right) = (a_{ij}) = A$. Then $dY = |A| dX$ or $J = |A|$. Hence, the following result:

Theorem 1.6.1. *Let X and Y be $p \times 1$ vectors of distinct real variables and $A = (a_{ij})$ be a constant nonsingular matrix. Then, the transformation $Y = AX$ is one to one and*

$$Y = AX, \quad |A| \neq 0 \Rightarrow dY = |A| dX. \quad (1.6.5)$$

Let us consider the complex case. Let $\tilde{X} = X_1 + iX_2$ where a tilde indicates that the matrix is in the complex domain, X_1 and X_2 are real $p \times 1$ vectors if \tilde{X} is $p \times 1$, and $i = \sqrt{-1}$. Then, the wedge product $d\tilde{X}$ is defined as $d\tilde{X} = dX_1 \wedge dX_2$. This is the general definition in the complex case whatever be the order of the matrix. If \tilde{Z} is $m \times n$ and if $\tilde{Z} = Z_1 + iZ_2$ where Z_1 and Z_2 are $m \times n$ and real, then $d\tilde{Z} = dZ_1 \wedge dZ_2$. Letting the constant $p \times p$ matrix $A = A_1 + iA_2$ where A_1 and A_2 are real and $p \times p$, and letting $\tilde{Y} = Y_1 + iY_2$ be $p \times 1$ where Y_1 and Y_2 are real and $p \times 1$, we have

$$\begin{aligned}
\tilde{Y} &= A\tilde{X} \Rightarrow Y_1 + iY_2 = [A_1 + iA_2][X_1 + iX_2] \\
&= [A_1X_1 - A_2X_2] + i[A_1X_2 + A_2X_1] \Rightarrow \\
Y_1 &= A_1X_1 - A_2X_2, \quad Y_2 = A_1X_2 + A_2X_1 \Rightarrow \\
\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} &= \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}. \quad (i)
\end{aligned}$$

Now, applying Result 1.6.1 on (i), it follows that

$$dY_1 \wedge dY_2 = \det \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} dX_1 \wedge dX_2. \quad (ii)$$

That is,

$$d\tilde{Y} = \det \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} d\tilde{X} \Rightarrow d\tilde{Y} = J d\tilde{X} \quad (iii)$$

where the Jacobian can be shown to be the absolute value of the determinant of A . If the determinant of A is denoted by $\det(A)$ and its absolute value, by $|\det(A)|$, and if $\det(A) = a + ib$ with a, b real and $i = \sqrt{-1}$ then the absolute value of the determinant is $+\sqrt{(a+ib)(a-ib)} = +\sqrt{(a^2+b^2)} = +\sqrt{[\det(A)][\det(A^*)]} = +\sqrt{[\det(AA^*)]}$. It can be easily seen that the above Jacobian is given by

$$\begin{aligned} J &= \det \begin{bmatrix} A_1 & -A_2 \\ A_2 & A_1 \end{bmatrix} = \det \begin{bmatrix} A_1 & -iA_2 \\ -iA_2 & A_1 \end{bmatrix} \\ &\quad (\text{multiplying the second row block by } -i \text{ and second column block by } i) \\ &= \det \begin{bmatrix} A_1 - iA_2 & A_1 - iA_2 \\ -iA_2 & A_1 \end{bmatrix} \quad (\text{adding the second row block to the first row block}) \\ &= \det(A_1 - iA_2) \det \begin{bmatrix} I & I \\ -iA_2 & A_1 \end{bmatrix} = \det(A_1 - iA_2) \det(A_1 + iA_2) \\ &\quad (\text{adding } (-1) \text{ times the first } p \text{ columns to the last } p \text{ columns}) \\ &= [\det(A)][\det(A^*)] = [\det(AA^*)] = |\det(A)|^2. \end{aligned}$$

Then, we have the following companion result of Theorem 1.6.1.

Theorem 1.6a.1. *Let \tilde{X} and \tilde{Y} be $p \times 1$ vectors in the complex domain, and let A be a $p \times p$ nonsingular constant matrix that may or may not be in the complex domain. If C is a constant $p \times 1$ vector, then*

$$\tilde{Y} = A\tilde{X} + C, \det(A) \neq 0 \Rightarrow d\tilde{Y} = |\det(A)|^2 d\tilde{X} = |\det(AA^*)| d\tilde{X}. \quad (1.6a.1)$$

For the results that follow, the complex case can be handled in a similar way and hence, only the final results will be stated. For details, the reader may refer to Mathai (1997). A more general result is the following:

Theorem 1.6.2. *Let X and Y be real $m \times n$ matrices with distinct real variables as elements. Let A be a $m \times m$ nonsingular constant matrix and C be a $m \times n$ constant matrix. Then*

$$Y = AX + C, \det(A) \neq 0 \Rightarrow dY = |A|^n dX. \quad (1.6.6)$$

The companion result is stated in the next theorem.

Theorem 1.6a.2. *Let \tilde{X} and \tilde{Y} be $m \times n$ matrices in the complex domain. Let A be a constant $m \times m$ nonsingular matrix that may or may not be in the complex domain, and C be a $m \times n$ constant matrix. Then*

$$\tilde{Y} = A\tilde{X} + C, \det(A) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^n d\tilde{X}. \quad (1.6a.2)$$

For proving the Theorems 1.6.2 and 1.6a.2, consider the columns of Y and X . Then apply Theorems 1.6.1 and 1.6a.1 to establish the results. If $X, \tilde{X}, Y, \tilde{Y}$ are as defined in Theorems 1.6.2 and 1.6a.2 and if B is a $n \times n$ nonsingular constant matrix, then we have the following results:

Theorems 1.6.3 and 1.6a.3. *Let $X, \tilde{X}, Y, \tilde{Y}$ and C be $m \times n$ matrices with distinct elements as previously defined, C be a constant matrix and B be a $n \times n$ nonsingular constant matrix. Then*

$$Y = XB + C, \det(B) \neq 0 \Rightarrow dY = |B|^m dX \quad (1.6.7)$$

and

$$\tilde{Y} = \tilde{X}B + C, \det(B) \neq 0 \Rightarrow d\tilde{Y} = |\det(BB^*)|^m d\tilde{X}. \quad (1.6a.3)$$

For proving these results, consider the rows of $Y, \tilde{Y}, X, \tilde{X}$ and then apply Theorems 1.6.1, 1.6a.1 to establish the results. Combining Theorems 1.6.2 and 1.6.3, as well as Theorems 1.6a.2 and 1.6a.3, we have the following results:

Theorems 1.6.4 and 1.6a.4. *Let $X, \tilde{X}, Y, \tilde{Y}$ be $m \times n$ matrices as previously defined, and let A be $m \times m$ and B be $n \times n$ nonsingular constant matrices. Then*

$$Y = AXB, \det(A) \neq 0, \det(B) \neq 0 \Rightarrow dY = |A|^n |B|^m dX \quad (1.6.8)$$

and

$$\tilde{Y} = A\tilde{X}B, \det(A) \neq 0, \det(B) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^n |\det(BB^*)|^m d\tilde{X}. \quad (1.6a.4)$$

We now consider the case of linear transformations involving symmetric and Hermitian matrices.

Theorems 1.6.5 and 1.6a.5. *Let $X = X', Y = Y'$ be real symmetric $p \times p$ matrices and let $\tilde{X} = \tilde{X}^*, \tilde{Y} = \tilde{Y}^*$ be $p \times p$ Hermitian matrices. If A is a $p \times p$ nonsingular constant matrix, then*

$$Y = AXA', Y = Y', X = X', \det(A) \neq 0 \Rightarrow dY = |A|^{p+1} dX \quad (1.6.9)$$

and

$$\tilde{Y} = A\tilde{X}A^*, \det(A) \neq 0 \Rightarrow d\tilde{Y} = |\det(AA^*)|^p d\tilde{X} \quad (1.6a.5)$$

for $\tilde{X} = \tilde{X}^*$ or $\tilde{X} = -\tilde{X}^*$.

The proof involves some properties of elementary matrices and elementary transformations. Elementary matrices were introduced in Sect. 1.2.1. There are two types of basic elementary matrices, the E and F types where the E type is obtained by multiplying any row (column) of an identity matrix by a nonzero scalar and the F type is obtained by adding any row to any other row of an identity matrix. A combination of E and F type matrices results in a G type matrix where a constant multiple of one row of an identity matrix is added to any other row. The G type is not a basic elementary matrix. By performing successive pre-multiplication with E , F and G type matrices, one can reduce a nonsingular matrix to a product of the basic elementary matrices of the E and F types, observing that the E and F type elementary matrices are nonsingular. This result is needed to establish Theorems 1.6.5 and 1.6a.5. Let $A = E_1 E_2 F_1 \cdots E_r F_s$ for some E_1, \dots, E_r and F_1, \dots, F_s . Then

$$AXA' = E_1 E_2 F_1 \cdots E_r F_s X F_s' E_r' \cdots E_2' E_1'.$$

Let $Y_1 = F_s X F_s'$ in which case the connection between dX and dY_1 can be determined from F_s . Now, letting $Y_2 = E_r Y_1 E_r'$, the connection between dY_2 and dY_1 can be similarly determined from E_r . Continuing in this manner, we finally obtain the connection between dY and dX , which will give the Jacobian as $|A|^{p+1}$ for the real case. In the complex case, the procedure is parallel.

We now consider two basic nonlinear transformations. In the first case, X is a $p \times p$ nonsingular matrix going to its inverse, that is, $Y = X^{-1}$.

Theorems 1.6.6 and 1.6a.6. *Let X and \tilde{X} be $p \times p$ real and complex nonsingular matrices, respectively. Let the regular inverses be denoted by $Y = X^{-1}$ and $\tilde{Y} = \tilde{X}^{-1}$, respectively. Then, ignoring the sign,*

$$Y = X^{-1} \Rightarrow dY = \begin{cases} |X|^{-2p} dX & \text{for a general } X \\ |X|^{-(p+1)} dX & \text{for } X = X' \\ |X|^{-(p-1)} dX & \text{for } X = -X' \end{cases} \quad (1.6.10)$$

and

$$\tilde{Y} = \tilde{X}^{-1} \Rightarrow d\tilde{Y} = \begin{cases} |\det(\tilde{X} \tilde{X}^*)|^{-2p} & \text{for a general } \tilde{X} \\ |\det(\tilde{X} \tilde{X}^*)|^{-p} & \text{for } \tilde{X} = \tilde{X}^* \text{ or } \tilde{X} = -\tilde{X}^*. \end{cases} \quad (1.6a.6)$$

The proof is based on the following observations: In the real case $XX^{-1} = I_p \Rightarrow (dX)X^{-1} + X(dX^{-1}) = O$ where (dX) represents the matrix of differentials in X . This means that

$$(dX^{-1}) = -X^{-1}(dX)X^{-1}.$$

The differentials are appearing only in the matrices of differentials. Hence this situation is equivalent to the general linear transformation considered in Theorems 1.6.4 and 1.6.5 where X and X^{-1} act as constants. The result is obtained upon taking the wedge product of differentials. The complex case is parallel.

The next results involve real positive definite matrices or Hermitian positive definite matrices that are expressible in terms of triangular matrices and the corresponding connection between the wedge product of differentials. Let X and \tilde{X} be complex $p \times p$ real positive definite and Hermitian positive definite matrices, respectively. Let $T = (t_{ij})$ be a real lower triangular matrix with $t_{ij} = 0$, $i < j$, $t_{jj} > 0$, $j = 1, \dots, p$, and the t_{ij} 's, $i \geq j$, be distinct real variables. Let $\tilde{T} = (\tilde{t}_{ij})$ be a lower triangular matrix with $\tilde{t}_{ij} = 0$, for $i < j$, the \tilde{t}_{ij} 's, $i > j$, be distinct complex variables, and \tilde{t}_{jj} , $j = 1, \dots, p$, be positive real variables. Then, the transformations $X = TT'$ in the real case and $\tilde{X} = \tilde{T}\tilde{T}'^*$ in the complex case can be shown to be one-to-one, which enables us to write dX in terms of dT and vice versa, uniquely, and $d\tilde{X}$ in terms of $d\tilde{T}$, uniquely. We first consider the real case. When $p = 2$,

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}, \quad x_{11} > 0, \quad x_{22} > 0, \quad x_{21} = x_{12}, \quad x_{11}x_{22} - x_{12}^2 > 0$$

due to positive definiteness of X , and

$$\begin{aligned} X = TT' &= \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{21}t_{11} \\ t_{21}t_{11} & t_{21}^2 + t_{22}^2 \end{bmatrix} \Rightarrow \\ \frac{\partial x_{11}}{\partial t_{11}} &= 2t_{11}, \quad \frac{\partial x_{11}}{\partial t_{21}} = 0, \quad \frac{\partial x_{11}}{\partial t_{22}} = 0 \\ \frac{\partial x_{22}}{\partial t_{11}} &= 0, \quad \frac{\partial x_{22}}{\partial t_{21}} = 2t_{21}, \quad \frac{\partial x_{22}}{\partial t_{22}} = 2t_{22} \\ \frac{\partial x_{12}}{\partial t_{11}} &= t_{21}, \quad \frac{\partial x_{12}}{\partial t_{21}} = t_{11}, \quad \frac{\partial x_{12}}{\partial t_{22}} = 0. \end{aligned}$$

Taking the x_{ij} 's in the order x_{11}, x_{12}, x_{22} and the t_{ij} 's in the order t_{11}, t_{21}, t_{22} , we form the following matrix of partial derivatives:

$$\begin{array}{cccc} & t_{11} & t_{12} & t_{22} \\ x_{11} & 2t_{11} & 0 & 0 \\ x_{21} & * & t_{11} & 0 \\ x_{22} & * & * & 2t_{22} \end{array}$$

where an asterisk indicates that an element may be present in that position; however, its value is irrelevant since the matrix is triangular and its determinant will simply be the

product of its diagonal elements. It can be observed from this pattern that for a general p , a diagonal element will be multiplied by 2 whenever x_{jj} is differentiated with respect to t_{jj} , $j = 1, \dots, p$. Then t_{11} will appear p times, t_{22} will appear $p - 1$ times, and so on, and t_{pp} will appear once along the diagonal. Hence the product of the diagonal elements will be $2^p t_{11}^p t_{22}^{p-1} \cdots t_{pp} = 2^p \{\prod_{j=1}^p t_{jj}^{p+1-j}\}$. A parallel procedure will yield the Jacobian in the complex case. Hence, the following results:

Theorems 1.6.7 and 1.6a.7. *Let X, \tilde{X}, T and \tilde{T} be $p \times p$ matrices where X is real positive definite, \tilde{X} is Hermitian positive definite, and T and \tilde{T} are lower triangular matrices whose diagonal elements are real and positive as described above. Then the transformations $X = TT'$ and $\tilde{X} = \tilde{T}\tilde{T}'^*$ are one-to-one, and*

$$dX = 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT \tag{1.6.11}$$

and

$$d\tilde{X} = 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T}. \tag{1.6a.7}$$

Given these introductory materials, we will explore multivariate statistical analysis from the perspective of Special Functions. As far as possible, the material in this chapter is self-contained. A few more Jacobians will be required when tackling transformations involving rectangular matrices or eigenvalue problems. These will be discussed in the respective chapters later on.

Example 1.6.2. Evaluate the following integrals: (1): $\int_X e^{-X'AX} dX$ where $A > O$ (real positive definite) is 3×3 and X is a 3×1 vector of distinct real scalar variables; (2): $\int_X e^{-\text{tr}(AXBX')}$ dX where X is a 2×3 matrix of distinct real scalar variables, $A > O$ (real positive definite), is 2×2 and $B > O$ (real positive definite) is 3×3 , A and B being constant matrices; (3): $\int_{X>O} e^{-\text{tr}(X)} dX$ where $X = X' > O$ is a 2×2 real positive definite matrix of distinct real scalar variables.

Solution 1.6.2.

(1) Let $X' = (x_1, x_2, x_3)$, $A > O$. Since $A > O$, we can uniquely define $A^{\frac{1}{2}} = (A^{\frac{1}{2}})'$. Then, write $X'AX = X'A^{\frac{1}{2}}A^{\frac{1}{2}}X = Y'Y$, $Y = A^{\frac{1}{2}}X$. It follows from Theorem 1.6.1 that $dX = |A|^{-\frac{1}{2}}dY$, and letting $Y' = (y_1, y_2, y_3)$, we have

$$\begin{aligned} \int_X e^{-X'AX} dX &= |A|^{-\frac{1}{2}} \int_Y e^{-(Y'Y)} dY \\ &= |A|^{-\frac{1}{2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y_1^2+y_2^2+y_3^2)} dy_1 \wedge dy_2 \wedge dy_3. \end{aligned}$$

Since

$$\int_{-\infty}^{\infty} e^{-y_j^2} dy_j = \sqrt{\pi}, \quad j = 1, 2, 3,$$

$$\int_X e^{-X'AX} dX = |A|^{-\frac{1}{2}} (\sqrt{\pi})^3.$$

(2) Since A is a 2×2 positive definite matrix, there exists a 2×2 matrix $A^{\frac{1}{2}}$ that is symmetric and positive definite. Similarly, there exists a 3×3 matrix $B^{\frac{1}{2}}$ that is symmetric and positive definite. Let $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}} \Rightarrow dY = |A|^{\frac{3}{2}}|B|^{\frac{2}{2}}dX$ or $dX = |A|^{-\frac{3}{2}}|B|^{-1}dY$ by Theorem 1.6.4. Moreover, given two matrices A_1 and A_2 , $\text{tr}(A_1A_2) = \text{tr}(A_2A_1)$ even if $A_1A_2 \neq A_2A_1$, as long as the products are defined. By making use of this property, we may write

$$\begin{aligned} \text{tr}(AXBX') &= \text{tr}(A^{\frac{1}{2}}A^{\frac{1}{2}}XB^{\frac{1}{2}}B^{\frac{1}{2}}X') = \text{tr}(A^{\frac{1}{2}}XB^{\frac{1}{2}}B^{\frac{1}{2}}X'A^{\frac{1}{2}}) \\ &= \text{tr}[(A^{\frac{1}{2}}XB^{\frac{1}{2}})(A^{\frac{1}{2}}XB^{\frac{1}{2}})'] = \text{tr}(YY') \end{aligned}$$

where $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$ and dY is given above. However, for any real matrix Y , whether square or rectangular, $\text{tr}(YY') = \text{tr}(Y'Y) =$ the sum of the squares of all the elements of Y . Thus, we have

$$\int_X e^{-\text{tr}(AXBX')} dX = |A|^{-\frac{3}{2}}|B|^{-1} \int_Y e^{-\text{tr}(YY')} dY.$$

Observe that since $\text{tr}(YY')$ is the sum of squares of 6 real scalar variables, the integral over Y reduces to a multiple integral involving six integrals where each variable is over the entire real line. Hence,

$$\int_Y e^{-\text{tr}(YY')} dY = \prod_{j=1}^6 \int_{-\infty}^{\infty} e^{-y_j^2} dy_j = \prod_{j=1}^6 (\sqrt{\pi}) = (\sqrt{\pi})^6.$$

Note that we have denoted the sum of the six y_{ij}^2 as $y_1^2 + \dots + y_6^2$ for convenience. Thus,

$$\int_X e^{-\text{tr}(AXBX')} dX = |A|^{-\frac{3}{2}}|B|^{-1}(\sqrt{\pi})^6.$$

(3) In this case, X is a 2×2 real positive definite matrix. Let $X = TT'$ where T is lower triangular with positive diagonal elements. Then,

$$T = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix}, \quad t_{11} > 0, \quad t_{22} > 0, \quad TT' = \begin{bmatrix} t_{11} & 0 \\ t_{21} & t_{22} \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} \\ 0 & t_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{11}t_{21} \\ t_{11}t_{21} & t_{21}^2 + t_{22}^2 \end{bmatrix},$$

and $\text{tr}(TT') = t_{11}^2 + (t_{21}^2 + t_{22}^2)$, $t_{11} > 0$, $t_{22} > 0$, $-\infty < t_{21} < \infty$. From Theorem 1.6.7, the Jacobian is

$$dX = 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT = 2^2 (t_{11}^2 t_{22}) dt_{11} \wedge dt_{21} \wedge dt_{22}.$$

Therefore

$$\begin{aligned} \int_{X>O} e^{-\text{tr}(X)} dX &= \int_T e^{-\text{tr}(TT')} [2^2 (t_{11}^2 t_{22})] dT \\ &= \left(\int_{-\infty}^{\infty} e^{-t_{21}^2} dt_{21} \right) \left(\int_0^{\infty} 2t_{11}^2 e^{-t_{11}^2} dt_{11} \right) \left(\int_0^{\infty} 2t_{22} e^{-t_{22}^2} dt_{22} \right) \\ &= [\sqrt{\pi}] [\Gamma(\frac{3}{2})] [\Gamma(1)] = \frac{\pi}{2}. \end{aligned}$$

Example 1.6.3. Let $A = A^* > O$ be a constant 2×2 Hermitian positive definite matrix. Let \tilde{X} be a 2×1 vector in the complex domain and $\tilde{X}_2 > O$ be a 2×2 Hermitian positive definite matrix. Then, evaluate the following integrals: (1): $\int_{\tilde{X}} e^{-(\tilde{X}^* A \tilde{X})} d\tilde{X}$; (2): $\int_{\tilde{X}_2 > O} e^{-\text{tr}(\tilde{X}_2)} d\tilde{X}_2$.

Solution 1.6.3.

(1): Since $A = A^* > O$, there exists a unique Hermitian positive definite square root $A^{\frac{1}{2}}$. Then,

$$\begin{aligned} \tilde{X}^* A \tilde{X} &= \tilde{X}^* A^{\frac{1}{2}} A^{\frac{1}{2}} \tilde{X} = \tilde{Y}^* \tilde{Y}, \\ \tilde{Y} &= A^{\frac{1}{2}} \tilde{X} \Rightarrow d\tilde{X} = |\det(A)|^{-1} d\tilde{Y} \end{aligned}$$

by Theorem 1.6a.1. But $\tilde{Y}^* \tilde{Y} = |\tilde{y}_1|^2 + |\tilde{y}_2|^2$ since $\tilde{Y}^* = (\tilde{y}_1^*, \tilde{y}_2^*)$. Since the \tilde{y}_j 's are scalar in this case, an asterisk means only the complex conjugate, the transpose being itself. However,

$$\int_{\tilde{y}_j} e^{-|\tilde{y}_j|^2} d\tilde{y}_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y_{j1}^2 + y_{j2}^2)} dy_{j1} \wedge dy_{j2} = (\sqrt{\pi})^2 = \pi$$

where $\tilde{y}_j = y_{j1} + iy_{j2}$, $i = \sqrt{-1}$, y_{j1} and y_{j2} being real. Hence

$$\begin{aligned} \int_{\tilde{X}} e^{-(\tilde{X}^* A \tilde{X})} d\tilde{X} &= |\det(A)|^{-1} \int_{\tilde{Y}} e^{-(|\tilde{y}_1|^2 + |\tilde{y}_2|^2)} d\tilde{y}_1 \wedge d\tilde{y}_2 \\ &= |\det(A)|^{-1} \left(\prod_{j=1}^2 \int_{\tilde{Y}_j} e^{-|\tilde{y}_j|^2} d\tilde{y}_j \right) = |\det(A)|^{-1} \prod_{j=1}^2 \pi \\ &= |\det(A)|^{-1} \pi^2. \end{aligned}$$

(2): Make the transformation $\tilde{X}_2 = \tilde{T}\tilde{T}^*$ where \tilde{T} is lower triangular with its diagonal elements being real and positive. That is,

$$\tilde{T} = \begin{bmatrix} t_{11} & 0 \\ \tilde{t}_{21} & t_{22} \end{bmatrix}, \quad \tilde{T}\tilde{T}^* = \begin{bmatrix} t_{11}^2 & t_{11}\tilde{t}_{21} \\ t_{11}\tilde{t}_{21} & |\tilde{t}_{21}|^2 + t_{22}^2 \end{bmatrix}$$

and the Jacobian is $d\tilde{X}_2 = 2^p \{\prod_{j=1}^p t_{jj}^{2(p-j)+1}\} d\tilde{T} = 2^2 t_{11}^3 t_{22} d\tilde{T}$ by Theorem 1.6a.7. Hence,

$$\int_{\tilde{X}_2 > O} e^{-\text{tr}(\tilde{X}_2)} d\tilde{X}_2 = \int_{\tilde{T}} 2^2 t_{11}^3 t_{22} e^{-(t_{11}^2 + t_{22}^2 + |\tilde{t}_{21}|^2)} dt_{11} \wedge dt_{22} \wedge d\tilde{t}_{21}.$$

But

$$\begin{aligned} 2 \int_{t_{11} > 0} t_{11}^3 e^{-t_{11}^2} dt_{11} &= \int_{u=0}^{\infty} u e^{-u} du = 1, \\ 2 \int_{t_{22} > 0} t_{22} e^{-t_{22}^2} dt_{22} &= \int_{v=0}^{\infty} e^{-v} dv = 1, \\ \int_{\tilde{t}_{21}} e^{-|\tilde{t}_{21}|^2} d\tilde{t}_{21} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(t_{211}^2 + t_{212}^2)} dt_{211} \wedge dt_{212} \\ &= \left(\int_{-\infty}^{\infty} e^{-t_{211}^2} dt_{211} \right) \left(\int_{-\infty}^{\infty} e^{-t_{212}^2} dt_{212} \right) \\ &= \sqrt{\pi} \sqrt{\pi} = \pi, \end{aligned}$$

where $\tilde{t}_{21} = t_{211} + it_{212}$, $i = \sqrt{-1}$ and t_{211} , t_{212} real. Thus

$$\int_{\tilde{X}_2 > O} e^{-\text{tr}(\tilde{X}_2)} d\tilde{X}_2 = \pi.$$

1.7. Differential Operators

Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad \frac{\partial}{\partial X} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_p} \end{bmatrix}, \quad \frac{\partial}{\partial X'} = \left[\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_p} \right]$$

where x_1, \dots, x_p are distinct real scalar variables, $\frac{\partial}{\partial X}$ is the partial differential operator and $\frac{\partial}{\partial X'}$ is the transpose operator. Then, $\frac{\partial}{\partial X} \frac{\partial}{\partial X'}$ is the configuration of all second order partial differential operators given by

$$\frac{\partial}{\partial X} \frac{\partial}{\partial X'} = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} & \frac{\partial^2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2}{\partial x_1 \partial x_p} \\ \frac{\partial^2}{\partial x_2 \partial x_1} & \frac{\partial^2}{\partial x_2^2} & \cdots & \frac{\partial^2}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_p \partial x_1} & \frac{\partial^2}{\partial x_p \partial x_2} & \cdots & \frac{\partial^2}{\partial x_p^2} \end{bmatrix}.$$

Let $f(X)$ be a real-valued scalar function of X . Then, this operator, operating on f will be defined as

$$\frac{\partial}{\partial X} f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{bmatrix}.$$

For example, if $f = x_1^2 + x_1 x_2 + x_2^3$, then $\frac{\partial f}{\partial x_1} = 2x_1 + x_2$, $\frac{\partial f}{\partial x_2} = x_1 + 3x_2^2$, and

$$\frac{\partial f}{\partial X} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + 3x_2^2 \end{bmatrix}.$$

Let $f = a_1 x_1 + a_2 x_2 + \cdots + a_p x_p = A'X = X'A$, $A' = (a_1, \dots, a_p)$, $X' = (x_1, \dots, x_p)$ where a_1, \dots, a_p are real constants and x_1, \dots, x_p are distinct real scalar variables. Then $\frac{\partial f}{\partial x_j} = a_j$ and we have the following result:

Theorem 1.7.1. *Let A, X and f be as defined above where $f = a_1 x_1 + \cdots + a_p x_p$ is a linear function of X , then*

$$\frac{\partial}{\partial X} f = A.$$

Letting $f = X'X = x_1^2 + \cdots + x_p^2$, $\frac{\partial f}{\partial x_j} = 2x_j$, and we have the following result.

Theorem 1.7.2. *Let X be a $p \times 1$ vector of real scalar variables so that $X'X = x_1^2 + \cdots + x_p^2$. Then*

$$\frac{\partial f}{\partial X} = 2X.$$

Now, let us consider a general quadratic form $f = X'AX$, $A = A'$, where X is a $p \times 1$ vector whose components are real scalar variables and A is a constant matrix. Then $\frac{\partial f}{\partial x_j} = (a_{j1}x_1 + \cdots + a_{jp}x_p) + (a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p)$ for $j = 1, \dots, p$. Hence we have the following result:

Theorem 1.7.3. *Let $f = X'AX$ be a real quadratic form where X is a $p \times 1$ real vector whose components are distinct real scalar variables and A is a constant matrix. Then*

$$\frac{\partial f}{\partial X} = \begin{cases} (A + A')X & \text{for a general } A \\ 2AX & \text{when } A = A' \end{cases} .$$

1.7.1. Some basic applications of the vector differential operator

Let X be a $p \times 1$ vector with real scalar elements x_1, \dots, x_p . Let $A = (a_{ij}) = A'$ be a constant matrix. Consider the problem of optimizing the real quadratic form $u = X'AX$. There is no unrestricted maximum or minimum. If $A = A' > O$ (positive definite), u can tend to $+\infty$ and similarly, if $A = A' < O$, u can go to $-\infty$. However, if we confine ourselves to the surface of a unit hypersphere or equivalently require that $X'X = 1$, then we can have a finite maximum and a finite minimum. Let $u_1 = X'AX - \lambda(X'X - 1)$ so that we have added zero to u and hence u_1 is the same as u , where λ is an arbitrary constant or a Lagrangian multiplier. Then, differentiating u_1 with respect to x_1, \dots, x_p , equating the resulting expressions to zero, and thereafter solving for critical points, is equivalent to solving the equation $\frac{\partial u_1}{\partial X} = O$ (null) and solving this single equation. That is,

$$\frac{\partial u_1}{\partial X} = O \Rightarrow 2AX - 2\lambda X = O \Rightarrow (A - \lambda I)X = O. \quad (i)$$

For (i) to have a non-null solution for X , the coefficient matrix $A - \lambda I$ has to be singular or its determinant must be zero. That is, $|A - \lambda I| = 0$ and $AX = \lambda X$ or λ is an eigenvalue of A and X is the corresponding eigenvector. But

$$AX = \lambda X \Rightarrow X'AX = \lambda X'X = \lambda \text{ since } X'X = 1. \quad (ii)$$

Hence the maximum value of $X'AX$ corresponds to the largest eigenvalue of A and the minimum value of $X'AX$, to the smallest eigenvalue of A . Observe that when $A = A'$ the eigenvalues are real. Hence we have the following result:

Theorem 1.7.4. *Let $u = X'AX$, $A = A'$, X be a $p \times 1$ vector of real scalar variables as its elements. Letting $X'X = 1$, then*

$$\begin{aligned} \max_{X'X=1} [X'AX] &= \lambda_1 = \text{the largest eigenvalue of } A \\ \min_{X'X=1} [X'AX] &= \lambda_p = \text{the smallest eigenvalue of } A. \end{aligned}$$

Principal Component Analysis where it is assumed that $A > O$ relies on this result. This will be elaborated upon in later chapters. Now, we consider the optimization of $u = X'AX$, $A = A'$ subject to the condition $X'BX = 1$, $B = B'$. Take λ as the Lagrangian multiplier and consider $u_1 = X'AX - \lambda(X'BX - 1)$. Then

$$\frac{\partial u_1}{\partial X} = O \Rightarrow AX = \lambda BX \Rightarrow |A - \lambda B| = 0. \quad (iii)$$

Note that $X'AX = \lambda X'BX = \lambda$ from (i). Hence, the maximum of $X'AX$ is the largest value of λ satisfying (i) and the minimum of $X'AX$ is the smallest value of λ satisfying (i). Note that when B is nonsingular, $|A - \lambda B| = 0 \Rightarrow |AB^{-1} - \lambda I| = 0$ or λ is an eigenvalue of AB^{-1} . Thus, this case can also be treated as an eigenvalue problem. Hence, the following result:

Theorem 1.7.5. *Let $u = X'AX$, $A = A'$ where the elements of X are distinct real scalar variables. Consider the problem of optimizing $X'AX$ subject to the condition $X'BX = 1$, $B = B'$, where A and B are constant matrices, then*

$$\begin{aligned} \max_{X'BX=1} [X'AX] &= \lambda_1 = \text{largest eigenvalue of } AB^{-1}, |B| \neq 0 \\ &= \text{the largest root of } |A - \lambda B| = 0; \\ \min_{X'BX=1} [X'AX] &= \lambda_p = \text{smallest eigenvalue of } AB^{-1}, |B| \neq 0 \\ &= \text{the smallest root of } |A - \lambda B| = 0. \end{aligned}$$

Now, consider the optimization of a real quadratic form subject to a linear constraint. Let $u = X'AX$, $A = A'$ be a quadratic form where X is $p \times 1$. Let $B'X = X'B = 1$ be a constraint where $B' = (b_1, \dots, b_p)$, $X' = (x_1, \dots, x_p)$ with b_1, \dots, b_p being real constants and x_1, \dots, x_p being real distinct scalar variables. Take 2λ as the Lagrangian multiplier and consider $u_1 = X'AX - 2\lambda(X'B - 1)$. The critical points are available from the following equation

$$\begin{aligned} \frac{\partial}{\partial X} u_1 = O &\Rightarrow 2AX - 2\lambda B = O \Rightarrow X = \lambda A^{-1}B, \text{ for } |A| \neq 0 \\ &\Rightarrow B'X = \lambda B'A^{-1}B \Rightarrow \lambda = \frac{1}{B'A^{-1}B}. \end{aligned}$$

In this problem, observe that the quadratic form is unbounded even under the restriction $B'X = 1$ and hence there is no maximum. The only critical point corresponds to a minimum. From $AX = \lambda B$, we have $X'AX = \lambda X'B = \lambda$. Hence the minimum value is $\lambda = [B'A^{-1}B]^{-1}$ where it is assumed that A is nonsingular. Thus following result:

Theorem 1.7.6. Let $u = X'AX$, $A = A'$, $|A| \neq 0$. Let $B'X = 1$ where $B' = (b_1, \dots, b_p)$ be a constant vector and X is $p \times 1$ vector of real distinct scalar variables. Then, the minimum of the quadratic form u , under the restriction $B'X = 1$ where B is a constant vector, is given by

$$\min_{B'X=1} [X'AX] = \frac{1}{B'A^{-1}B}.$$

Such problems arise for instance in regression analysis and model building situations. We could have eliminated one of the variables with the linear constraint; however, the optimization would still involve all other variables, and thus not much simplification would be achieved by eliminating one variable. Hence, operating with the vector differential operator is the most convenient procedure in this case.

We will now consider the mathematical part of a general problem in prediction analysis where some variables are predicted from another set of variables. This topic is related to *Canonical Correlation Analysis*. We will consider the optimization part of the problem in this section. The problem consists in optimizing a bilinear form subject to quadratic constraints. Let X be a $p \times 1$ vector of real scalar variables x_1, \dots, x_p , and Y be a $q \times 1$ vector of real scalar variables y_1, \dots, y_q , where q need not be equal to p . Consider the bilinear form $u = X'AY$ where A is a $p \times q$ rectangular constant matrix. We would like to optimize this bilinear form subject to the quadratic constraints $X'BX = 1$, $Y'CY = 1$, $B = B'$ and $C = C'$ where B and C are constant matrices. In Canonical Correlation Analysis, B and C are constant real positive definite matrices. Take λ_1 and λ_2 as Lagrangian multipliers and let $u_1 = X'AY - \frac{\lambda_1}{2}(X'BX - 1) - \frac{\lambda_2}{2}(Y'CY - 1)$. Then

$$\begin{aligned} \frac{\partial}{\partial X} u_1 = 0 &\Rightarrow AY - \lambda_1 BX = 0 \Rightarrow AY = \lambda_1 BX \\ &\Rightarrow X'AY = \lambda_1 X'BX = \lambda_1; \end{aligned} \quad (i)$$

$$\begin{aligned} \frac{\partial}{\partial Y} u_1 = 0 &\Rightarrow A'X - \lambda_2 CY = 0 \Rightarrow A'X = \lambda_2 CY \\ &\Rightarrow Y'A'X = \lambda_2 Y'CY = \lambda_2. \end{aligned} \quad (ii)$$

It follows from (i) and (ii) that $\lambda_1 = \lambda_2 = \lambda$, say. Observe that $X'AY$ is 1×1 so that $X'AY = Y'A'X$. After substituting λ to λ_1 and λ_2 , we can combine equations (i) and (ii) in a single matrix equation as follows:

$$\begin{aligned} \begin{bmatrix} -\lambda B & A \\ A' & -\lambda C \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} &= 0 \Rightarrow \\ \begin{vmatrix} -\lambda B & A \\ A' & -\lambda C \end{vmatrix} &= 0. \end{aligned} \quad (iii)$$

Opening up the determinant by making use of a result on partitioned matrices from Sect. 1.3, we have

$$\begin{aligned} |-\lambda B| |-\lambda C - A'(-\lambda B)^{-1}A| = 0, |B| \neq 0 \Rightarrow \\ |A'B^{-1}A - \lambda^2 C| = 0. \end{aligned} \quad (iv)$$

Then $\nu = \lambda^2$ is a root obtained from Eq. (iv). We can also obtain a parallel result by opening up the determinant in (iii) as

$$|-\lambda C| |-\lambda B - A(-\lambda C)^{-1}A'| = 0 \Rightarrow |AC^{-1}A' - \lambda^2 B| = 0, |C| \neq 0. \quad (v)$$

Hence we have the following result.

Theorem 1.7.7. *Let X and Y be respectively $p \times 1$ and $q \times 1$ real vectors whose components are distinct scalar variables. Consider the bilinear form $X'AY$ and the quadratic forms $X'BX$ and $Y'CY$ where $B = B'$, $C = C'$, and B and C are nonsingular constant matrices. Then,*

$$\begin{aligned} \max_{X'BX=1, Y'CY=1} [X'AY] &= |\lambda_1| \\ \min_{X'BX=1, Y'CY=1} [X'AY] &= |\lambda_p| \end{aligned}$$

where λ_1^2 is the largest root resulting from equation (iv) or (v) and λ_p^2 is the smallest root resulting from equation (iv) or (v).

Observe that if $p < q$, we may utilize equation (v) to solve for λ^2 and if $q < p$, then we may use equation (iv) to solve for λ^2 , and both will lead to the same solution. In the above derivation, we assumed that B and C are nonsingular. In Canonical Correlation Analysis, both B and C are real positive definite matrices corresponding to the variances $X'BX$ and $Y'CY$ of the linear forms and then, $X'AY$ corresponds to covariance between these linear forms.

Note 1.7.1. We have confined ourselves to results in the real domain in this subsection since only real cases are discussed in connection with the applications that are considered in later chapters, such as Principal Component Analysis and Canonical Correlation Analysis. The corresponding complex cases do not appear to have practical applications.

Accordingly, optimizations of Hermitian forms will not be discussed. However, parallel results to Theorems 1.7.1–1.7.7 could similarly be worked out in the complex domain.

References

A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A.M. Mathai and Hans J. Haubold (2017a): *Linear Algebra: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai and Hans J. Haubold (2017b): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 2

The Univariate Gaussian and Related Distributions



2.1. Introduction

It is assumed that the reader has had adequate exposure to basic concepts in Probability, Statistics, Calculus and Linear Algebra. This chapter will serve as a review of the basic ideas about the univariate Gaussian, or normal, distribution as well as related distributions. We will begin with a discussion of the univariate Gaussian density. We will adopt the following notation: real scalar mathematical or random variables will be denoted by lower-case letters such as x, y, z , whereas vector or matrix-variate mathematical or random variables will be denoted by capital letters such as X, Y, Z, \dots . Statisticians usually employ the double notation X and x where it is claimed that x is a realization of X . Since x can vary, it is a variable in the mathematical sense. Treating mathematical and random variables the same way will simplify the notation and possibly reduce the confusion. Complex variables will be written with a tilde such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$, etc. Constant scalars and matrices will be written without a tilde unless for stressing that the constant matrix is in the complex domain. In such a case, a tilde will be also be utilized for the constant. Constant matrices will be denoted by A, B, C, \dots .

The numbering will first indicate the chapter and then the section. For example, Eq. (2.1.9) will be the ninth equation in Sect. 2.1 of this chapter. Local numbering for sub-sections will be indicated as (i), (ii), and so on.

Let x_1 be a real univariate Gaussian, or normal, random variable whose parameters are μ_1 and σ_1^2 ; this will be written as $x_1 \sim N_1(\mu_1, \sigma_1^2)$, the associated density being given by

$$f(x_1) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2}, \quad -\infty < x_1 < \infty, \quad -\infty < \mu_1 < \infty, \quad \sigma_1 > 0.$$

In this instance, the subscript 1 in $N_1(\cdot)$ refers to the univariate case. Incidentally, a density is a real-valued scalar function of x such that $f(x) \geq 0$ for all x and $\int_x f(x) dx = 1$. The moment generating function (mgf) of this Gaussian random variable x_1 , with t_1 as

its parameter, is given by the following expected value, where $E[\cdot]$ denotes the expected value of $[\cdot]$:

$$E[e^{t_1 x_1}] = \int_{-\infty}^{\infty} e^{t_1 x_1} f(x_1) dx_1 = e^{t_1 \mu_1 + \frac{1}{2} t_1^2 \sigma_1^2}. \quad (2.1.1)$$

2.1a. The Complex Scalar Gaussian Variable

Let $\tilde{x} = x_1 + ix_2$, $i = \sqrt{-1}$, x_1, x_2 real scalar variables. Let $E[x_1] = \mu_1$, $E[x_2] = \mu_2$, $\text{Var}(x_1) = \sigma_1^2$, $\text{Var}(x_2) = \sigma_2^2$, $\text{Cov}(x_1, x_2) = \sigma_{12}$. By definition, the variance of the complex random variable \tilde{x} is defined as

$$\text{Var}(\tilde{x}) = E[\tilde{x} - E(\tilde{x})][\tilde{x} - E(\tilde{x})]^*$$

where $*$ indicates a conjugate transpose in general; in this case, it simply means the conjugate since \tilde{x} is a scalar. Since $\tilde{x} - E(\tilde{x}) = x_1 + ix_2 - \mu_1 - i\mu_2 = (x_1 - \mu_1) + i(x_2 - \mu_2)$ and $[\tilde{x} - E(\tilde{x})]^* = (x_1 - \mu_1) - i(x_2 - \mu_2)$,

$$\begin{aligned} \text{Var}(\tilde{x}) &= E[\tilde{x} - E(\tilde{x})][\tilde{x} - E(\tilde{x})]^* \\ &= E[(x_1 - \mu_1) + i(x_2 - \mu_2)][(x_1 - \mu_1) - i(x_2 - \mu_2)] = E[(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2] \\ &= \sigma_1^2 + \sigma_2^2 \equiv \sigma^2. \end{aligned} \quad (i)$$

Observe that $\text{Cov}(x_1, x_2)$ does not appear in the scalar case. However, the covariance will be present in the vector/matrix case as will be explained in the coming chapters. The complex Gaussian density is given by

$$f(\tilde{x}) = \frac{1}{\pi \sigma^2} e^{-\frac{1}{\sigma^2} (\tilde{x} - \tilde{\mu})^* (\tilde{x} - \tilde{\mu})} \quad (ii)$$

for $\tilde{x} = x_1 + ix_2$, $\tilde{\mu} = \mu_1 + i\mu_2$, $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $\sigma^2 > 0$, $j = 1, 2$. We will write this as $\tilde{x} \sim \tilde{N}_1(\tilde{\mu}, \sigma^2)$. It can be shown that the two parameters appearing in the density in (ii) are the mean value of \tilde{x} and the variance of \tilde{x} . We now establish that the density in (ii) is equivalent to a real bivariate Gaussian density with $\sigma_1^2 = \frac{1}{2}\sigma^2$, $\sigma_2^2 = \frac{1}{2}\sigma^2$ and zero correlation. In the real bivariate normal density, the exponent is the following, with Σ as given below:

$$\begin{aligned} &-\frac{1}{2}[(x_1 - \mu_1), (x_2 - \mu_2)] \Sigma^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \frac{1}{2}\sigma^2 & 0 \\ 0 & \frac{1}{2}\sigma^2 \end{bmatrix} \\ &= -\frac{1}{\sigma^2} \{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2\} = -\frac{1}{\sigma^2} (\tilde{x} - \tilde{\mu})^* (\tilde{x} - \tilde{\mu}). \end{aligned}$$

This exponent agrees with that appearing in the complex case. Now, consider the constant part in the real bivariate case:

$$(2\pi)|\Sigma|^{\frac{1}{2}} = (2\pi) \left| \begin{array}{cc} \frac{1}{2}\sigma^2 & 0 \\ 0 & \frac{1}{2}\sigma^2 \end{array} \right|^{\frac{1}{2}} = \pi\sigma^2,$$

which also coincides with that of the complex Gaussian. Hence, a complex scalar Gaussian is equivalent to a real bivariate Gaussian case whose parameters are as described above.

Let us consider the mgf of the complex Gaussian scalar case. Let $\tilde{t} = t_1 + it_2$, $i = \sqrt{-1}$, with t_1 and t_2 being real parameters, so that $\tilde{t}^* = \bar{\tilde{t}} = t_1 - it_2$ is the conjugate of \tilde{t} . Then $\tilde{t}^*\tilde{x} = t_1x_1 + t_2x_2 + i(t_1x_2 - t_2x_1)$. Note that $t_1x_1 + t_2x_2$ contains the necessary number of parameters (that is, 2) and hence, to be consistent with the definition of the mgf in a real bivariate case, the imaginary part should not be taken into account; thus, we should define the mgf as $E[e^{\Re(\tilde{t}^*\tilde{x})}]$, where $\Re(\cdot)$ denotes the real part of (\cdot) . Accordingly, in the complex case, the mgf is obtained as follows:

$$\begin{aligned} M_{\tilde{x}}(\tilde{t}) &= E[e^{\Re(\tilde{t}^*\tilde{x})}] \\ &= \frac{1}{\pi\sigma^2} \int_{\tilde{x}} e^{\Re(\tilde{t}^*\tilde{x}) - \frac{1}{\sigma^2}(\tilde{x}-\tilde{\mu})^*(\tilde{x}-\tilde{\mu})} d\tilde{x} \\ &= \frac{e^{\Re(\tilde{t}^*\tilde{\mu})}}{\pi\sigma^2} \int_{\tilde{x}} e^{\Re(\tilde{t}^*(\tilde{x}-\tilde{\mu})) - \frac{1}{\sigma^2}(\tilde{x}-\tilde{\mu})^*(\tilde{x}-\tilde{\mu})} d\tilde{x}. \end{aligned}$$

Let us simplify the exponent:

$$\begin{aligned} \Re(\tilde{t}^*(\tilde{x}-\tilde{\mu})) - \frac{1}{\sigma^2}(\tilde{x}-\tilde{\mu})^*(\tilde{x}-\tilde{\mu}) \\ &= -\left\{ \frac{1}{\sigma^2}(x_1-\mu_1)^2 + \frac{1}{\sigma^2}(x_2-\mu_2)^2 - t_1(x_1-\mu_1) - t_2(x_2-\mu_2) \right\} \\ &= \frac{\sigma^2}{4}(t_1^2+t_2^2) - \left\{ (y_1 - \frac{\sigma}{2}t_1)^2 + (y_2 - \frac{\sigma}{2}t_2)^2 \right\} \end{aligned}$$

where $y_1 = \frac{x_1-\mu_1}{\sigma}$, $y_2 = \frac{x_2-\mu_2}{\sigma}$, $dy_j = \frac{1}{\sigma}dx_j$, $j = 1, 2$. But

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(y_j - \frac{\sigma}{2}t_j)^2} dy_j = 1, \quad j = 1, 2.$$

Hence,

$$M_{\tilde{x}}(\tilde{t}) = e^{\Re(\tilde{t}^*\tilde{\mu}) + \frac{\sigma^2}{4}\tilde{t}^*\tilde{t}} = e^{t_1\mu_1 + t_2\mu_2 + \frac{\sigma^2}{4}(t_1^2 + t_2^2)}, \quad (2.1a.1)$$

which is the mgf of the equivalent real bivariate Gaussian distribution.

Note 2.1.1. A statistical density is invariably a real-valued scalar function of the variables involved, be they scalar, vector or matrix variables, real or complex.

2.1.1. Linear functions of Gaussian variables in the real domain

If x_1, \dots, x_k are statistically independently distributed real scalar Gaussian variables with parameters (μ_j, σ_j^2) , $j = 1, \dots, k$ and if a_1, \dots, a_k are real scalar constants then the mgf of a linear function $u = a_1x_1 + \dots + a_kx_k$ is given by

$$\begin{aligned} M_u(t) &= E[e^{tu}] = E[e^{ta_1x_1 + \dots + ta_kx_k}] = M_{x_1}(a_1t) \cdots M_{x_k}(a_kt), \text{ as } M_{ax}(t) = M_x(at), \\ &= e^{(ta_1\mu_1 + \dots + ta_k\mu_k) + \frac{1}{2}t^2(a_1^2\sigma_1^2 + \dots + a_k^2\sigma_k^2)} \\ &= e^{t(\sum_{j=1}^k a_j\mu_j) + \frac{1}{2}t^2(\sum_{j=1}^k a_j^2\sigma_j^2)}, \end{aligned}$$

which is the mgf of a real normal random variable whose parameters are $(\sum_{j=1}^k a_j\mu_j, \sum_{j=1}^k a_j^2\sigma_j^2)$. Hence, the following result:

Theorem 2.1.1. *Let the real scalar random variable x_j have a real univariate normal (Gaussian) distribution, that is, $x_j \sim N_1(\mu_j, \sigma_j^2)$, $j = 1, \dots, k$ and let x_1, \dots, x_k be statistically independently distributed. Then, any linear function $u = a_1x_1 + \dots + a_kx_k$, where a_1, \dots, a_k are real constants, has a real normal distribution with mean value $\sum_{j=1}^k a_j\mu_j$ and variance $\sum_{j=1}^k a_j^2\sigma_j^2$, that is, $u \sim N_1(\sum_{j=1}^k a_j\mu_j, \sum_{j=1}^k a_j^2\sigma_j^2)$.*

Vector/matrix notation enables one to express this result in a more convenient form. Let

$$L = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k^2 \end{bmatrix}.$$

Then denoting the transposes by primes, $u = L'X = X'L$, $E(u) = L'\mu = \mu'L$, and

$$\begin{aligned} \text{Var}(u) &= E[(u - E(u))(u - E(u))'] = L'E[(X - E(X))(X - E(X))']L \\ &= L'\text{Cov}(X)L = L'\Sigma L \end{aligned}$$

where, in this case, Σ is the diagonal matrix $\text{diag}(\sigma_1^2, \dots, \sigma_k^2)$. If x_1, \dots, x_k is a simple random sample from x_1 , that is, from the normal population specified by the density of x_1 or, equivalently, if x_1, \dots, x_k are iid (independently and identically distributed) random

variables having as a common distribution that of x_1 , then $E(u) = \mu_1 L'J = \mu_1 J'L$ and $\text{Var}(u) = \sigma_1^2 L'L$ where u is the linear function defined in Theorem 2.1.1 and $J' = (1, 1, \dots, 1)$ is a vector of unities.

Example 2.1.1. Let $x_1 \sim N_1(-1, 1)$ and $x_2 \sim N_1(2, 2)$ be independently distributed real normal variables. Determine the density of the linear function $u = 5x_1 - 2x_2 + 7$.

Solution 2.1.1. Since u is a linear function of independently distributed real scalar normal variables, it is real scalar normal whose parameters $E(u)$ and $\text{Var}(u)$ are

$$\begin{aligned} E(u) &= 5E(x_1) - 2E(x_2) + 7 = 5(-1) - 2(2) + 7 = -2 \\ \text{Var}(u) &= 25\text{Var}(x_1) + 4\text{Var}(x_2) + 0 = 25(1) + 4(2) = 33, \end{aligned}$$

the covariance being zero since x_1 and x_2 are independently distributed. Thus, $u \sim N_1(-2, 33)$.

2.1a.1. Linear functions in the complex domain

We can also look into the distribution of linear functions of independently distributed complex Gaussian variables. Let a be a constant and \tilde{x} a complex random variable, where a may be real or complex. Then, from the definition of the variance in the complex domain, one has

$$\begin{aligned} \text{Var}(a\tilde{x}) &= E[(a\tilde{x} - E(a\tilde{x}))(a\tilde{x} - E(a\tilde{x}))^*] = aE[(\tilde{x} - E(\tilde{x}))(\tilde{x} - E(\tilde{x}))^*]a^* \\ &= a\text{Var}(\tilde{x})a^* = aa^*\text{Var}(\tilde{x}) = |a|^2\text{Var}(\tilde{x}) = |a|^2\sigma^2 \end{aligned}$$

when the variance of \tilde{x} is σ^2 , where $|a|$ denotes the absolute value of a . As well, $E[a\tilde{x}] = aE[\tilde{x}] = a\tilde{\mu}$. Then, a companion to Theorem 2.1.1 is obtained.

Theorem 2.1a.1. Let $\tilde{x}_1, \dots, \tilde{x}_k$ be independently distributed scalar complex Gaussian variables, $\tilde{x}_j \sim \tilde{N}_1(\tilde{\mu}_j, \sigma_j^2)$, $j = 1, \dots, k$. Let a_1, \dots, a_k be real or complex constants and $\tilde{u} = a_1\tilde{x}_1 + \dots + a_k\tilde{x}_k$ be a linear function. Then, \tilde{u} has a univariate complex Gaussian distribution given by $\tilde{u} \sim \tilde{N}_1(\sum_{j=1}^k a_j\tilde{\mu}_j, \sum_{j=1}^k |a_j|^2\text{Var}(\tilde{x}_j))$.

Example 2.1a.1. Let $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ be independently distributed complex Gaussian univariate random variables with expected values $\tilde{\mu}_1 = -1 + 2i$, $\tilde{\mu}_2 = i$, $\tilde{\mu}_3 = -1 - i$ respectively. Let $\tilde{x}_j = x_{1j} + ix_{2j}$, $j = 1, 2, 3$. Let $[\text{Var}(x_{1j}), \text{Var}(x_{2j})] = [(1, 1), (1, 2), (2, 3)]$, respectively. Let $a_1 = 1 + i$, $a_2 = 2 - 3i$, $a_3 = 2 + i$, $a_4 = 3 + 2i$. Determine the density of the linear function $\tilde{u} = a_1\tilde{x}_1 + a_2\tilde{x}_2 + a_3\tilde{x}_3 + a_4$.

Solution 2.1a.1.

$$\begin{aligned}
E(\tilde{u}) &= a_1 E(\tilde{x}_1) + a_2 E(\tilde{x}_2) + a_3 E(\tilde{x}_3) + a_4 \\
&= (1+i)(-1+2i) + (2-3i)(i) + (2+i)(-1-i) + (3+2i) \\
&= (-3+i) + (3+2i) + (-1-3i) + (3+2i) = 2+2i; \\
\text{Var}(\tilde{u}) &= |a_1|^2 \text{Var}(\tilde{x}_1) + |a_2|^2 \text{Var}(\tilde{x}_2) + |a_3|^2 \text{Var}(\tilde{x}_3)
\end{aligned}$$

and the covariances are equal to zero since the variables are independently distributed. Note that $\tilde{x}_1 = x_{11} + ix_{21}$ and hence, for example,

$$\begin{aligned}
\text{Var}(\tilde{x}_1) &= E[(\tilde{x}_1 - E(\tilde{x}_1))(\tilde{x}_1 - E(\tilde{x}_1))^*] \\
&= E\{[(x_{11} - E(x_{11})) + i(x_{21} - E(x_{21}))][(x_{11} - E(x_{11})) - i(x_{21} - E(x_{21}))]\} \\
&= E[(x_{11} - E(x_{11}))^2] - (i)^2 E[(x_{21} - E(x_{21}))^2] \\
&= \text{Var}(x_{11}) + \text{Var}(x_{21}) = 1 + 1 = 2.
\end{aligned}$$

Similarly, $\text{Var}(\tilde{x}_2) = 1 + 2 = 3$, $\text{Var}(\tilde{x}_3) = 2 + 3 = 5$. Moreover, $|a_1|^2 = (1)^2 + (1)^2 = 2$, $|a_2|^2 = (2)^2 + (3)^2 = 13$, $|a_3|^2 = (2)^2 + (1)^2 = 5$. Accordingly, $\text{Var}(\tilde{u}) = 2(2) + (13)(3) + (5)(5) = 68$. Thus, $\tilde{u} \sim \tilde{N}_1(2+2i, 68)$. Note that the constant a_4 only affects the mean value. Had a_4 been absent from \tilde{u} , its mean value would have been real and equal to -1 .

2.1.2. The chisquare distribution in the real domain

Suppose that x_1 follows a real standard normal distribution, that is, $x_1 \sim N_1(0, 1)$, whose mean value is zero and variance, 1. What is then the density of x_1^2 , the square of a real standard normal variable? Let the distribution function or cumulative distribution function of x_1 be $F_{x_1}(t) = \text{Pr}\{x_1 \leq t\}$ and that of $y_1 = x_1^2$ be $F_{y_1}(t) = \text{Pr}\{y_1 \leq t\}$. Note that since $y_1 > 0$, t must be positive. Then,

$$\begin{aligned}
F_{y_1}(t) &= \text{Pr}\{y_1 \leq t\} = \text{Pr}\{x_1^2 \leq t\} = \text{Pr}\{|x_1| \leq \sqrt{t}\} = \text{Pr}\{-\sqrt{t} \leq x_1 \leq \sqrt{t}\} \\
&= \text{Pr}\{x_1 \leq \sqrt{t}\} - \text{Pr}\{x_1 \leq -\sqrt{t}\} = F_{x_1}(\sqrt{t}) - F_{x_1}(-\sqrt{t}). \tag{i}
\end{aligned}$$

Denoting the density of y_1 by $g(y_1)$, this density at $y_1 = t$ is available by differentiating the distribution function $F_{y_1}(t)$ with respect to t . As for the density of x_1 , which is the standard normal density, it can be obtained by differentiating $F_{x_1}(\sqrt{t})$ with respect to \sqrt{t} . Thus, differentiating (i) throughout with respect to t , we have

$$\begin{aligned}
g(t)|_{t=y_1} &= \left[\frac{d}{dt} F_{x_1}(\sqrt{t}) - \frac{d}{dt} F_{x_1}(-\sqrt{t}) \right] \Big|_{t=y_1} \\
&= \left[\frac{d}{d\sqrt{t}} F_{x_1}(\sqrt{t}) \frac{d\sqrt{t}}{dt} - \frac{d}{d\sqrt{t}} F_{x_1}(-\sqrt{t}) \frac{d(-\sqrt{t})}{dt} \right] \Big|_{t=y_1} \\
&= \frac{1}{2} t^{\frac{1}{2}-1} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t} \Big|_{t=y_1} + \frac{1}{2} t^{\frac{1}{2}-1} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t} \Big|_{t=y_1} \\
&= \frac{1}{2^{\frac{1}{2}} \Gamma(1/2)} y_1^{\frac{1}{2}-1} e^{-\frac{1}{2}y_1}, \quad 0 \leq y_1 < \infty, \quad \text{with } \Gamma(1/2) = \sqrt{\pi}. \quad (ii)
\end{aligned}$$

Accordingly, the density of $y_1 = x_1^2$ or the square of a real standard normal variable, is a two-parameter real gamma with $\alpha = \frac{1}{2}$ and $\beta = 2$ or a real chisquare with one degree of freedom. A two-parameter real gamma density with the parameters (α, β) is given by

$$f_1(y_1) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y_1^{\alpha-1} e^{-\frac{y_1}{\beta}}, \quad 0 \leq y_1 < \infty, \quad \alpha > 0, \quad \beta > 0, \quad (2.1.2)$$

and $f_1(y_1) = 0$ elsewhere. When $\alpha = \frac{n}{2}$ and $\beta = 2$, we have a real chisquare density with n degrees of freedom. Hence, the following result:

Theorem 2.1.2. *The square of a real scalar standard normal random variable is a real chisquare variable with one degree of freedom. A real chisquare with n degrees of freedom has the density given in (2.1.2) with $\alpha = \frac{n}{2}$ and $\beta = 2$.*

A real scalar chisquare random variable with m degrees of freedom is denoted as χ_m^2 . From (2.1.2), by computing the mgf we can see that the mgf of a real scalar gamma random variable y is $M_y(t) = (1 - \beta t)^{-\alpha}$ for $1 - \beta t > 0$. Hence, a real chisquare with m degrees of freedom has the mgf $M_{\chi_m^2}(t) = (1 - 2t)^{-\frac{m}{2}}$ for $1 - 2t > 0$. The condition $1 - \beta t > 0$ is required for the convergence of the integral when evaluating the mgf of a real gamma random variable. If $y_j \sim \chi_{m_j}^2$, $j = 1, \dots, k$ and if y_1, \dots, y_k are independently distributed, then the sum $y = y_1 + \dots + y_k \sim \chi_{m_1 + \dots + m_k}^2$, a real chisquare with $m_1 + \dots + m_k$ degrees of freedom, with mgf $M_y(t) = (1 - 2t)^{-\frac{1}{2}(m_1 + \dots + m_k)}$ for $1 - 2t > 0$.

Example 2.1.2. Let $x_1 \sim N_1(-1, 4)$, $x_2 \sim N_1(2, 2)$ be independently distributed. Let $u = x_1^2 + 2x_2^2 + 2x_1 - 8x_2 + 5$. Compute the density of u .

Solution 2.1.2.

$$\begin{aligned}
u &= x_1^2 + 2x_2^2 + 2x_1 - 8x_2 + 5 = (x_1 + 1)^2 + 2(x_2 - 2)^2 - 4 \\
&= 4 \left[\frac{(x_1 + 1)^2}{4} + \frac{(x_2 - 2)^2}{2} \right] - 4.
\end{aligned}$$

Since $x_1 \sim N_1(-1, 4)$ and $x_2 \sim N_1(2, 2)$ are independently distributed, so are $\frac{(x_1+1)^2}{4} \sim \chi_1^2$ and $\frac{(x_2-2)^2}{2} \sim \chi_1^2$, and hence the sum is a real χ_2^2 random variable. Then, $u = 4y - 4$ with $y = \chi_2^2$. But the density of y , denoted by $f_y(y)$, is

$$f_y(y) = \frac{1}{2}e^{-\frac{y}{2}}, \quad 0 \leq y < \infty,$$

and $f_y(y) = 0$ elsewhere. Then, $z = 4y$ has the density

$$f_z(z) = \frac{1}{8}e^{-\frac{z}{8}}, \quad 0 \leq z < \infty,$$

and $f_z(z) = 0$ elsewhere. However, since $u = z - 4$, its density is

$$f_u(u) = \frac{1}{8}e^{-\frac{(u+4)}{8}}, \quad -4 \leq u < \infty,$$

and zero elsewhere.

2.1a.2. The chisquare distribution in the complex domain

Let us consider the distribution of $\tilde{z}_1 \tilde{z}_1^*$ of a scalar standard complex normal variable \tilde{z}_1 . The density of \tilde{z}_1 is given by

$$f_{\tilde{z}_1}(\tilde{z}_1) = \frac{1}{\pi}e^{-\tilde{z}_1^* \tilde{z}_1}, \quad \tilde{z}_1 = z_{11} + iz_{12}, \quad -\infty < z_{1j} < \infty, \quad j = 1, 2.$$

Let $\tilde{u}_1 = \tilde{z}_1^* \tilde{z}_1$. Note that $\tilde{z}_1^* \tilde{z}_1$ is real and hence we may associate a real parameter t to the mgf. Note that $\tilde{z}_1 \tilde{z}_1^*$ in the scalar complex case corresponds to z^2 in the real scalar case where $z \sim N_1(0, 1)$. Then, the mgf of \tilde{u}_1 is given by

$$M_{\tilde{u}_1}(t) = E[e^{\Re(\tilde{t}\tilde{u}_1)}] = \frac{1}{\pi} \int_{\tilde{z}_1} e^{-\Re[(1-t)\tilde{z}_1^* \tilde{z}_1]} d\tilde{z}_1.$$

However, $\tilde{z}_1^* \tilde{z}_1 = z_{11}^2 + z_{12}^2$ as $\tilde{z}_1 = z_{11} + iz_{12}$, $i = \sqrt{-1}$, where z_{11} and z_{12} are real. Thus, the above integral gives $(1-t)^{-\frac{1}{2}}(1-t)^{-\frac{1}{2}} = (1-t)^{-1}$ for $1-t > 0$, which is the mgf of a real scalar gamma variable with parameters $\alpha = 1$ and $\beta = 1$. Let $\tilde{z}_j \sim \tilde{N}_1(\tilde{\mu}_j, \sigma^2)$, $j = 1, \dots, k$, be scalar complex normal random variables that are independently distributed. Letting

$$\tilde{u} = \sum_{j=1}^k \left(\frac{\tilde{z}_j - \tilde{\mu}_j}{\sigma_j} \right)^* \left(\frac{\tilde{z}_j - \tilde{\mu}_j}{\sigma_j} \right) \sim \text{real scalar gamma with parameters } \alpha = k, \beta = 1,$$

whose density is

$$f_{\tilde{u}}(u) = \frac{1}{\Gamma(k)} u^{k-1} e^{-u}, \quad 0 \leq u < \infty, \quad k = 1, 2, \dots, \quad (2.1a.2)$$

\tilde{u} is referred to as a scalar chisquare in the complex domain having k degrees of freedom, which is denoted $\tilde{u} \sim \tilde{\chi}_k^2$. Hence, the following result:

Theorem 2.1a.2. Let $\tilde{z}_j \sim \tilde{N}_1(\tilde{\mu}_j, \sigma_j^2)$, $j = 1, \dots, k$, be independently distributed and $\tilde{u} = \sum_{j=1}^k (\frac{\tilde{z}_j - \tilde{\mu}_j}{\sigma_j})^* (\frac{\tilde{z}_j - \tilde{\mu}_j}{\sigma_j})$. Then \tilde{u} is called a scalar chisquare having k degrees of freedom in the complex domain whose density as given in (2.1a.2) is that of a real scalar gamma random variable with parameters $\alpha = k$ and $\beta = 1$.

Example 2.1a.2. Let $\tilde{x}_1 \sim \tilde{N}_1(i, 2)$, $\tilde{x}_2 \sim \tilde{N}_1(1 - i, 1)$ be independently distributed complex Gaussian univariate random variables. Let $\tilde{u} = \tilde{x}_1^* \tilde{x}_1 + 2\tilde{x}_2^* \tilde{x}_2 - 2\tilde{x}_2^* - 2\tilde{x}_2 + i(\tilde{x}_1 + 2\tilde{x}_2^*) - i(\tilde{x}_1^* + 2\tilde{x}_2) + 5$. Evaluate the density of \tilde{u} .

Solution 2.1a.2. Let us simplify \tilde{u} , keeping in mind the parameters in the densities of \tilde{x}_1 and \tilde{x}_2 . Since terms of the type $\tilde{x}_1^* \tilde{x}_1$ and $\tilde{x}_2^* \tilde{x}_2$ are present in \tilde{u} , we may simplify into factors involving \tilde{x}_j^* and \tilde{x}_j for $j = 1, 2$. From the density of \tilde{x}_1 we have

$$\frac{(\tilde{x}_1 - i)^* (\tilde{x}_1 - i)}{2} \sim \tilde{\chi}_1^2$$

where

$$(\tilde{x}_1 - i)^* (\tilde{x}_1 - i) = (\tilde{x}_1^* + i)(\tilde{x}_1 - i) = \tilde{x}_1^* \tilde{x}_1 + i\tilde{x}_1 - i\tilde{x}_1^* + 1. \quad (i)$$

After removing the elements in (i) from \tilde{u} , the remainder is

$$\begin{aligned} & 2\tilde{x}_2^* \tilde{x}_2 - 2\tilde{x}_2^* - 2\tilde{x}_2 + 2i\tilde{x}_2^* - 2i\tilde{x}_2 + 4 \\ &= 2[(\tilde{x}_2 - 1)^* (\tilde{x}_2 - 1) - i\tilde{x}_2 + i\tilde{x}_2^* + 1] \\ &= 2[(\tilde{x}_2 - 1 + i)^* (\tilde{x}_2 - 1 + i)]. \end{aligned}$$

Accordingly,

$$\begin{aligned} \tilde{u} &= 2 \left[\frac{(\tilde{x}_1 - i)^* (\tilde{x}_1 - i)}{2} + (\tilde{x}_2 - 1 + i)^* (\tilde{x}_2 - 1 + i) \right] \\ &= 2[\tilde{\chi}_1^2 + \tilde{\chi}_1^2] = 2\tilde{\chi}_2^2 \end{aligned}$$

where $\tilde{\chi}_2^2$ is a scalar chisquare of degree 2 in the complex domain or, equivalently, a real scalar gamma with parameters $(\alpha = 2, \beta = 1)$. Letting $y = \tilde{\chi}_2^2$, the density of y , denoted by $f_y(y)$, is

$$f_y(y) = y e^{-y}, \quad 0 \leq y < \infty,$$

and $f_y(y) = 0$ elsewhere. Then, the density of $u = 2y$, denoted by $f_u(u)$, which is given by

$$f_u(u) = \frac{u}{2} e^{-\frac{u}{2}}, \quad 0 \leq u < \infty,$$

and $f_u(u) = 0$ elsewhere, is that of a real scalar gamma with the parameters ($\alpha = 2$, $\beta = 2$).

2.1.3. The type-2 beta and F distributions in the real domain

What about the distribution of the ratio of two independently distributed real scalar chisquare random variables? Let $y_1 \sim \chi_m^2$ and $y_2 \sim \chi_n^2$, that is, y_1 and y_2 are real chisquare random variables with m and n degrees of freedom respectively, and assume that y_1 and y_2 are independently distributed. Let us determine the density of $u = y_1/y_2$. Let $v = y_2$ and consider the transformation (y_1, y_2) onto (u, v) . Noting that

$$\frac{\partial u}{\partial y_1} = \frac{1}{y_2}, \quad \frac{\partial v}{\partial y_2} = 1, \quad \frac{\partial v}{\partial y_1} = 0,$$

one has

$$\begin{aligned} du \wedge dv &= \begin{vmatrix} \frac{\partial u}{\partial y_1} & \frac{\partial u}{\partial y_2} \\ \frac{\partial v}{\partial y_1} & \frac{\partial v}{\partial y_2} \end{vmatrix} dy_1 \wedge dy_2 = \begin{vmatrix} \frac{1}{y_2} & * \\ 0 & 1 \end{vmatrix} dy_1 \wedge dy_2 \\ &= \frac{1}{y_2} dy_1 \wedge dy_2 \Rightarrow dy_1 \wedge dy_2 = v du \wedge dv \end{aligned}$$

where the asterisk indicates the presence of some element in which we are not interested owing to the triangular pattern for the Jacobian matrix. Letting the joint density of y_1 and y_2 be denoted by $f_{12}(y_1, y_2)$, one has

$$f_{12}(y_1, y_2) = \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} y_1^{\frac{m}{2}-1} y_2^{\frac{n}{2}-1} e^{-\frac{y_1+y_2}{2}}$$

for $0 \leq y_1 < \infty$, $0 \leq y_2 < \infty$, $m, n = 1, 2, \dots$, and $f_{12}(y_1, y_2) = 0$ elsewhere. Let the joint density of u and v be denoted by $g_{12}(u, v)$ and the marginal density of u be denoted by $g_1(u)$. Then,

$$\begin{aligned}
 g_{12}(u, v) &= c v(uv)^{\frac{m}{2}-1} v^{\frac{n}{2}-1} e^{-\frac{1}{2}(uv+v)}, \quad c = \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})}, \\
 g_1(u) &= c u^{\frac{m}{2}-1} \int_{v=0}^{\infty} v^{\frac{m+n}{2}-1} e^{-v\frac{(1+u)}{2}} dv \\
 &= c u^{\frac{m}{2}-1} \Gamma\left(\frac{m+n}{2}\right) \left(\frac{1+u}{2}\right)^{-\frac{m+n}{2}} \\
 &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} u^{\frac{m}{2}-1} (1+u)^{-\frac{m+n}{2}} \tag{2.1.3}
 \end{aligned}$$

for $m, n = 1, 2, \dots$, $0 \leq u < \infty$ and $g_1(u) = 0$ elsewhere. Note that $g_1(u)$ is a type-2 real scalar beta density. Hence, we have the following result:

Theorem 2.1.3. *Let the real scalar $y_1 \sim \chi_m^2$ and $y_2 \sim \chi_n^2$ be independently distributed, then the ratio $u = \frac{y_1}{y_2}$ is a type-2 real scalar beta random variable with the parameters $\frac{m}{2}$ and $\frac{n}{2}$ where $m, n = 1, 2, \dots$, whose density is provided in (2.1.3).*

This result also holds for general real scalar gamma random variables $x_1 > 0$ and $x_2 > 0$ with parameters (α_1, β) and (α_2, β) , respectively, where β is a common scale parameter and it is assumed that x_1 and x_2 are independently distributed. Then, $u = \frac{x_1}{x_2}$ is a type-2 beta with parameters α_1 and α_2 .

If u as defined in Theorem 2.1.3 is replaced by $\frac{m}{n} F_{m,n}$ or $F = F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n} = \frac{n}{m} u$ is known as the F -random variable with m and n degrees of freedom, where the degrees of freedom indicate those of the numerator and denominator chisquare random variables which are independently distributed. Denoting the density of F by $f_F(F)$ we have the following result:

Theorem 2.1.4. *Letting $F = F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n}$ where the two real scalar chisquares are independently distributed, the real scalar F -density is given by*

$$f_F(F) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{F^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n} F\right)^{\frac{m+n}{2}}} \tag{2.1.4}$$

for $0 \leq F < \infty$, $m, n = 1, 2, \dots$, and $f_F(F) = 0$ elsewhere.

Example 2.1.3. Let x_1 and x_2 be independently distributed real scalar gamma random variables with parameters (α_1, β) and (α_2, β) , respectively, β being a common parameter, whose densities are as specified in (2.1.2). Let $u_1 = \frac{x_1}{x_1+x_2}$, $u_2 = \frac{x_1}{x_2}$, $u_3 = x_1 + x_2$.

Show that (1): u_3 has a real scalar gamma density as given in (2.1.2) with the parameters $(\alpha_1 + \alpha_2, \beta)$; (2): u_1 and u_3 as well as u_2 and u_3 are independently distributed; (3): u_2 is a real scalar type-2 beta with parameters (α_1, α_2) whose density is specified in (2.1.3); (4): u_1 has a real scalar type-1 beta density given as

$$f_1(u_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u_1^{\alpha_1-1} (1 - u_1)^{\alpha_2-1}, \quad 0 \leq u_1 \leq 1, \quad (2.1.5)$$

for $\Re(\alpha_1) > 0$, $\Re(\alpha_2) > 0$ and zero elsewhere. [In a statistical density, the parameters are usually real; however, since the integrals exist for complex parameters, the conditions are given for complex parameters as the real parts of α_1 and α_2 , which must be positive. When they are real, the conditions will be simply $\alpha_1 > 0$ and $\alpha_2 > 0$.]

Solution 2.1.3. Since x_1 and x_2 are independently distributed, their joint density is the product of the marginal densities, which is given by

$$f_{12}(x_1, x_2) = c x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-\frac{1}{\beta}(x_1+x_2)}, \quad 0 \leq x_j < \infty, \quad j = 1, 2, \quad (i)$$

for $\Re(\alpha_j) > 0$, $\Re(\beta) > 0$, $j = 1, 2$ and zero elsewhere, where

$$c = \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2)}.$$

Since the sum $x_1 + x_2$ is present in the exponent and both x_1 and x_2 are positive, a convenient transformation is $x_1 = r \cos^2 \theta$, $x_2 = r \sin^2 \theta$, $0 \leq r < \infty$, $0 \leq \theta \leq \frac{\pi}{2}$. Then, the Jacobian is available from the detailed derivation of Jacobian given in the beginning of Sect. 2.1.3 or from Example 1.6.1. That is,

$$dx_1 \wedge dx_2 = 2r \sin \theta \cos \theta dr \wedge d\theta. \quad (ii)$$

Then from (i) and (ii), the joint density of r and θ , denoted by $f_{r,\theta}(r, \theta)$, is the following:

$$f_{r,\theta}(r, \theta) = c (\cos^2 \theta)^{\alpha_1-1} (\sin^2 \theta)^{\alpha_2-1} 2 \cos \theta \sin \theta r^{\alpha_1+\alpha_2-1} e^{-\frac{1}{\beta}r} \quad (iii)$$

and zero elsewhere. As $f_{r,\theta}(r, \theta)$ is a product of positive integrable functions involving solely r and θ , r and θ are independently distributed. Since $u_3 = x_1 + x_2 = r \cos^2 \theta + r \sin^2 \theta = r$ is solely a function of r and $u_1 = \frac{x_1}{x_1+x_2} = \cos^2 \theta$ and $u_2 = \frac{\cos^2 \theta}{\sin^2 \theta}$ are solely functions of θ , it follows that u_1 and u_3 as well as u_2 and u_3 are independently distributed. From (iii), upon multiplying and dividing by $\Gamma(\alpha_1 + \alpha_2)$, we obtain the density of u_3 as

$$f_1(u_3) = \frac{1}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1 + \alpha_2)} u_3^{\alpha_1+\alpha_2-1} e^{-\frac{u_3}{\beta}}, \quad 0 \leq u_3 < \infty, \quad (iv)$$

and zero elsewhere, which is a real scalar gamma density with parameters $(\alpha_1 + \alpha_2, \beta)$. From (iii), the density of θ , denoted by $f_2(\theta)$, is

$$f_2(\theta) = c (\cos^2 \theta)^{\alpha_1-1} (\sin^2 \theta)^{\alpha_2-1}, \quad 0 \leq \theta \leq \frac{\pi}{2} \quad (v)$$

and zero elsewhere, for $\Re(\alpha_j) > 0$, $j = 1, 2, \dots$. From this result, we can obtain the density of $u_1 = \cos^2 \theta$. Then, $du_1 = -2 \cos \theta \sin \theta d\theta$. Moreover, when $\theta \rightarrow 0$, $u_1 \rightarrow 1$ and when $\theta \rightarrow \frac{\pi}{2}$, $u_1 \rightarrow 0$. Hence, the minus sign in the Jacobian is needed to obtain the limits in the natural order, $0 \leq u_1 \leq 1$. Substituting in (v), the density of u_1 denoted by $f_3(u_1)$, is as given in (2.1.5), u_1 being a real scalar type-1 beta random variable with parameters (α_1, α_2) . Now, observe that

$$u_2 = \frac{\cos^2 \theta}{\sin^2 \theta} = \frac{\cos^2 \theta}{1 - \cos^2 \theta} = \frac{u_1}{1 - u_1}. \quad (vi)$$

Given the density of u_1 as specified in (2.1.5), we can obtain the density of u_2 as follows. As $u_2 = \frac{u_1}{1-u_1}$, we have $u_1 = \frac{u_2}{1+u_2} \Rightarrow du_1 = \frac{1}{(1+u_2)^2} du_2$; then substituting these values in the density of u_1 , we have the following density for u_2 :

$$f_3(u_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u_2^{\alpha_1-1} (1 + u_2)^{-(\alpha_1+\alpha_2)}, \quad 0 \leq u_2 < \infty, \quad (2.1.6)$$

and zero elsewhere, for $\Re(\alpha_j) > 0$, $j = 1, 2$, which is a real scalar type-2 beta density with parameters (α_1, α_2) . The results associated with the densities (2.1.5) and (2.1.6) are now stated as a theorem.

Theorem 2.1.5. *Let x_1 and x_2 be independently distributed real scalar gamma random variables with the parameters (α_1, β) , (α_2, β) , respectively, β being a common scale parameter. [If $x_1 \sim \chi_m^2$ and $x_2 \sim \chi_n^2$, then $\alpha_1 = \frac{m}{2}$, $\alpha_2 = \frac{n}{2}$ and $\beta = 2$.] Then $u_1 = \frac{x_1}{x_1+x_2}$ is a real scalar type-1 beta whose density is as specified in (2.1.5) with the parameters (α_1, α_2) , and $u_2 = \frac{x_1}{x_2}$ is a real scalar type-2 beta whose density is as given in (2.1.6) with the parameters (α_1, α_2) .*

2.1a.3. The type-2 beta and F distributions in the complex domain

It follows that in the complex domain, if $\tilde{\chi}_m^2$ and $\tilde{\chi}_n^2$ are independently distributed, then the sum is a chisquare with $m + n$ degrees of freedom, that is, $\tilde{\chi}_m^2 + \tilde{\chi}_n^2 = \tilde{\chi}_{m+n}^2$. We now look into type-2 beta variables and F -variables and their connection to chisquare variables in the complex domain. Since, in the complex domain, the chisquares are actually real variables, the density of the ratio of two independently distributed chisquares with m and n degrees of freedom in the complex domain, remains the same as the density given in

(2.1.3) with $\frac{m}{2}$ and $\frac{n}{2}$ replaced by m and n , respectively. Thus, letting $\tilde{u} = \tilde{\chi}_m^2 / \tilde{\chi}_n^2$ where the two chisquares in the complex domain are independently distributed, the density of \tilde{u} , denoted by $\tilde{g}_1(u)$, is

$$\tilde{g}_1(u) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} u^{m-1} (1+u)^{-(m+n)} \quad (2.1a.3)$$

for $0 \leq u < \infty$, $m, n = 1, 2, \dots$, and $\tilde{g}_1(u) = 0$ elsewhere.

Theorem 2.1a.3. *Let $\tilde{y}_1 \sim \tilde{\chi}_m^2$ and $\tilde{y}_2 \sim \tilde{\chi}_n^2$ be independently distributed where \tilde{y}_1 and \tilde{y}_2 are in the complex domain; then, $\tilde{u} = \frac{\tilde{y}_1}{\tilde{y}_2}$ is a real type-2 beta whose density is given in (2.1a.3).*

If the F random variable in the complex domain is defined as $\tilde{F}_{m,n} = \frac{\tilde{\chi}_m^2/m}{\tilde{\chi}_n^2/n}$ where the two chisquares in the complex domain are independently distributed, then the density of \tilde{F} is that of the real F -density with m and n replaced by $2m$ and $2n$ in (2.1.4), respectively.

Theorem 2.1a.4. *Let $\tilde{F} = \tilde{F}_{m,n} = \frac{\tilde{\chi}_m^2/m}{\tilde{\chi}_n^2/n}$ where the two chisquares in the complex domain are independently distributed; then, \tilde{F} is referred to as an F random variable in the complex domain and it has a real F -density with the parameters m and n , which is given by*

$$\tilde{g}_2(F) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} \left(\frac{m}{n}\right)^m F^{m-1} \left(1 + \frac{m}{n}F\right)^{-(m+n)} \quad (2.1a.4)$$

for $0 \leq F < \infty$, $m, n = 1, 2, \dots$, and $\tilde{g}_2(F) = 0$ elsewhere.

A type-1 beta representation in the complex domain can similarly be obtained from Theorem 2.1a.3. This will be stated as a theorem.

Theorem 2.1a.5. *Let $\tilde{x}_1 \sim \tilde{\chi}_m^2$ and $\tilde{x}_2 \sim \tilde{\chi}_n^2$ be independently distributed scalar chisquare variables in the complex domain with m and n degrees of freedom, respectively. Let $\tilde{u}_1 = \frac{\tilde{x}_1}{\tilde{x}_1 + \tilde{x}_2}$, which is a real variable that we will call u_1 . Then, \tilde{u}_1 is a scalar type-1 beta random variable in the complex domain with the parameters m, n , whose real scalar density is*

$$\tilde{f}_1(\tilde{u}_1) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} u_1^{m-1} (1-u_1)^{n-1}, \quad 0 \leq u_1 \leq 1, \quad (2.1a.5)$$

and zero elsewhere, for $m, n = 1, 2, \dots$

2.1.4. Power transformation of type-1 and type-2 beta random variables

Let us make a power transformation of the type $u_1 = ay^\delta$, $a > 0$, $\delta > 0$. Then, $du_1 = a\delta y^{\delta-1}dy$. For convenience, let the parameters in (2.1.5) be $\alpha_1 = \alpha$ and $\alpha_2 = \beta$. Then, the density given in (2.1.5) becomes

$$f_{11}(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \delta a^\alpha y^{\alpha\delta-1} (1 - ay^\delta)^{\beta-1}, \quad 0 \leq y \leq \frac{1}{a^{\frac{1}{\delta}}}, \quad (2.1.7)$$

and zero elsewhere, for $a > 0$, $\delta > 0$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$. We can extend the support to $-a^{-\frac{1}{\delta}} \leq y \leq a^{-\frac{1}{\delta}}$ by replacing y by $|y|$ and multiplying the normalizing constant by $\frac{1}{2}$. Such power transformed models are useful in practical applications. Observe that a power transformation has the following effect: for $y < 1$, the density is reduced if $\delta > 1$ or raised if $\delta < 1$, whereas for $y > 1$, the density increases if $\delta > 1$ or diminishes if $\delta < 1$. For instance, the particular case $\alpha = 1$ is highly useful in reliability theory and stress-strength analysis. Thus, letting $\alpha = 1$ in the original real scalar type-1 beta density (2.1.7) and denoting the resulting density by $f_{12}(y)$, one has

$$f_{12}(y) = a\delta\beta y^{\delta-1} (1 - ay^\delta)^{\beta-1}, \quad 0 \leq y \leq a^{-\frac{1}{\delta}}, \quad (2.1.8)$$

for $a > 0$, $\delta > 0$, $\Re(\beta) > 0$, and zero elsewhere. In the model in (2.1.8), the reliability, that is, $Pr\{y \geq t\}$, for some t , can be easily determined. As well, the hazard function $\frac{f_{12}(y=t)}{Pr\{y \geq t\}}$ is readily available. Actually, the reliability or survival function is

$$Pr\{y \geq t\} = (1 - at^\delta)^\beta, \quad a > 0, \delta > 0, t > 0, \beta > 0, \quad (i)$$

and the hazard function is

$$\frac{f_{12}(y=t)}{Pr\{y \geq t\}} = \frac{a\delta\beta t^{\delta-1}}{1 - at^\delta}. \quad (ii)$$

Observe that the free parameters a , δ and β allow for much versatility in model building situations. If $\beta = 1$ in the real scalar type-1 beta model in (2.1.7), then the density reduces to $\alpha y^{\alpha-1}$, $0 \leq y \leq 1$, $\alpha > 0$, which is a simple power function. The most popular power function model in the statistical literature is the Weibull model, which is a power transformed exponential density. Consider the real scalar exponential density

$$g(x) = \theta e^{-\theta x}, \quad \theta > 0, x \geq 0, \quad (iii)$$

and zero elsewhere, and let $x = y^\delta$, $\delta > 0$. Then the model in (iii) becomes the real scalar Weibull density, denoted by $g_1(y)$:

$$g_1(y) = \theta\delta y^{\delta-1} e^{-\theta y^\delta}, \quad \theta > 0, \delta > 0, y \geq 0, \quad (iv)$$

and zero elsewhere.

Now, let us consider power transformations in a real scalar type-2 beta density given in (2.1.6). For convenience let $\alpha_1 = \alpha$ and $\alpha_2 = \beta$. Letting $y_2 = ay^\delta$, $a > 0$, $\delta > 0$, the model specified by (2.1.6) then becomes

$$f_{21}(y) = a^\alpha \delta \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha\delta-1} (1 + ay^\delta)^{-(\alpha+\beta)} \quad (v)$$

for $a > 0$, $\delta > 0$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$, and zero elsewhere. As in the type-1 beta case, the most interesting special case occurs when $\alpha = 1$. Denoting the resulting density by $f_{22}(y)$, we have

$$f_{22}(y) = a\delta\beta y^{\delta-1} (1 + ay^\delta)^{-(\beta+1)}, \quad 0 \leq y < \infty, \quad (2.1.9)$$

for $a > 0$, $\delta > 0$, $\Re(\beta) > 0$, $\alpha = 1$, and zero elsewhere. In this case as well, the reliability and hazard functions can easily be determined:

$$\text{Reliability function} = Pr\{y \geq t\} = (1 + at^\delta)^{-\beta}, \quad (vi)$$

$$\text{Hazard function} = \frac{f_{22}(y=t)}{Pr\{y \geq t\}} = \frac{a\delta\beta t^{\delta-1}}{1 + at^\delta}. \quad (vii)$$

Again, for application purposes, the forms in (vi) and (vii) are seen to be very versatile due to the presence of the free parameters a , δ and β .

2.1.5. Exponentiation of real scalar type-1 and type-2 beta variables

Let us consider the real scalar type-1 beta model in (2.1.5) where, for convenience, we let $\alpha_1 = \alpha$ and $\alpha_2 = \beta$. Letting $u_1 = ae^{-by}$, we denote the resulting density by $f_{13}(y)$ where

$$f_{13}(y) = a^\alpha b \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{-b\alpha y} (1 - ae^{-by})^{\beta-1}, \quad y \geq \ln a^{\frac{1}{b}}, \quad (2.1.10)$$

for $a > 0$, $b > 0$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$, and zero elsewhere. Again, for practical application the special case $\alpha = 1$ is the most useful one. Let the density corresponding to this special case be denoted by $f_{14}(y)$. Then,

$$f_{14}(y) = ab\beta e^{-by} (1 - ae^{-by})^{\beta-1}, \quad y \geq \ln a^{\frac{1}{b}}, \quad (i)$$

for $a > 0$, $b > 0$, $\beta > 0$, and zero elsewhere. In this case,

$$\text{Reliability function} = Pr\{y \geq t\} = (1 - ae^{-bt})^\beta, \quad (ii)$$

$$\text{Hazard function} = \frac{f_{14}(y=t)}{Pr\{y \geq t\}} = \frac{ab\beta e^{-bt}}{[1 - ae^{-bt}]}. \quad (iii)$$

Now, consider exponentiating a real scalar type-2 beta random variable whose density is given in (2.1.6). For convenience, we will let the parameters be $(\alpha_1 = \alpha$ and $\alpha_2 = \beta)$. Letting $u_2 = e^{-by}$ in (2.1.6), we obtain the following density:

$$f_{21}(y) = a^\alpha b \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{-b\alpha y} (1 + ae^{-by})^{-(\alpha+\beta)}, \quad -\infty < y < \infty, \quad (2.1.11)$$

for $a > 0$, $b > 0$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$, and zero elsewhere. The model in (2.1.11) is in fact the *generalized logistic model* introduced by Mathai and Provost (2006). For the special case $\alpha = 1$, $\beta = 1$, $a = 1$, $b = 1$ in (2.1.11), we have the following density:

$$f_{22}(y) = \frac{e^{-y}}{(1 + e^{-y})^2} = \frac{e^y}{(1 + e^y)^2}, \quad -\infty < y < \infty. \quad (iv)$$

This is the famous *logistic model* which is utilized in industrial applications.

2.1.6. The Student- t distribution in the real domain

A real Student- t variable with ν degrees of freedom, denoted by t_ν , is defined as $t_\nu = \frac{z}{\sqrt{\chi_\nu^2/\nu}}$ where $z \sim N_1(0, 1)$ and χ_ν^2 is a real scalar chisquare with ν degrees of freedom, z and χ_ν^2 being independently distributed. It follows from the definition of a real $F_{m,n}$ random variable, that $t_\nu^2 = \frac{z^2}{\chi_\nu^2/\nu} = F_{1,\nu}$, an F random variable with 1 and ν degrees of freedom. Thus, the density of t_ν^2 is available from that of an $F_{1,\nu}$. On substituting the values $m = 1$, $n = \nu$ in the F -density appearing in (2.1.4), we obtain the density of $t^2 = w$, denoted by $f_w(w)$, as

$$f_w(w) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left(\frac{1}{\nu}\right)^{\frac{1}{2}} \frac{w^{\frac{1}{2}-1}}{(1 + \frac{w}{\nu})^{\frac{\nu+1}{2}}}, \quad 0 \leq w < \infty, \quad (2.1.12)$$

for $w = t^2$, $\nu = 1, 2, \dots$ and $f_w(w) = 0$ elsewhere. Since $w = t^2$, then the part of the density for $t > 0$ is available from (2.1.12) by observing that $\frac{1}{2}w^{\frac{1}{2}-1}dw = dt$ for $t > 0$. Hence for $t > 0$ that part of the Student- t density is available from (2.1.12) as

$$f_{1t}(t) = 2 \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad 0 \leq t < \infty, \quad (2.1.13)$$

and zero elsewhere. Since (2.1.13) is symmetric, we extend it over $(-\infty, \infty)$ and so, obtain the real Student- t density, denoted by $f_t(t)$. This is stated in the next theorem.

Theorem 2.1.6. Consider a real scalar standard normal variable z , which is divided by the square root of a real chisquare variable with ν degrees of freedom divided by its

number of degrees of freedom ν , that is, $t = \frac{z}{\sqrt{\chi_{\nu}^2/\nu}}$, where z and χ_{ν}^2 are independently distributed; then t is known as the real scalar Student- t variable and its density is given by

$$f_t(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty, \quad (2.1.14)$$

for $\nu = 1, 2, \dots$

2.1a.4. The Student- t distribution in the complex domain

Let $\tilde{z} \sim \tilde{N}_1(0, 1)$ and $\tilde{y} \sim \tilde{\chi}_{\nu}^2$ in the complex domain or equivalently \tilde{y} is distributed as a real gamma with the parameters ($\alpha = \nu, \beta = 1$), and let these random variables be independently distributed. Then, we will define Student- t with ν degrees of freedom in the complex domain as follows:

$$\tilde{t} = \tilde{t}_{\nu} = \frac{|\tilde{z}|}{\sqrt{\tilde{\chi}_{\nu}^2/\nu}}, \quad |\tilde{z}| = (z_1^2 + z_2^2)^{\frac{1}{2}}, \quad \tilde{z} = z_1 + iz_2$$

with z_1, z_2 real and $i = \sqrt{-1}$. What is then the density of \tilde{t}_{ν} ? The joint density of \tilde{z} and \tilde{y} , denoted by $\tilde{f}(\tilde{y}, \tilde{z})$, is

$$\tilde{f}(\tilde{z}, \tilde{y})d\tilde{y} \wedge d\tilde{z} = \frac{1}{\pi\Gamma(\nu)} y^{\nu-1} e^{-y-|\tilde{z}|^2} d\tilde{y} \wedge d\tilde{z}.$$

Let $\tilde{z} = z_1 + iz_2$, $i = \sqrt{-1}$, where $z_1 = r \cos \theta$ and $z_2 = r \sin \theta$, $0 \leq r < \infty$, $0 \leq \theta \leq 2\pi$. Then, $dz_1 \wedge dz_2 = r dr \wedge d\theta$, and the joint density of r and \tilde{y} , denoted by $f_1(r, y)$, is the following after integrating out θ , observing that y has a real gamma density:

$$f_1(r, y)dr \wedge dy = \frac{2}{\Gamma(\nu)} y^{\nu-1} e^{-y-r^2} r dr \wedge dy.$$

Let $u = t^2 = \frac{vr^2}{y}$ and $y = w$. Then, $du \wedge dw = \frac{2vr}{w} dr \wedge dy$ and so, $r dr \wedge dy = \frac{w}{2v} du \wedge dw$. Letting the joint density of u and w be denoted by $f_2(u, w)$, we have

$$f_2(u, w) = \frac{1}{v\Gamma(\nu)} w^{\nu} e^{-(w+\frac{uw}{v})}$$

and the marginal density of u , denoted by $g(u)$, is as follows:

$$g(u) = \int_{w=0}^{\infty} f_2(u, w) dw = \int_0^{\infty} \frac{w^{\nu}}{\Gamma(\nu+1)} e^{-w(1+\frac{u}{v})} dw = \left(1 + \frac{u}{v}\right)^{-(\nu+1)}$$

for $0 \leq u < \infty$, $\nu = 1, 2, \dots$, $u = t^2$ and zero elsewhere. Thus the part of the density of t , for $t > 0$ denoted by $f_{1t}(t)$ is as follows, observing that $du = 2tdt$ for $t > 0$:

$$f_{1t}(t) = 2t \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)}, \quad 0 \leq t < \infty, \quad \nu = 1, 2, \dots \quad (2.1a.6)$$

Extending this density over the real line, we obtain the following density of \tilde{t} in the complex case:

$$\tilde{f}_\nu(t) = |t| \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)}, \quad -\infty < t < \infty, \quad \nu = 1, \dots \quad (2.1a.7)$$

Thus, the following result:

Theorem 2.1a.6. *Let $\tilde{z} \sim \tilde{N}_1(0, 1)$, $\tilde{y} \sim \tilde{\chi}_\nu^2$, a scalar chisquare in the complex domain and let \tilde{z} and \tilde{y} in the complex domain be independently distributed. Consider the real variable $t = t_\nu = \frac{|\tilde{z}|}{\sqrt{\tilde{y}/\nu}}$. Then this t will be called a Student- t with ν degrees of freedom in the complex domain and its density is given by (2.1a.7).*

2.1.7. The Cauchy distribution in the real domain

We have already seen a ratio distribution in Sect. 2.1.3, namely the real type-2 beta distribution and, as particular cases, the real F-distribution and the real t^2 distribution. We now consider a ratio of two independently distributed real standard normal variables. Let $z_1 \sim N_1(0, 1)$ and $z_2 \sim N_1(0, 1)$ be independently distributed. The joint density of z_1 and z_2 , denoted by $f(z_1, z_2)$, is given by

$$f(z_1, z_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)}, \quad -\infty < z_j < \infty, \quad j = 1, 2.$$

Consider the quadrant $z_1 > 0, z_2 > 0$ and the transformation $u = \frac{z_1}{z_2}$, $v = z_2$. Then $dz_1 \wedge dz_2 = vdu \wedge dv$, see Sect. 2.1.3. Note that $u > 0$ covers the quadrants $z_1 > 0, z_2 > 0$ and $z_1 < 0, z_2 < 0$. The part of the density of u in the quadrant $u > 0, v > 0$, denoted as $g(u, v)$, is given by

$$g(u, v) = \frac{v}{2\pi} e^{-\frac{1}{2}v^2(1+u^2)}$$

and that part of the marginal density of u , denoted by $g_1(u)$, is

$$g_1(u) = \frac{1}{2\pi} \int_0^\infty v e^{-v^2 \frac{(1+u^2)}{2}} dv = \frac{1}{2\pi(1+u^2)}.$$

The other two quadrants $z_1 > 0, z_2 < 0$ and $z_1 < 0, z_2 > 0$, which correspond to $u < 0$, will yield the same form as above. Accordingly, the density of the ratio $u = \frac{z_1}{z_2}$, known as the real Cauchy density, is as specified in the next theorem.

Theorem 2.1.7. *Consider the independently distributed real standard normal variables $z_1 \sim N_1(0, 1)$ and $z_2 \sim N_1(0, 1)$. Then the ratio $u = \frac{z_1}{z_2}$ has the real Cauchy distribution having the following density:*

$$g_u(u) = \frac{1}{\pi(1+u^2)}, \quad -\infty < u < \infty. \quad (2.1.15)$$

By integrating out in each interval $(-\infty, 0)$ and $(0, \infty)$, with the help of a type-2 beta integral, it can be established that (2.1.15) is indeed a density. Since $g_u(u)$ is symmetric, $Pr\{u \leq 0\} = Pr\{u \geq 0\} = \frac{1}{2}$, and one could posit that the mean value of u may be zero. However, observe that

$$\int_0^{\infty} \frac{u}{1+u^2} du = \frac{1}{2} \ln(1+u^2) \Big|_0^{\infty} \rightarrow \infty.$$

Thus, $E(u)$, the mean value of a real Cauchy random variable, does not exist, which implies that the higher moments do not exist either.

Exercises 2.1

2.1.1. Consider someone throwing dart at a board to hit a point on the board. Taking this target point as the origin, consider a rectangular coordinate system. If (x, y) is a point of hit, then compute the densities of x and y under the following assumptions: (1): There is no bias in the horizontal and vertical directions or x and y are independently distributed; (2): The joint density is a function of the distance from the origin $\sqrt{x^2 + y^2}$. That is, if $f_1(x)$ and $f_2(y)$ are the densities of x and y then it is given that $f_1(x)f_2(y) = g(\sqrt{x^2 + y^2})$ where f_1, f_2, g are unknown functions. Show that f_1 and f_2 are identical and real normal densities.

2.1.2. Generalize Exercise 2.1.1 to 3-dimensional Euclidean space or

$$g(\sqrt{x^2 + y^2 + z^2}) = f_1(x)f_2(y)f_3(z).$$

2.1.3. Generalize Exercise 2.1.2 to k -space, $k \geq 3$.

2.1.4. Let $f(x)$ be an arbitrary density. Then Shannon's measure of entropy or uncertainty is $S = -k \int_x f(x) \ln f(x) dx$ where k is a constant. Optimize S , subject to the conditions

(a): $\int_{-\infty}^{\infty} f(x)dx = 1$; (b): Condition in (a) plus $\int_{-\infty}^{\infty} xf(x)dx =$ given quantity; (c): The conditions in (b) plus $\int_{-\infty}^{\infty} x^2 f(x)dx =$ a given quantity. Show that under (a), f is a uniform density; under (b), f is an exponential density and under (c), f is a Gaussian density. Hint: Use Calculus of Variation.

2.1.5. Let the error of measurement ϵ satisfy the following conditions: (1) $\epsilon = \epsilon_1 + \epsilon_2 + \dots$ or it is a sum of infinitely many infinitesimal contributions ϵ_j 's where the ϵ_j 's are independently distributed. (2): Suppose that ϵ_j can only take two values δ with probability $\frac{1}{2}$ and $-\delta$ with probability $\frac{1}{2}$ for all j . (3): $\text{Var}(\epsilon) = \sigma^2 < \infty$. Then show that this error density is real Gaussian. Hint: Use mgf. [This is Gauss' derivation of the normal law and hence it is called the error curve or Gaussian density also.]

2.1.6. The pathway model of Mathai (2005) has the following form in the case of real positive scalar variable x :

$$f_1(x) = c_1 x^\gamma [1 - a(1 - q)x^\delta]^{-\frac{1}{1-q}}, q < 1, 0 \leq x \leq [a(1 - q)]^{-\frac{1}{\delta}},$$

for $\delta > 0, a > 0, \gamma > -1$ and $f_1(x) = 0$ elsewhere. Show that this generalized type-1 beta form changes to generalized type-2 beta form for $q > 1$,

$$f_2(x) = c_2 x^\gamma [1 + a(q - 1)x^\delta]^{-\frac{1}{q-1}}, q > 1, x \geq 0, \delta > 0, a > 0$$

and $f_2(x) = 0$ elsewhere, and for $q \rightarrow 1$, the model goes into a generalized gamma form given by

$$f_3(x) = c_3 x^\gamma e^{-ax^\delta}, a > 0, \delta > 0, x \geq 0$$

and zero elsewhere. Evaluate the normalizing constants c_1, c_2, c_3 . All models are available either from $f_1(x)$ or from $f_2(x)$ where q is the pathway parameter.

2.1.7. Make a transformation $x = e^{-t}$ in the generalized gamma model of $f_3(x)$ of Exercise 2.1.6. Show that an extreme-value density for t is available.

2.1.8. Consider the type-2 beta model

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1 + x)^{-(\alpha+\beta)}, x \geq 0, \Re(\alpha) > 0, \Re(\beta) > 0$$

and zero elsewhere. Make the transformation $x = e^y$ and then show that y has a generalized logistic distribution and as a particular case there one gets the logistic density.

2.1.9. Show that for $0 \leq x < \infty, \beta > 0, f(x) = c[1 + e^{\alpha+\beta x}]^{-1}$ is a density, which is known as Fermi-Dirac density. Evaluate the normalizing constant c .

2.1.10. Let $f(x) = c[e^{\alpha+\beta x} - 1]^{-1}$ for $0 \leq x < \infty$, $\beta > 0$. Show that $f(x)$ is a density, known as Bose-Einstein density. Evaluate the normalizing constant c .

2.1.11. Evaluate the incomplete gamma integral $\gamma(\alpha; b) = \int_0^b x^{\alpha-1} e^{-x} dx$ and show that it can be written in terms of the confluent hypergeometric series

$${}_1F_1(\beta; \delta; y) = \sum_{k=0}^{\infty} \frac{(\beta)_k}{(\delta)_k} \frac{y^k}{k!},$$

$(\alpha)_k = \alpha(\alpha + 1) \cdots (\alpha + k - 1)$, $\alpha \neq 0$, $(\alpha)_0 = 1$ is the Pochhammer symbol. Evaluate the normalizing constant c if $f(x) = cx^{\alpha-1} e^{-x}$, $0 \leq x \leq a$, $\alpha > 0$ and zero elsewhere, is a density.

2.1.12. Evaluate the incomplete beta integral $b(\alpha; \beta; b) = \int_0^b x^{\alpha-1} (1-x)^{\beta-1}$, $\alpha > 0$, $\beta > 0$, $0 \leq b \leq 1$. Show that it is available in terms of a Gauss' hypergeometric series of the form ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$, $|z| < 1$.

2.1.13. For the pathway model in Exercise 2.1.6 compute the reliability function $Pr\{x \geq t\}$ when $\gamma = 0$ for all the cases $q < 1$, $q > 1$, $q \rightarrow 1$.

2.1.14. *Weibull density:* In the generalized gamma density $f(x) = cx^{\gamma-1} e^{-ax^\delta}$, $x \geq 0$, $\gamma > 0$, $a > 0$, $\delta > 0$ and zero elsewhere, if $\delta = \gamma$ then $f(x)$ is called a Weibull density. For a Weibull density, evaluate the hazard function $h(t) = f(t)/Pr\{x \geq t\}$.

2.1.15. Consider a type-1 beta density $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$, $0 \leq x \leq 1$, $\alpha > 0$, $\beta > 0$ and zero elsewhere. Let $\alpha = 1$. Consider a power transformation $x = y^\delta$, $\delta > 0$. Let this model be $g(y)$. Compute the reliability function $Pr\{y \geq t\}$ and the hazard function $h(t) = g(t)/Pr\{y \geq t\}$.

2.1.16. Verify that if z is a real standard normal variable, $E(e^{tz^2}) = (1-2t)^{-1/2}$, $t < 1/2$, which is the mgf of a chi-square random variable having one degree of freedom. Owing to the uniqueness of the mgf, this result establishes that $z^2 \sim \chi_1^2$.

2.2. Quadratic Forms, Chisquaredness and Independence in the Real Domain

Let x_1, \dots, x_p be iid (independently and identically distributed) real scalar random variables distributed as $N_1(0, 1)$ and X be a $p \times 1$ vector whose components are x_1, \dots, x_p , that is, $X' = (x_1, \dots, x_p)$. Consider the real quadratic form $u_1 = X'AX$ for some $p \times p$ real constant symmetric matrix $A = A'$. Then, we have the following result:

Theorem 2.2.1. *The quadratic form $u_1 = X'AX$, $A = A'$, where the components of X are iid $N_1(0, 1)$, is distributed as a real chisquare with r , $r \leq p$, degrees of freedom if and only if A is idempotent, that is, $A = A^2$, and A of rank r .*

Proof: When $A = A'$ is real, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that $P'AP = \text{diag}(\lambda_1, \dots, \lambda_p)$, where the λ_j 's are the eigenvalues of A . Consider the transformation $X = PY$ or $Y = P'X$. Then

$$X'AX = Y'P'APY = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_p y_p^2 \quad (i)$$

where y_1, \dots, y_p are the components of Y and $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A . We have already shown in Theorem 2.1.1 that all linear functions of independent real normal variables are also real normal and hence, all the y_j 's are normally distributed. The expectation of Y is $E[Y] = E[P'X] = P'E(X) = P'O = O$ and the covariance matrix associated with Y is

$$\text{Cov}(Y) = E[Y - E(Y)][Y - E(Y)]' = E[YY'] = P'\text{Cov}(X)P = P'IP = P'P = I$$

which means that the y_j 's are real standard normal variables that are mutually independently distributed. Hence, $y_j^2 \sim \chi_1^2$ or each y_j^2 is a real chisquare with one degree of freedom each and the y_j 's are all mutually independently distributed. If $A = A^2$ and the rank of A is r , then r of the eigenvalues of A are unities and the remaining ones are equal to zero as the eigenvalues of an idempotent matrix can only be equal to zero or one, the number of ones being equal to the rank of the idempotent matrix. Then the representation in (i) becomes sum of r independently distributed real chisquares of one degree of freedom each and hence the sum is a real chisquare of r degrees of freedom. Hence, the sufficiency of the result is proved. For the necessity, we assume that $X'AX \sim \chi_r^2$ and we must prove that $A = A^2$ and A is of rank r . Note that it is assumed throughout that $A = A'$. If $X'AX$ is a real chisquare having r degrees of freedom, then the mgf of $u_1 = X'AX$ is given by $M_{u_1}(t) = (1 - 2t)^{-\frac{r}{2}}$. From the representation given in (i), the mgf's are as follows: $M_{y_j^2}(t) = (1 - 2t)^{-\frac{1}{2}} \Rightarrow M_{\lambda_j y_j^2}(t) = (1 - 2\lambda_j t)^{-\frac{1}{2}}$, $j = 1, \dots, p$, the y_j 's being independently distributed. Thus, the mgf of the right-hand side of (i) is $M_{u_1}(t) = \prod_{j=1}^p (1 - 2\lambda_j t)^{-\frac{1}{2}}$. Hence, we have

$$(1 - 2t)^{-\frac{r}{2}} = \prod_{j=1}^p (1 - 2\lambda_j t)^{-\frac{1}{2}}, \quad 1 - 2t > 0, \quad 1 - 2\lambda_j t > 0, \quad j = 1, \dots, p. \quad (ii)$$

Taking the natural logarithm of each side of (ii), expanding the terms and then comparing the coefficients of $\frac{(2t)^n}{n}$ on both sides for $n = 1, 2, \dots$, we obtain equations of the type

$$r = \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \lambda_j^2 = \sum_{j=1}^p \lambda_j^3 = \dots \quad (iii)$$

The only solution resulting from (iii) is that r of the λ_j 's are unities and the remaining ones are zeros. This result, combined with the property that $A = A'$ guarantees that A is idempotent of rank r .

Observe that the eigenvalues of a matrix being ones and zeros need not imply that the matrix is idempotent; take for instance triangular matrices whose diagonal elements are unities and zeros. However, this property combined with the symmetry assumption will guarantee that the matrix is idempotent.

Corollary 2.2.1. *If the simple random sample or the iid variables came from a real $N_1(0, \sigma^2)$ distribution, then the modification needed in Theorem 2.2.1 is that $\frac{1}{\sigma^2} X'AX \sim \chi_r^2$, $A = A'$, if and only if $A = A^2$ and A is of rank r .*

The above result, Theorem 2.2.1, coupled with another result on the independence of quadratic forms, are quite useful in the areas of Design of Experiment, Analysis of Variance and Regression Analysis, as well as in model building and hypotheses testing situations. This result on the independence of quadratic forms is stated next.

Theorem 2.2.2. *Let x_1, \dots, x_p be iid variables from a real $N_1(0, 1)$ population. Consider two real quadratic forms $u_1 = X'AX$, $A = A'$ and $u_2 = X'BX$, $B = B'$, where the components of the $p \times 1$ vector X are the x_1, \dots, x_p . Then, u_1 and u_2 are independently distributed if and only if $AB = O$.*

Proof: Let us assume that $AB = O$. Then $AB = O = O' = (AB)' = B'A' = BA$. When $AB = BA$, there exists a single orthonormal matrix P , $PP' = I$, $P'P = I$, such that both the quadratic forms are reduced to their canonical forms by the same P . Let

$$u_1 = X'AX = \lambda_1 y_1^2 + \dots + \lambda_p y_p^2 \quad (i)$$

and

$$u_2 = X'BX = \nu_1 y_1^2 + \dots + \nu_p y_p^2 \quad (ii)$$

where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A and v_1, \dots, v_p are the eigenvalues of B . Since $A = A'$, the eigenvalues λ_j 's are all real. Moreover,

$$AB = O \Rightarrow P'ABP = P'APP'BP = D_1D_2 = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_p \end{bmatrix} = O, \quad (iii)$$

which means that $\lambda_j v_j = 0$ for all $j = 1, \dots, p$. Thus, whenever a λ_j is not zero, the corresponding v_j is zero and vice versa. Accordingly, the λ_j 's and v_j 's are separated in (i) and (ii), that is, the independent components are mathematically separated and hence u_1 and u_2 are statistically independently distributed. The converse which can be stated as follows: if u_1 and u_2 are independently distributed, $A = A'$, $B = B'$ and the x_j 's are real iid $N_1(0, 1)$, then $AB = O$, is more difficult to establish. The proof which requires additional properties of matrices, will not be herein presented. Note that there are several incorrect or incomplete “proofs” in the literature. A correct derivation may be found in Mathai and Provost (1992).

When x_1, \dots, x_p are iid $N_1(0, \sigma^2)$, the above result on the independence of quadratic forms still holds since the independence is not altered by multiplying the quadratic forms by $\frac{1}{\sigma^2}$.

Example 2.2.1. Construct two 3×3 matrices A and B such that $A = A'$, $B = B'$ [both are symmetric], $A = A^2$ [A is idempotent], $AB = O$ [A and B are orthogonal to each other], and A has rank 2. Then (1): verify Theorem 2.2.1; (2): verify Theorem 2.2.2.

Solution 2.2.1. Consider the following matrices:

$$A = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Note that both A and B are symmetric, that is, $A = A'$, $B = B'$. Further, the rank of A is 2 since the first and second row vectors are linearly independent and the third row is a multiple of the first one. Note that $A^2 = A$ and $AB = O$. Now, consider the quadratic forms $u = X'AX$ and $v = X'BX$. Then $u = \frac{1}{2}x_1^2 + x_2^2 + \frac{1}{2}x_3^2 - x_1x_3 = x_2^2 + [\frac{1}{\sqrt{2}}(x_1 - x_3)]^2$. Our initial assumption is that $x_j \sim N_1(0, 1)$, $j = 1, 2, 3$ and the x_j 's are independently distributed. Let $y_1 = \frac{1}{\sqrt{2}}(x_1 - x_3)$. Then, $E[y_1] = 0$, $\text{Var}(y_1) = +\frac{1}{2}[\text{Var}(x_1) + \text{Var}(x_3)] =$

$\frac{1}{2}[1 + 1] = 1$. Since y_1 is a linear function of normal variables, y_1 is normal with the parameters $E[y_1] = 0$ and $\text{Var}(y_1) = 1$, that is, $y_1 \sim N_1(0, 1)$, and hence $y_1^2 \sim \chi_1^2$; as well, $x_2^2 \sim \chi_1^2$. Thus, $u \sim \chi_2^2$ since x_2 and y_1 are independently distributed given that the variables are separated, noting that y_1 does not involve x_2 . This verifies Theorem 2.2.1. Now, having already determined that $AB = O$, it remains to show that u and v are independently distributed where $v = X'BX = x_1^2 + x_3^2 + 2x_1x_3 = (x_1 + x_3)^2$. Let $y_2 = \frac{1}{\sqrt{2}}(x_1 + x_3) \Rightarrow y_2 \sim N_1(0, 1)$ as y_2 is a linear function of normal variables and hence normal with parameters $E[y_2] = 0$ and $\text{Var}(y_2) = 1$. On noting that v does not contain x_2 , we need only consider the parts of u and v containing x_1 and x_3 . Thus, our question reduces to: are y_1 and y_2 independently distributed? Since both y_1 and y_2 are linear functions of normal variables, both y_1 and y_2 are normal. Since the covariance between y_1 and y_2 , that is, $\text{Cov}(y_1, y_2) = \frac{1}{2}\text{Cov}(x_1 - x_3, x_1 + x_3) = \frac{1}{2}[\text{Var}(x_1) - \text{Var}(x_3)] = \frac{1}{2}[1 - 1] = 0$, the two normal variables are uncorrelated and hence, independently distributed. That is, y_1 and y_2 are independently distributed, thereby implying that u and v are also independently distributed, which verifies Theorem 2.2.2.

2.2a. Hermitian Forms, Chisquaredness and Independence in the Complex Domain

Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k$ be independently and identically distributed standard univariate Gaussian variables in the complex domain and let \tilde{X} be a $k \times 1$ vector whose components are $\tilde{x}_1, \dots, \tilde{x}_k$. Consider the Hermitian form $\tilde{X}^*A\tilde{X}$, $A = A^*$ (Hermitian) where A is a $k \times k$ constant Hermitian matrix. Then, there exists a unitary matrix Q , $QQ^* = I$, $Q^*Q = I$, such that $Q^*AQ = \text{diag}(\lambda_1, \dots, \lambda_k)$. Note that the λ_j 's are real since A is Hermitian. Consider the transformation $\tilde{X} = Q\tilde{Y}$. Then,

$$\tilde{X}^*A\tilde{X} = \lambda_1|\tilde{y}_1|^2 + \dots + \lambda_k|\tilde{y}_k|^2$$

where the \tilde{y}_j 's are iid standard normal in the complex domain, $\tilde{y}_j \sim \tilde{N}_1(0, 1)$, $j = 1, \dots, k$. Then, $\tilde{y}_j^*\tilde{y}_j = |\tilde{y}_j|^2 \sim \tilde{\chi}_1^2$, a chisquare having one degree of freedom in the complex domain or, equivalently, a real gamma random variable with the parameters ($\alpha = 1, \beta = 1$), the $|\tilde{y}_j|^2$'s being independently distributed for $j = 1, \dots, k$. Thus, we can state the following result whose proof parallels that in the real case.

Theorem 2.2a.1. *Let $\tilde{x}_1, \dots, \tilde{x}_k$ be iid $\tilde{N}_1(0, 1)$ variables in the complex domain. Consider the Hermitian form $u = \tilde{X}^*A\tilde{X}$, $A = A^*$ where \tilde{X} is a $k \times 1$ vector whose components are $\tilde{x}_1, \dots, \tilde{x}_k$. Then, u is distributed as a chisquare in the complex domain with r degrees of freedom or a real gamma with the parameters ($\alpha = r, \beta = 1$), if and only if A is of rank r and $A = A^2$.*

Theorem 2.2a.2. Let the \tilde{x}_j 's and \tilde{X} be as in Theorem 2.2a.1. Consider two Hermitian forms $u_1 = \tilde{X}^* A \tilde{X}$, $A = A^*$ and $u_2 = \tilde{X}^* B \tilde{X}$, $B = B^*$. Then, u_1 and u_2 are independently distributed if and only if $AB = O$ (null matrix).

Example 2.2a.1. Construct two 3×3 Hermitian matrices A and B , that is $A = A^*$, $B = B^*$, such that $A = A^2$ [idempotent] and is of rank 2 with $AB = O$. Then (1): verify Theorems 2.2a.1, (2): verify Theorem 2.2a.2.

Solution 2.2a.1. Consider the following matrices

$$A = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{(1+i)}{\sqrt{8}} \\ 0 & 1 & 0 \\ -\frac{(1-i)}{\sqrt{8}} & 0 & \frac{1}{2} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{2} & 0 & \frac{(1+i)}{\sqrt{8}} \\ 0 & 0 & 0 \\ \frac{(1-i)}{\sqrt{8}} & 0 & \frac{1}{2} \end{bmatrix}.$$

It can be readily verified that $A = A^*$, $B = B^*$, $A = A^2$, $AB = O$. Further, on multiplying the first row of A by $-\frac{2(1-i)}{\sqrt{8}}$, we obtain the third row, and since the third row is a multiple of the first one and the first and second rows are linearly independent, the rank of A is 2. Our initial assumption is that $\tilde{x}_j \sim \tilde{N}_1(0, 1)$, $j = 1, 2, 3$, that is, they are univariate complex Gaussian, and they are independently distributed. Then, $\tilde{x}_j^* \tilde{x}_j \sim \tilde{\chi}_1^2$, a chisquare with one degree of freedom in the complex domain or a real gamma random variable with the parameters ($\alpha = 1, \beta = 1$) for each $j = 1, 2, 3$. Let us consider the Hermitian forms $u = \tilde{X}^* A \tilde{X}$ and $v = \tilde{X}^* B \tilde{X}$, $\tilde{X}' = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$. Then

$$\begin{aligned} u &= \frac{1}{2} \tilde{x}_1^* \tilde{x}_1 - \frac{(1+i)}{\sqrt{8}} \tilde{x}_1^* \tilde{x}_3 - \frac{(1-i)}{\sqrt{8}} \tilde{x}_3^* \tilde{x}_1 + \frac{1}{2} \tilde{x}_3^* \tilde{x}_3 + \tilde{x}_2^* \tilde{x}_2 \\ &= \tilde{x}_2^* \tilde{x}_2 + \frac{1}{2} [\tilde{x}_1^* \tilde{x}_1 - 4 \frac{(1+i)}{\sqrt{8}} \tilde{x}_1^* \tilde{x}_3 + \tilde{x}_3^* \tilde{x}_3] \\ &= \tilde{\chi}_1^2 + \left[\frac{1}{\sqrt{2}} (\tilde{y}_1 - \tilde{x}_3) \right]^* \left[\frac{1}{\sqrt{2}} (\tilde{y}_1 - \tilde{x}_3) \right] \end{aligned} \quad (i)$$

where

$$\begin{aligned} \tilde{y}_1 &= 2 \frac{(1+i)}{\sqrt{8}} \tilde{x}_1 \Rightarrow E[\tilde{y}_1] = 0, \quad \text{Var}(\tilde{y}_1) = \left| 2 \frac{(1+i)}{\sqrt{8}} \right|^2 \text{Var}(\tilde{x}_1) \\ \text{Var}(\tilde{y}_1) &= E \left\{ \left[2 \frac{(1+i)}{\sqrt{8}} \right]^* \left[2 \frac{(1+i)}{\sqrt{8}} \right] \tilde{x}_1^* \tilde{x}_1 \right\} = E\{\tilde{x}_1^* \tilde{x}_1\} = \text{Var}(\tilde{x}_1) = 1. \end{aligned} \quad (ii)$$

Since \tilde{y}_1 is a linear function of \tilde{x}_1 , it is a univariate normal in the complex domain with parameters 0 and 1 or $\tilde{y}_1 \sim \tilde{N}_1(0, 1)$. The part not containing the $\tilde{\chi}_1^2$ in (i) can be written as follows:

$$\left[\frac{1}{\sqrt{2}} (\tilde{y}_1 - \tilde{x}_3) \right]^* \left[\frac{1}{\sqrt{2}} (\tilde{y}_1 - \tilde{x}_3) \right] \sim \tilde{\chi}_1^2 \quad (iii)$$

since $\tilde{y}_1 - \tilde{x}_3 \sim \tilde{N}_1(0, 2)$ as $\tilde{y}_1 - \tilde{x}_3$ is a linear function of the normal variables \tilde{y}_1 and \tilde{x}_3 . Therefore $u = \tilde{\chi}_1^2 + \tilde{\chi}_1^2 = \tilde{\chi}_2^2$, that is, a chisquare having two degrees of freedom in the complex domain or a real gamma with the parameters ($\alpha = 2, \beta = 1$). Observe that the two chisquares are independently distributed because one of them contains only \tilde{x}_2 and the other, \tilde{x}_1 and \tilde{x}_3 . This establishes (1). In order to verify (2), we first note that the Hermitian form v can be expressed as follows:

$$v = \frac{1}{2}\tilde{x}_1^*\tilde{x}_1 + \frac{(1+i)}{\sqrt{8}}\tilde{x}_1^*\tilde{x}_3 + \frac{(1-i)}{\sqrt{8}}\tilde{x}_3^*\tilde{x}_1 + \frac{1}{2}\tilde{x}_3^*\tilde{x}_3$$

which can be written in the following form by making use of steps similar to those leading to (iii):

$$v = \left[2\frac{(1+i)}{\sqrt{8}}\frac{\tilde{x}_1}{\sqrt{2}} + \frac{\tilde{x}_3}{\sqrt{2}}\right]^* \left[2\frac{(1+i)}{\sqrt{8}}\frac{\tilde{x}_1}{\sqrt{2}} + \frac{\tilde{x}_3}{\sqrt{2}}\right] = \tilde{\chi}_1^2 \quad (iv)$$

or v is a chisquare with one degree of freedom in the complex domain. Observe that \tilde{x}_2 is absent in (iv), so that we need only compare the terms containing \tilde{x}_1 and \tilde{x}_3 in (iii) and (iv). These terms are $\tilde{y}_2 = 2\frac{(1+i)}{\sqrt{8}}\tilde{x}_1 + \tilde{x}_3$ and $\tilde{y}_3 = 2\frac{(1+i)}{\sqrt{8}}\tilde{x}_1 - \tilde{x}_3$. Noting that the covariance between \tilde{y}_2 and \tilde{y}_3 is zero:

$$\text{Cov}(\tilde{y}_2, \tilde{y}_3) = \left|2\frac{(1+i)}{\sqrt{8}}\right|^2 \text{Var}(\tilde{x}_1) - \text{Var}(\tilde{x}_3) = 1 - 1 = 0,$$

and that \tilde{y}_2 and \tilde{y}_3 are linear functions of normal variables and hence normal, the fact that they are uncorrelated implies that they are independently distributed. Thus, u and v are indeed independently distributed, which establishes (2).

2.2.1. Extensions of the results in the real domain

Let $x_j \sim N_1(\mu_j, \sigma_j^2)$, $j = 1, \dots, k$, be independently distributed. Then, $\frac{x_j}{\sigma_j} \sim N_1\left(\frac{\mu_j}{\sigma_j}, 1\right)$, $\sigma_j > 0$, $j = 1, \dots, k$. Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k^2 \end{bmatrix}, \quad \Sigma^{\frac{1}{2}} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix}.$$

Then, let

$$Y = \Sigma^{-\frac{1}{2}}X = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}, \quad E[Y] = \Sigma^{-\frac{1}{2}}E[X] = \Sigma^{-\frac{1}{2}}\mu.$$

If $\mu = O$, it has already been shown that $Y'Y \sim \chi_k^2$. If $\mu \neq O$, then $Y'Y \sim \chi_k^2(\lambda)$, $\lambda = \frac{1}{2}\mu'\Sigma^{-1}\mu$. It is assumed that the noncentral chisquare distribution has already been discussed in a basic course in Statistics. It is defined for instance in Mathai and Haubold (2017a, 2017b) and will be briefly discussed in Sect. 2.3.1. If $\mu = O$, then for any $k \times k$ symmetric matrix $A = A'$, $Y'AY \sim \chi_r^2$ if and only if $A = A^2$ and A is of rank r . This result has already been established. Now, if $\mu \neq O$, then $X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y \sim \chi_k^2$ if and only if $\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}A\Sigma A\Sigma^{\frac{1}{2}} \Rightarrow A = A\Sigma A$ and $\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$ is of rank r or A is of rank r since $\Sigma > O$. Hence, we have the following result:

Theorem 2.2.3. *Let the real scalars $x_j \sim N_1(\mu_j, \sigma_j^2)$, $j = 1, \dots, k$, be independently distributed. Let*

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = \Sigma^{-\frac{1}{2}}X, E(Y) = \Sigma^{-\frac{1}{2}}\mu, \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}.$$

Then for any $k \times k$ symmetric matrix $A = A'$,

$$X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y \sim \begin{cases} \chi_r^2 & \text{if } \mu = O \\ \chi_k^2(\lambda) & \text{if } \mu \neq O, \lambda = \frac{1}{2}\mu'\Sigma^{-\frac{1}{2}}A\Sigma^{-\frac{1}{2}}\mu \end{cases}$$

if and only if $A = A\Sigma A$ and A is of rank r .

Independence is not altered if the variables are relocated. Consider two quadratic forms $X'AX$ and $X'BX$, $A = A'$, $B = B'$. Then, $X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y$ and $X'BX = Y'\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}Y$, and we have the following result:

Theorem 2.2.4. *Let x_j , X , Y , Σ be as defined in Theorem 2.2.3. Then, the quadratic forms $X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y$ and $X'BX = Y'\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}Y$, $A = A'$, $B = B'$, are independently distributed if and only if $A\Sigma B = O$.*

Let X , A and Σ be as defined in Theorem 2.2.3, Z be a standard normal vector whose components z_i , $i = 1, \dots, k$ are iid $N_1(0, 1)$, and P be an orthonormal matrix such that $P'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}P = \text{diag}(\lambda_1, \dots, \lambda_k)$; then, a general quadratic form $X'AX$ can be expressed as follows:

$$X'AX = (Z'\Sigma^{\frac{1}{2}} + \mu')A(\Sigma^{\frac{1}{2}}Z + \mu) = (Z + \Sigma^{-\frac{1}{2}}\mu)'PP'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}PP'(Z + \Sigma^{-\frac{1}{2}}\mu)$$

where $\lambda_1, \dots, \lambda_k$ are the eigenvalues of $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$. Hence, the following decomposition of the quadratic form:

$$X'AX = \lambda_1(u_1 + b_1)^2 + \dots + \lambda_k(u_k + b_k)^2, \quad (2.2.1)$$

where $\begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} = P' \Sigma^{-\frac{1}{2}} \mu$, $\begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} = P' \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}$, and hence the u_i 's are iid $N_1(0, 1)$.

Thus, $X'AX$ can be expressed as a linear combination of independently distributed non-central chisquare random variables, each having one degree of freedom, whose non-centrality parameters are respectively $b_j^2/2$, $j = 1, \dots, k$. Of course, the k chisquares will be central when $\mu = O$.

2.2a.1. Extensions of the results in the complex domain

Let the complex scalar variables $\tilde{x}_j \sim \tilde{N}_1(\tilde{\mu}_j, \sigma_j^2)$, $j = 1, \dots, k$, be independently distributed and $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_k^2)$. As well, let

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_k \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k^2 \end{bmatrix}, \quad \Sigma^{-\frac{1}{2}} \tilde{X} = \tilde{Y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_k \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_1 \\ \vdots \\ \tilde{\mu}_k \end{bmatrix}$$

where $\Sigma^{\frac{1}{2}}$ is the Hermitian positive definite square root of Σ . In this case, $\tilde{y}_j \sim \tilde{N}_1(\frac{\tilde{\mu}_j}{\sigma_j}, 1)$, $j = 1, \dots, k$ and the \tilde{y}_j 's are assumed to be independently distributed. Hence, for any Hermitian form $\tilde{X}^* A \tilde{X}$, $A = A^*$, we have $\tilde{X}^* A \tilde{X} = \tilde{Y}^* \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} \tilde{Y}$. Hence if $\tilde{\mu} = O$ (null vector), then from the previous result on chisquarredness, we have:

Theorem 2.2a.3. *Let \tilde{X} , Σ , \tilde{Y} , $\tilde{\mu}$ be as defined above. Let $u = \tilde{X}^* A \tilde{X}$, $A = A^*$ be a Hermitian form. Then $u \sim \tilde{\chi}_r^2$ in the complex domain if and only if A is of rank r , $\tilde{\mu} = O$ and $A = A \Sigma A$. [A chisquare with r degrees of freedom in the complex domain is a real gamma with parameters $(\alpha = r, \beta = 1)$.]*

If $\tilde{\mu} \neq O$, then we have a noncentral chisquare in the complex domain. A result on the independence of Hermitian forms can be obtained as well.

Theorem 2.2a.4. *Let \tilde{X} , \tilde{Y} , Σ be as defined above. Consider the Hermitian forms $u_1 = \tilde{X}^* A \tilde{X}$, $A = A^*$ and $u_2 = \tilde{X}^* B \tilde{X}$, $B = B^*$. Then u_1 and u_2 are independently distributed if and only if $A \Sigma B = O$.*

The proofs of Theorems 2.2a.3 and 2.2a.4 parallel those presented in the real case and are hence omitted.

Exercises 2.2

2.2.1. Give a proof to the second part of Theorem 2.2.2, namely, given that $X'AX$, $A = A'$ and $X'BX$, $B = B'$ are independently distributed where the components of the $p \times 1$ vector X are mutually independently distributed as real standard normal variables, then show that $AB = O$.

2.2.2. Let the real scalar $x_j \sim N_1(0, \sigma^2)$, $\sigma^2 > 0$, $j = 1, 2, \dots, k$ and be independently distributed. Let $X' = (x_1, \dots, x_k)$ or X is the $k \times 1$ vector where the elements are x_1, \dots, x_k . Then the joint density of the real scalar variables x_1, \dots, x_k , denoted by $f(X)$, is

$$f(X) = \frac{1}{(\sqrt{2\pi})^k} e^{-\frac{1}{2\sigma^2} X'X}, \quad -\infty < x_j < \infty, \quad j = 1, \dots, k.$$

Consider the quadratic form $u = X'AX$, $A = A'$ and X is as defined above. (1): Compute the mgf of u ; (2): Compute the density of u if A is of rank r and all eigenvalues of A are equal to $\lambda > 0$; (3): If the eigenvalues are $\lambda > 0$ for m of the eigenvalues and the remaining n of them are $\lambda < 0$, $m + n = r$, compute the density of u .

2.2.3. In Exercise 2.2.2 compute the density of u if (1): r_1 of the eigenvalues are λ_1 each and r_2 of the eigenvalues are λ_2 each, $r_1 + r_2 = r$. Consider all situations $\lambda_1 > 0$, $\lambda_2 > 0$ etc.

2.2.4. In Exercise 2.2.2 compute the density of u for the general case with no restrictions on the eigenvalues.

2.2.5. Let $x_j \sim N_1(0, \sigma^2)$, $j = 1, 2$ and be independently distributed. Let $X' = (x_1, x_2)$. Let $u = X'AX$ where $A = A'$. Compute the density of u if the eigenvalues of A are (1): 2 and 1, (2): 2 and -1 ; (3): Construct a real 2×2 matrix $A = A'$ where the eigenvalues are 2 and 1.

2.2.6. Show that the results on chisquaredness and independence in the real or complex domain need not hold if $A \neq A^*$, $B \neq B^*$.

2.2.7. Construct a 2×2 Hermitian matrix $A = A^*$ such that $A = A^2$ and verify Theorem 2.2a.3. Construct 2×2 Hermitian matrices A and B such that $AB = O$, and then verify Theorem 2.2a.4.

2.2.8. Let $\tilde{x}_1, \dots, \tilde{x}_m$ be a simple random sample of size m from a complex normal population $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$. Let $\tilde{y}_1, \dots, \tilde{y}_n$ be iid $\tilde{N}(\tilde{\mu}_2, \sigma_2^2)$. Let the two complex normal populations be independent. Let

$$s_1^2 = \sum_{j=1}^m (\tilde{x}_j - \bar{\tilde{x}})^* (\tilde{x}_j - \bar{\tilde{x}}) / \sigma_1^2, \quad s_2^2 = \sum_{j=1}^n (\tilde{y}_j - \bar{\tilde{y}})^* (\tilde{y}_j - \bar{\tilde{y}}) / \sigma_2^2,$$

$$s_{11}^2 = \frac{1}{\sigma_1^2} \sum_{j=1}^m (\tilde{x}_j - \tilde{\mu}_1)^* (\tilde{x}_j - \tilde{\mu}_1), \quad s_{21}^2 = \frac{1}{\sigma_2^2} \sum_{j=1}^n (\tilde{y}_j - \tilde{\mu}_2)^* (\tilde{y}_j - \tilde{\mu}_2)$$

Then, show that

$$\frac{s_{11}^2/m}{s_{21}^2/n} \sim \tilde{F}_{m,n}, \quad \frac{s_1^2/(m-1)}{s_2^2/(n-1)} \sim \tilde{F}_{m-1,n-1}$$

for $\sigma_1^2 = \sigma_2^2$.

2.2.9. In Exercise 2.2.8 show that $\frac{s_{11}^2}{s_{21}^2}$ is a type-2 beta with the parameters m and n , and $\frac{s_1^2}{s_2^2}$ is a type-2 beta with the parameters $m-1$ and $n-1$ for $\sigma_1^2 = \sigma_2^2$.

2.2.10. In Exercise 2.2.8 if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then show that

$$\frac{1}{\sigma^2} \left[\sum_{j=1}^m (\tilde{x}_j - \bar{\tilde{x}})^* (\tilde{x}_j - \bar{\tilde{x}}) + \sum_{j=1}^n (\tilde{y}_j - \bar{\tilde{y}})^* (\tilde{y}_j - \bar{\tilde{y}}) \right] \sim \tilde{\chi}_{m+n-2}^2.$$

2.2.11. In Exercise 2.2.10 if $\bar{\tilde{x}}$ and $\bar{\tilde{y}}$ are replaced by $\tilde{\mu}_1$ and $\tilde{\mu}_2$ respectively then show that the degrees of freedom of the chisquare is $m+n$.

2.2.12. Derive the representation of the general quadratic form $X'AX$ given in (2.2.1).

2.3. Simple Random Samples from Real Populations and Sampling Distributions

For practical applications, an important result is that on the independence of the sample mean and sample variance when the sample comes from a normal (Gaussian) population. Let x_1, \dots, x_n be a simple random sample of size n from a real $N_1(\mu_1, \sigma_1^2)$ or, equivalently, x_1, \dots, x_n are iid $N_1(\mu_1, \sigma_1^2)$. Recall that we have established that any linear function $L'X = X'L$, $L' = (a_1, \dots, a_n)$, $X' = (x_1, \dots, x_n)$ remains normally distributed (Theorem 2.1.1). Now, consider two linear forms $y_1 = L_1'X$, $y_2 = L_2'X$, with $L_1' = (a_1, \dots, a_n)$, $L_2' = (b_1, \dots, b_n)$ where a_1, \dots, a_n , b_1, \dots, b_n are real scalar constants. Let us examine the conditions that are required for assessing the independence of

the linear forms y_1 and y_2 . Since x_1, \dots, x_n are iid, we can determine the joint mgf of x_1, \dots, x_n . We take a $n \times 1$ parameter vector T , $T' = (t_1, \dots, t_n)$ where the t_j 's are scalar parameters. Then, by definition, the joint mgf is given by

$$E[e^{T'X}] = \prod_{j=1}^n M_{x_j}(t_j) = \prod_{j=1}^n e^{t_j\mu_1 + \frac{1}{2}t_j^2\sigma_1^2} = e^{\mu_1 T'J + \frac{\sigma_1^2}{2} T'T} \quad (2.3.1)$$

since the x_j 's are iid, $J' = (1, \dots, 1)$. Since every linear function of x_1, \dots, x_n is a univariate normal, we have $y_1 \sim N_1(\mu_1 L_1' J, \frac{\sigma_1^2 t_1^2}{2} L_1' L_1)$ and hence the mgf of y_1 , taking t_1 as the parameter for the mgf, is $M_{y_1}(t_1) = e^{t_1\mu_1 L_1' J + \frac{\sigma_1^2 t_1^2}{2} L_1' L_1}$. Now, let us consider the joint mgf of y_1 and y_2 taking t_1 and t_2 as the respective parameters. Let the joint mgf be denoted by $M_{y_1, y_2}(t_1, t_2)$. Then,

$$\begin{aligned} M_{y_1, y_2}(t_1, t_2) &= E[e^{t_1 y_1 + t_2 y_2}] = E[e^{(t_1 L_1' + t_2 L_2')X}] \\ &= e^{\mu_1 (t_1 L_1' + t_2 L_2')J + \frac{\sigma_1^2}{2} (t_1 L_1 + t_2 L_2)'(t_1 L_1 + t_2 L_2)} \\ &= e^{\mu_1 (L_1' + L_2')J + \frac{\sigma_1^2}{2} (t_1^2 L_1' L_1 + t_2^2 L_2' L_2 + 2t_1 t_2 L_1' L_2)} \\ &= M_{y_1}(t_1) M_{y_2}(t_2) e^{\sigma_1^2 t_1 t_2 L_1' L_2}. \end{aligned}$$

Hence, the last factor on the right-hand side has to vanish for y_1 and y_2 to be independently distributed, and this can happen if and only if $L_1' L_2 = L_2' L_1 = 0$ since t_1 and t_2 are arbitrary. Thus, we have the following result:

Theorem 2.3.1. *Let x_1, \dots, x_n be iid $N_1(\mu_1, \sigma_1^2)$. Let $y_1 = L_1' X$ and $y_2 = L_2' X$ where $X' = (x_1, \dots, x_n)$, $L_1' = (a_1, \dots, a_n)$ and $L_2' = (b_1, \dots, b_n)$, the a_j 's and b_j 's being scalar constants. Then, y_1 and y_2 are independently distributed if and only if $L_1' L_2 = L_2' L_1 = 0$.*

Example 2.3.1. Let x_1, x_2, x_3, x_4 be a simple random sample of size 4 from a real normal population $N_1(\mu = 1, \sigma^2 = 2)$. Consider the following statistics: (1): u_1, v_1, w_1 , (2): u_2, v_2, w_2 . Check for independence of various statistics in (1): and (2): where

$$\begin{aligned} u_1 &= \bar{x} = \frac{1}{4}(x_1 + x_2 + x_3 + x_4), \quad v_1 = 2x_1 - 3x_2 + x_3 + x_4, \quad w_1 = x_1 - x_2 + x_3 - x_4; \\ u_2 &= \bar{x} = \frac{1}{4}(x_1 + x_2 + x_3 + x_4), \quad v_2 = x_1 - x_2 + x_3 - x_4, \quad w_2 = x_1 - x_2 - x_3 + x_4. \end{aligned}$$

Solution 2.3.1. Let $X' = (x_1, x_2, x_3, x_4)$ and let the coefficient vectors in (1) be denoted by L_1, L_2, L_3 and those in (2) be denoted by M_1, M_2, M_3 . Thus they are as follows :

$$L_1 = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, L_2 = \begin{bmatrix} 2 \\ -3 \\ 1 \\ 1 \end{bmatrix}, L_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \Rightarrow L_1' L_2 = \frac{1}{4}, L_1' L_3 = 0, L_2' L_3 = 5.$$

This means that u_1 and w_1 are independently distributed and that the other pairs are not independently distributed. The coefficient vectors in (2) are

$$M_1 = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, M_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, M_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \Rightarrow M_1' M_2 = 0, M_1' M_3 = 0, M_2' M_3 = 0.$$

This means that all the pairs are independently distributed, that is, u_2, v_2 and w_2 are mutually independently distributed.

We can extend Theorem 2.3.1 to sets of linear functions. Let $Y_1 = AX$ and $Y_2 = BX$ where A of dimension $m_1 \times n$, $m_1 \leq n$ and B of dimension $m_2 \times n$, $m_2 \leq n$ are constant matrices and $X' = (x_1, \dots, x_n)$ where the x_j 's are iid $N_1(\mu_1, \sigma_1^2)$. Let the parameter vectors T_1 and T_2 be of dimensions $m_1 \times 1$ and $m_2 \times 1$, respectively. Then, the mgf of Y_1 is $M_{Y_1}(T_1) = E[e^{T_1' Y_1}] = E[e^{T_1' A_1 X}]$, which can be evaluated by integration over the joint density of x_1, \dots, x_n , individually, or over the vector $X' = (x_1, \dots, x_n)$ with $E[X'] = [\mu_1, \mu_1, \dots, \mu_1] = \mu_1 [1, 1, \dots, 1] = \mu_1 J'$, $J' = [1, \dots, 1] \Rightarrow E[Y_1] = \mu_1 A_1 J$. The mgf of Y_1 is then

$$M_{Y_1}(T_1) = E[e^{\mu_1 T_1' A_1 J + T_1' A_1 [X - E(X)]}] = E[e^{\mu_1 T_1' A_1 J + T_1' A_1 Z}], Z = X - E(X), \quad (i)$$

and the exponent in the expected value, not containing μ_1 , simplifies to

$$-\frac{1}{2\sigma_1^2} \{Z'Z - 2\sigma_1^2 T_1' A_1 Z\} = -\frac{1}{2\sigma_1^2} \{(Z' - \sigma_1 T_1' A_1)(Z - \sigma_1 A_1' T_1) - \sigma_1^2 T_1' A_1 A_1' T_1\}.$$

Integration over Z or individually over the elements of Z , that is, z_1, \dots, z_n , yields 1 since the total probability is 1, which leaves the factor not containing Z . Thus,

$$M_{Y_1}(T_1) = e^{\mu_1 T_1' A_1 J + \frac{1}{2} T_1' A_1 A_1' T_1}, \quad (ii)$$

and similarly,

$$M_{Y_2}(T_2) = e^{\mu_1 T_2' A_2 J + \frac{1}{2} T_2' A_2 A_2' T_2}. \quad (iii)$$

The joint mgf of Y_1 and Y_2 is then

$$\begin{aligned} M_{Y_1, Y_2}(T_1, T_2) &= e^{\mu_1(T_1' A_1 J + T_2' A_2 J) + \frac{1}{2}(T_1' A_1 + T_2' A_2)(T_1' A_1 + T_2' A_2)'} \\ &= M_{Y_1}(T_1) M_{Y_2}(T_2) e^{T_1' A_1 A_2' T_2}. \end{aligned} \quad (iv)$$

Accordingly, Y_1 and Y_2 will be independent if and only if $A_1 A_2' = O \Rightarrow A_2 A_1' = O$ since T_1 and T_2 are arbitrary parameter vectors, the two null matrices having different orders. Then, we have

Theorem 2.3.2. Let $Y_1 = A_1 X$ and $Y_2 = A_2 X$, with $X' = (x_1, \dots, x_n)$, the x_j 's being iid $N_1(\mu_1, \sigma_1^2)$, $j = 1, \dots, n$, be two sets of linear forms where A_1 is $m_1 \times n$ and A_2 is $m_2 \times n$, $m_1 \leq n$, $m_2 \leq n$, are constant matrices. Then, Y_1 and Y_2 are independently distributed if and only if $A_1 A_2' = O$ or $A_2 A_1' = O$.

Example 2.3.2. Consider a simple random sample of size 4 from a real scalar normal population $N_1(\mu_1 = 0, \sigma_1^2 = 4)$. Let $X' = (x_1, x_2, x_3, x_4)$. Verify whether the sets of linear functions $U = A_1 X$, $V = A_2 X$, $W = A_3 X$ are pairwise independent, where

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & -1 & 1 & 3 \\ 1 & 2 & -1 & -2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}.$$

Solution 2.3.2. Taking the products, we have $A_1 A_2' \neq O$, $A_1 A_3' = O$, $A_2 A_3' \neq O$. Hence, the pair U and W are independently distributed and other pairs are not.

We can apply Theorems 2.3.1 and 2.3.2 to prove several results involving sample statistics. For instance, let x_1, \dots, x_n be iid $N_1(\mu_1, \sigma_1^2)$ or a simple random sample of size n from a real $N_1(\mu_1, \sigma_1^2)$ and $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$. Consider the vectors

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \vdots \\ \mu_1 \end{bmatrix}, \quad \bar{X} = \begin{bmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix}.$$

Note that when the x_j 's are iid $N_1(\mu_1, \sigma_1^2)$, $x_j - \mu_1 \sim N_1(0, \sigma_1^2)$, and that since $X - \bar{X} = (X - \mu) - (\bar{X} - \mu)$, we may take x_j 's as coming from $N_1(0, \sigma_1^2)$ for all operations involving (X, \bar{X}) . Moreover, $\bar{x} = \frac{1}{n} J' X$, $J' = (1, 1, \dots, 1)$ where J is a $n \times 1$ vector of unities. Then,

$$X - \bar{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} = \begin{bmatrix} x_1 - \frac{1}{n}J'X \\ x_2 - \frac{1}{n}J'X \\ \vdots \\ x_n - \frac{1}{n}J'X \end{bmatrix} = (I - \frac{1}{n}JJ')X, \quad (i)$$

and on letting $A = \frac{1}{n}JJ'$, we have

$$A = A^2, \quad I - A = (I - A)^2, \quad A(I - A) = O. \quad (ii)$$

Also note that

$$(X - \bar{X})'(X - \bar{X}) = \sum_{j=1}^n (x_j - \bar{x})^2 \text{ and } s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \quad (iii)$$

where s^2 is the sample variance and $\frac{1}{n}J'X = \bar{x}$ is the sample mean. Now, observe that in light of Theorem 2.3.2, $Y_1 = (I - A)X$ and $Y_2 = AX$ are independently distributed, which implies that $X - \bar{X}$ and \bar{X} are independently distributed. But \bar{X} contains only $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ and hence $X - \bar{X}$ and \bar{x} are independently distributed. We now will make use of the following result: If w_1 and w_2 are independently distributed real scalar random variables, then the pairs (w_1, w_2^2) , (w_1^2, w_2) , (w_1^2, w_2^2) are independently distributed when w_1 and w_2 are real scalar random variables; the converses need not be true. For example, w_1^2 and w_2^2 being independently distributed need not imply the independence of w_1 and w_2 . If w_1 and w_2 are real vectors or matrices and if w_1 and w_2 are independently distributed then the following pairs are also independently distributed wherever the quantities are defined: $(w_1, w_2 w_2')$, $(w_1, w_2' w_2)$, $(w_1 w_1', w_2)$, $(w_1' w_1, w_2)$, $(w_1' w_1, w_2' w_2)$. It then follows from (iii) that \bar{x} and $(X - \bar{X})'(X - \bar{X}) = \sum_{j=1}^n (x_j - \bar{x})^2$ are independently distributed. Hence, the following result:

Theorem 2.3.3. *Let x_1, \dots, x_n be iid $N_1(\mu_1, \sigma_1^2)$ or a simple random sample of size n from a univariate real normal population $N_1(\mu_1, \sigma_1^2)$. Let $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ be the sample mean and $s^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2$ be the sample variance. Then \bar{x} and s^2 are independently distributed.*

This result has several corollaries. When x_1, \dots, x_n are iid $N_1(\mu_1, \sigma_1^2)$, then the sample sum of products, which is also referred to as the corrected sample sum of products (corrected in the sense that \bar{x} is subtracted), is given by

$$\sum_{j=1}^n (x_j - \bar{x})^2 = X'(I - A)X, \quad A = \frac{1}{n} \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

where both A and $I - A$ are idempotent. In this case, $\text{tr}(A) = \frac{1}{n}(1 + \dots + 1) = 1$ and $\text{tr}(I - A) = n - 1$ and hence, the ranks of A and $I - A$ are 1 and $n - 1$, respectively. When a matrix is idempotent, its eigenvalues are either zero or one, the number of ones corresponding to its rank. As has already been pointed out, when X and \bar{X} are involved, it can be equivalently assumed that the sample is coming from a $N_1(0, \sigma_1^2)$ population. Hence

$$\frac{ns^2}{\sigma_1^2} = \frac{1}{\sigma_1^2} \sum_{j=1}^n (x_j - \bar{x})^2 \sim \chi_{n-1}^2 \quad (2.3.2)$$

is a real chisquare with $n - 1$ (the rank of the idempotent matrix of the quadratic form) degrees of freedom as per Theorem 2.2.1. Observe that when the sample comes from a real $N_1(\mu_1, \sigma_1^2)$ distribution, we have $\bar{x} \sim N_1(\mu_1, \frac{\sigma_1^2}{n})$ so that $z = \frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma_1} \sim N_1(0, 1)$ or z is a real standard normal, and that $\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi_{n-1}^2$ where $s_1^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}$. Recall that \bar{x} and s^2 are independently distributed. Hence, the ratio

$$\frac{z}{s_1/\sigma_1} \sim t_{n-1}$$

has a real Student-t distribution with $n - 1$ degrees of freedom, where $z = \frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma_1}$ and $\frac{z}{s_1/\sigma_1} = \frac{\sqrt{n}(\bar{x} - \mu_1)}{s_1}$. Hence, we have the following result:

Theorem 2.3.4. *Let x_1, \dots, x_n be iid $N_1(\mu_1, \sigma_1^2)$. Let $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ and $s_1^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n-1}$. Then,*

$$\frac{\sqrt{n}(\bar{x} - \mu_1)}{s_1} \sim t_{n-1} \quad (2.3.3)$$

where t_{n-1} is a real Student-t with $n - 1$ degrees of freedom.

It should also be noted that when x_1, \dots, x_n are iid $N_1(\mu_1, \sigma_1^2)$, then

$$\frac{\sum_{j=1}^n (x_j - \mu_1)^2}{\sigma_1^2} \sim \chi_n^2, \quad \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{\sigma_1^2} \sim \chi_{n-1}^2, \quad \frac{\sqrt{n}(\bar{x} - \mu_1)^2}{\sigma_1^2} \sim \chi_1^2,$$

wherefrom the following decomposition is obtained:

$$\frac{1}{\sigma_1^2} \sum_{j=1}^n (x_j - \mu_1)^2 = \frac{1}{\sigma_1^2} \left[\sum_{j=1}^n (x_j - \bar{x})^2 + n(\bar{x} - \mu_1)^2 \right] \Rightarrow \chi_n^2 = \chi_{n-1}^2 + \chi_1^2, \quad (2.3.4)$$

the two chisquare random variables on the right-hand side of the last equation being independently distributed.

2.3a. Simple Random Samples from a Complex Gaussian Population

The definition of a simple random sample from any population remains the same as in the real case. A set of complex scalar random variables $\tilde{x}_1, \dots, \tilde{x}_n$, which are iid as $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$ is called a simple random sample from this complex Gaussian population. Let \tilde{X} be the $n \times 1$ vector whose components are these sample variables, $\bar{\tilde{x}} = \frac{1}{n}(\tilde{x}_1 + \dots + \tilde{x}_n)$ denote the sample average, and $\bar{\tilde{X}}' = (\bar{\tilde{x}}, \dots, \bar{\tilde{x}})$ be the $1 \times n$ vector of sample means; then \tilde{X} , $\bar{\tilde{X}}$ and the sample sum of products matrix \tilde{s} are respectively,

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}, \quad \bar{\tilde{X}} = \begin{bmatrix} \bar{\tilde{x}} \\ \vdots \\ \bar{\tilde{x}} \end{bmatrix} \quad \text{and} \quad \tilde{s} = (\tilde{X} - \bar{\tilde{X}})^*(\tilde{X} - \bar{\tilde{X}}).$$

These quantities can be simplified as follows with the help of the vector of unities $J' = (1, 1, \dots, 1)$: $\bar{\tilde{x}} = \frac{1}{n}J'\tilde{X}$, $\tilde{X} - \bar{\tilde{X}} = [I - \frac{1}{n}JJ']\tilde{X}$, $\bar{\tilde{X}} = \frac{1}{n}JJ'\tilde{X}$, $\tilde{s} = \tilde{X}^*[I - \frac{1}{n}JJ']\tilde{X}$. Consider the Hermitian form

$$\tilde{X}^*[I - \frac{1}{n}JJ']\tilde{X} = \sum_{j=1}^n (\tilde{x}_j - \bar{\tilde{x}})^*(\tilde{x}_j - \bar{\tilde{x}}) = n\tilde{s}^2$$

where \tilde{s}^2 is the sample variance in the complex scalar case, given a simple random sample of size n .

Consider the linear forms $\tilde{u}_1 = L_1^*\tilde{X} = \bar{a}_1\tilde{x}_1 + \dots + \bar{a}_n\tilde{x}_n$ and $\tilde{u}_2 = L_2^*\tilde{X} = \bar{b}_1\tilde{x}_1 + \dots + \bar{b}_n\tilde{x}_n$ where the a_j 's and b_j 's are scalar constants that may be real or complex, \bar{a}_j and \bar{b}_j denoting the complex conjugates of a_j and b_j , respectively.

$$E[\tilde{X}]' = (\tilde{\mu}_1)[1, 1, \dots, 1] = (\tilde{\mu}_1)J', \quad J' = [1, \dots, 1],$$

since the \tilde{x}_j 's are iid $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$, $j = 1, \dots, n$. The mgf of \tilde{u}_1 and \tilde{u}_2 , denoted by $M_{\tilde{u}_j}(\tilde{t}_j)$, $j = 1, 2$ and the joint mgf of \tilde{u}_1 and \tilde{u}_2 , denoted by $M_{\tilde{u}_1, \tilde{u}_2}(\tilde{t}_1, \tilde{t}_2)$ are the following, where $\Re(\cdot)$ denotes the real part of (\cdot) :

$$\begin{aligned} M_{\tilde{u}_1}(\tilde{t}_1) &= E[e^{\Re(\tilde{t}_1^* \tilde{u}_1)}] = E[e^{\Re(\tilde{t}_1^* L_1^* \tilde{X})}] \\ &= e^{\Re(\tilde{\mu}_1 \tilde{t}_1^* L_1^* J)} E[e^{\Re(\tilde{t}_1^* L_1^* (\tilde{X} - E(\tilde{X})))}] = e^{\Re(\tilde{\mu}_1 \tilde{t}_1^* L_1^* J)} e^{\frac{\sigma_1^2}{4} \tilde{t}_1^* L_1^* L_1 \tilde{t}_1} \end{aligned} \quad (i)$$

$$M_{\tilde{u}_2}(\tilde{t}_2) = e^{\Re(\tilde{\mu}_1 \tilde{t}_2^* L_2^* J) + \frac{\sigma_1^2}{4} \tilde{t}_2^* L_2^* L_2 \tilde{t}_2} \quad (ii)$$

$$M_{\tilde{u}_1, \tilde{u}_2}(\tilde{t}_1, \tilde{t}_2) = M_{\tilde{u}_1}(\tilde{t}_1) M_{\tilde{u}_2}(\tilde{t}_2) e^{\frac{\sigma_1^2}{2} \tilde{t}_1^* L_1^* L_2 \tilde{t}_2}. \quad (iii)$$

Consequently, \tilde{u}_1 and \tilde{u}_2 are independently distributed if and only if the exponential part is 1 or equivalently $\tilde{t}_1^* L_1^* L_2 \tilde{t}_2 = 0$. Since \tilde{t}_1 and \tilde{t}_2 are arbitrary, this means $L_1^* L_2 = 0 \Rightarrow L_2^* L_1 = 0$. Then we have the following result:

Theorem 2.3a.1. *Let $\tilde{x}_1, \dots, \tilde{x}_n$ be a simple random sample of size n from a univariate complex Gaussian population $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$. Consider the linear forms $\tilde{u}_1 = L_1^* \tilde{X}$ and $\tilde{u}_2 = L_2^* \tilde{X}$ where L_1, L_2 and \tilde{X} are the previously defined $n \times 1$ vectors, and a star denotes the conjugate transpose. Then, \tilde{u}_1 and \tilde{u}_2 are independently distributed if and only if $L_1^* L_2 = 0$.*

Example 2.3a.1. Let \tilde{x}_j , $j = 1, 2, 3, 4$ be iid univariate complex normal $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$. Consider the linear forms

$$\begin{aligned} \tilde{u}_1 &= L_1^* \tilde{X} = (1 + i)\tilde{x}_1 + 2i\tilde{x}_2 - (1 - i)\tilde{x}_3 + 2\tilde{x}_4 \\ \tilde{u}_2 &= L_2^* \tilde{X} = (1 + i)\tilde{x}_1 + (2 + 3i)\tilde{x}_2 - (1 - i)\tilde{x}_3 - i\tilde{x}_4 \\ \tilde{u}_3 &= L_3^* \tilde{X} = -(1 + i)\tilde{x}_1 + i\tilde{x}_2 + (1 - i)\tilde{x}_3 + \tilde{x}_4. \end{aligned}$$

Verify whether the three linear forms are pairwise independent.

Solution 2.3a.1. With the usual notations, the coefficient vectors are as follows:

$$\begin{aligned} L_1^* &= [1 + i, 2i, -1 + i, 2] \Rightarrow L_1' = [1 - i, -2i, -1 - i, 2] \\ L_2^* &= [1 + i, 2 + 3i, 1 - i, -i] \Rightarrow L_2' = [1 - i, 2 - 3i, 1 + i, i] \\ L_3^* &= [-(1 + i), i, 1 - i, 1] \Rightarrow L_3' = [-(1 - i), -i, 1 + i, 1]. \end{aligned}$$

Taking the products we have $L_1^* L_2 = 6 + 6i \neq 0$, $L_1^* L_3 = 0$, $L_2^* L_3 = 3 - 3i \neq 0$. Hence, only \tilde{u}_1 and \tilde{u}_3 are independently distributed.

We can extend the result stated in Theorem 2.3a.1 to sets of linear forms. Let $\tilde{U}_1 = A_1 \tilde{X}$ and $\tilde{U}_2 = A_2 \tilde{X}$ where A_1 and A_2 are constant matrices that may or may not be in

the complex domain, A_1 is $m_1 \times n$ and A_2 is $m_2 \times n$, with $m_1 \leq n$, $m_2 \leq n$. As was previously the case, $\tilde{X}' = (\tilde{x}_1, \dots, \tilde{x}_n)$, \tilde{x}_j , $j = 1, \dots, n$, are iid $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$. Let \tilde{T}_1 and \tilde{T}_2 be parameter vectors of orders $m_1 \times 1$ and $m_2 \times 1$, respectively. Then, on following the steps leading to (iii), the mgf of \tilde{U}_1 and \tilde{U}_2 and their joint mgf are obtained as follows:

$$M_{\tilde{U}_1}(\tilde{T}_1) = E[e^{\Re(\tilde{T}_1^* A_1 \tilde{X})}] = e^{\Re(\tilde{\mu}_1 \tilde{T}_1^* A_1 J) + \frac{\sigma_1^2}{4} \tilde{T}_1^* A_1 A_1^* \tilde{T}_1} \quad (iv)$$

$$M_{\tilde{U}_2}(\tilde{T}_2) = e^{\Re(\tilde{\mu}_1 \tilde{T}_2^* A_2 J) + \frac{\sigma_1^2}{4} \tilde{T}_2^* A_2 A_2^* \tilde{T}_2} \quad (v)$$

$$M_{\tilde{U}_1, \tilde{U}_2}(\tilde{T}_1, \tilde{T}_2) = M_{\tilde{U}_1}(\tilde{T}_1) M_{\tilde{U}_2}(\tilde{T}_2) e^{\frac{\sigma_1^2}{2} \tilde{T}_1^* A_1 A_2^* \tilde{T}_2}. \quad (vi)$$

Since \tilde{T}_1 and \tilde{T}_2 are arbitrary, the exponential part in (vi) is 1 if and only if $A_1 A_2^* = O$ or $A_2 A_1^* = O$, the two null matrices having different orders. Then, we have:

Theorem 2.3a.2. *Let the \tilde{x}_j 's and \tilde{X} be as defined in Theorem 2.3a.1. Let A_1 be a $m_1 \times n$ constant matrix and A_2 be a $m_2 \times n$ constant matrix, $m_1 \leq n$, $m_2 \leq n$, and the constant matrices may or may not be in the complex domain. Consider the general linear forms $\tilde{U}_1 = A_1 \tilde{X}$ and $\tilde{U}_2 = A_2 \tilde{X}$. Then \tilde{U}_1 and \tilde{U}_2 are independently distributed if and only if $A_1 A_2^* = O$ or, equivalently, $A_2 A_1^* = O$.*

Example 2.3a.2. Let \tilde{x}_j , $j = 1, 2, 3, 4$, be iid univariate complex Gaussian $\tilde{N}_1(\tilde{\mu}_1, \sigma_1^2)$. Consider the following sets of linear forms $\tilde{U}_1 = A_1 \tilde{X}$, $\tilde{U}_2 = A_2 \tilde{X}$, $\tilde{U}_3 = A_3 \tilde{X}$ with $X' = (x_1, x_2, x_3, x_4)$, where

$$A_1 = \begin{bmatrix} 2+3i & 2+3i & 2+3i & 2+3i \\ 1+i & -(1+i) & -(1+i) & 1+i \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 2i & -2i & 2i & -2i \\ 1-i & 1-i & -1+i & -1+i \\ -(1+2i) & 1+2i & -(1+2i) & 1+2i \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1+2i & 0 & 1-2i & -2 \\ -2 & -1+i & -1+i & 2i \end{bmatrix}.$$

Verify whether the pairs in $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3$ are independently distributed.

Solution 2.3a.2. Since the products $A_1 A_2^* = O$, $A_1 A_3^* \neq O$, $A_2 A_3^* \neq O$, only \tilde{U}_1 and \tilde{U}_2 are independently distributed.

As a corollary of Theorem 2.3a.2, one has that the sample mean $\bar{\tilde{x}}$ and the sample sum of products \tilde{s} are also independently distributed in the complex Gaussian case, a result parallel to the corresponding one in the real case. This can be seen by taking $A_1 = \frac{1}{n} J J'$ and $A_2 = I - \frac{1}{n} J J'$. Then, since $A_1 = A_1^2$, $A_2 = A_2^2$, and $A_1 A_2 = O$, we have

$\frac{1}{\sigma_1^2} \tilde{X}^* A_1 \tilde{X} \sim \tilde{\chi}_1^2$ for $\tilde{\mu}_1 = 0$ and $\frac{1}{\sigma_1^2} \tilde{X}^* A_2 \tilde{X} \sim \tilde{\chi}_{n-1}^2$, and both of these chisquares in the complex domain are independently distributed. Then,

$$\frac{1}{\sigma_1^2} n\tilde{s} = \frac{1}{\sigma_1^2} \sum_{j=1}^n (\tilde{x}_j - \bar{\tilde{x}})^* (\tilde{x}_j - \bar{\tilde{x}}) \sim \tilde{\chi}_{n-1}^2. \quad (2.3a.1)$$

The Student-t with $n - 1$ degrees of freedom can be defined in terms of the standardized sample mean and sample variance in the complex case.

2.3.1. Noncentral chisquare having n degrees of freedom in the real domain

Let $x_j \sim N_1(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$ and the x_j 's be independently distributed. Then, $\frac{x_j - \mu_j}{\sigma_j} \sim N_1(0, 1)$ and $\sum_{j=1}^n \frac{(x_j - \mu_j)^2}{\sigma_j^2} \sim \chi_n^2$ where χ_n^2 is a real chisquare with n degrees of freedom. Then, when at least one of the μ_j 's is nonzero, $\sum_{j=1}^n \frac{x_j^2}{\sigma_j^2}$ is referred to as a *real non-central chisquare with n degrees of freedom and non-centrality parameter λ , which is denoted $\chi_n^2(\lambda)$* , where

$$\lambda = \frac{1}{2} \mu' \Sigma^{-1} \mu, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}.$$

Let $u = \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2}$. In order to derive the distribution of u , let us determine its mgf. Since u is a function of the x_j 's where $x_j \sim N_1(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$, we can integrate out over the joint density of the x_j 's. Then, with t as the mgf parameter,

$$\begin{aligned} M_u(t) &= E[e^{tu}] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{t \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} - \frac{1}{2} \sum_{j=1}^n \frac{(x_j - \mu_j)^2}{\sigma_j^2}} dx_1 \wedge \dots \wedge dx_n. \end{aligned}$$

The exponent, excluding $-\frac{1}{2}$ can be simplified as follows:

$$-2t \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} + \sum_{j=1}^n \frac{(x_j - \mu_j)^2}{\sigma_j^2} = (1 - 2t) \sum_{j=1}^n \frac{x_j^2}{\sigma_j^2} - 2 \sum_{j=1}^n \frac{\mu_j x_j}{\sigma_j^2} + \sum_{j=1}^n \frac{\mu_j^2}{\sigma_j^2}.$$

Let $y_j = \sqrt{(1-2t)}x_j$. Then, $(1-2t)^{-\frac{n}{2}}dy_1 \wedge \dots \wedge dy_n = dx_1 \wedge \dots \wedge dx_n$, and

$$\begin{aligned} (1-2t) \sum_{j=1}^n \left(\frac{x_j^2}{\sigma_j^2} \right) - 2 \sum_{j=1}^n \left(\frac{\mu_j x_j}{\sigma_j^2} \right) &= \sum_{j=1}^n \frac{y_j^2}{\sigma_j^2} - 2 \sum_{j=1}^n \frac{\mu_j y_j}{\sigma_j^2 \sqrt{(1-2t)}} \\ &\quad + \sum_{j=1}^n \left(\frac{\mu_j}{\sigma_j \sqrt{(1-2t)}} \right)^2 - \sum_{j=1}^n \frac{\mu_j^2}{\sigma_j^2 (1-2t)} \\ &= \sum_{j=1}^n \left(\frac{\left(y_j - \frac{\mu_j}{\sqrt{(1-2t)}} \right)^2}{\sigma_j^2} \right) - \sum_{j=1}^n \frac{\mu_j^2}{\sigma_j^2 (1-2t)}. \end{aligned}$$

But

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\sum_{j=1}^n \frac{1}{2\sigma_j^2} \left(y_j - \frac{\mu_j}{\sqrt{(1-2t)}} \right)^2} dy_1 \wedge \dots \wedge dy_n = 1.$$

Hence, for $\lambda = \frac{1}{2} \sum_{j=1}^n \frac{\mu_j^2}{\sigma_j^2} = \frac{1}{2} \mu' \Sigma^{-1} \mu$,

$$\begin{aligned} M_u(t) &= \frac{1}{(1-2t)^{\frac{n}{2}}} [e^{-\lambda + \frac{\lambda}{(1-2t)}}] \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \frac{1}{(1-2t)^{\frac{n}{2}+k}}. \end{aligned} \tag{2.3.5}$$

However, $(1-2t)^{-\left(\frac{n}{2}+k\right)}$ is the mgf of a real scalar gamma with parameters $(\alpha = \frac{n}{2}+k, \beta = 2)$ or a real chisquare with $n+2k$ degrees of freedom or χ_{n+2k}^2 . Hence, the density of a non-central chisquare with n degrees of freedom and non-centrality parameter λ , denoted by $g_{u,\lambda}(u)$, is obtained by term by term inversion as follow:

$$g_{u,\lambda}(u) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \frac{u^{\frac{n+2k}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{n+2k}{2}} \Gamma\left(\frac{n}{2}+k\right)} \tag{2.3.6}$$

$$= \frac{u^{\frac{n}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \frac{u^k}{\left(\frac{n}{2}\right)_k} \tag{2.3.7}$$

where $\left(\frac{n}{2}\right)_k$ is the Pochhammer symbol given by

$$(a)_k = a(a+1) \dots (a+k-1), \quad a \neq 0, \quad (a)_0 \text{ being equal to } 1, \tag{2.3.8}$$

and, in general, $\Gamma(\alpha + k) = \Gamma(\alpha)(\alpha)_k$ for $k = 1, 2, \dots$, whenever the gamma functions are defined. Hence, provided $(1 - 2t) > 0$, (2.3.6) can be looked upon as a weighted sum of chisquare densities whose weights are Poisson distributed, that is, (2.3.6) is a Poisson mixture of chisquare densities. As well, we can view (2.3.7) as a chisquare density having n degrees of freedom appended with a Bessel series. In general, a Bessel series is of the form

$${}_0F_1(; b; x) = \sum_{k=0}^{\infty} \frac{1}{(b)_k} \frac{x^k}{k!}, \quad b \neq 0, -1, -2, \dots, \quad (2.3.9)$$

which is convergent for all x .

2.3.1.1. Mean value and variance, real central and non-central chisquare

The mgf of a real χ_ν^2 is $(1 - 2t)^{-\frac{\nu}{2}}$, $1 - 2t > 0$. Thus,

$$E[\chi_\nu^2] = \frac{d}{dt}(1 - 2t)^{-\frac{\nu}{2}}|_{t=0} = \left(-\frac{\nu}{2}\right)(-2)(1 - 2t)^{-\frac{\nu}{2}-1}|_{t=0} = \nu$$

$$E[\chi_\nu^2]^2 = \frac{d^2}{dt^2}(1 - 2t)^{-\frac{\nu}{2}}|_{t=0} = \left(-\frac{\nu}{2}\right)(-2)\left(-\frac{\nu}{2} - 1\right)(-2) = \nu(\nu + 2).$$

That is,

$$E[\chi_\nu^2] = \nu \text{ and } \text{Var}(\chi_\nu^2) = \nu(\nu + 2) - \nu^2 = 2\nu. \quad (2.3.10)$$

What are then the mean and the variance of a real non-central $\chi_\nu^2(\lambda)$? They can be derived either from the mgf or from the density. Making use of the density, we have

$$E[\chi_\nu^2(\lambda)] = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \int_0^{\infty} u \frac{u^{\frac{\nu}{2}+k-1} e^{-\frac{u}{2}}}{2^{\frac{\nu}{2}+k} \Gamma(\frac{\nu}{2} + k)} du,$$

the integral part being equal to

$$\frac{\Gamma(\frac{\nu}{2} + k + 1) 2^{\frac{\nu}{2}+k+1}}{\Gamma(\frac{\nu}{2} + k) 2^{\frac{\nu}{2}+k}} = 2\left(\frac{\nu}{2} + k\right) = \nu + 2k.$$

Now, the remaining summation over k can be looked upon as the expected value of $\nu + 2k$ in a Poisson distribution. In this case, we can write the expected values as the expected value of a conditional expectation: $E[u] = E[E(u|k)]$, $u = \chi_\nu^2(\lambda)$. Thus,

$$E[\chi_\nu^2(\lambda)] = \nu + 2 \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \nu + 2E[k] = \nu + 2\lambda.$$

Moreover,

$$E[\chi_v^2(\lambda)]^2 = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \int_0^{\infty} u^2 \frac{u^{\frac{v}{2}+k-1}}{2^{\frac{v}{2}+k} \Gamma(\frac{v}{2}+k)} du,$$

the integral part being

$$\begin{aligned} \frac{\Gamma(\frac{v}{2}+k+2)}{\Gamma(\frac{v}{2}+k)} \frac{2^{\frac{v}{2}+k+2}}{2^{\frac{v}{2}+k}} &= 2^2 \left(\frac{v}{2}+k+1\right) \left(\frac{v}{2}+k\right) \\ &= (v+2k+2)(v+2k) = v^2 + 2vk + 2v(k+1) + 4k(k+1). \end{aligned}$$

Since $E[k] = \lambda$, $E[k^2] = \lambda^2 + \lambda$ for a Poisson distribution,

$$E[\chi_v^2(\lambda)]^2 = v^2 + 2v + 4v\lambda + 4(\lambda^2 + 2\lambda).$$

Thus,

$$\begin{aligned} \text{Var}(\chi_v^2(\lambda)) &= E[\chi_v^2(\lambda)]^2 - [E(\chi_v^2(\lambda))]^2 \\ &= v^2 + 2v + 4v\lambda + 4(\lambda^2 + 2\lambda) - (v + 2\lambda)^2 \\ &= 2v + 8\lambda. \end{aligned}$$

To summarize,

$$E[\chi_v^2(\lambda)] = v + 2\lambda \text{ and } \text{Var}(\chi_v^2(\lambda)) = 2v + 8\lambda. \quad (2.3.11)$$

Example 2.3.3. Let $x_1 \sim N_1(-1, 2)$, $x_2 \sim N_1(1, 3)$ and $x_3 \sim N_1(-2, 2)$ be independently distributed and $u = \frac{x_1^2}{2} + \frac{x_2^2}{3} + \frac{x_3^2}{2}$. Provide explicit expressions for the density of u , $E[u]$ and $\text{Var}(u)$.

Solution 2.3.3. This u has a noncentral chisquare distribution with non-centrality parameter λ where

$$\lambda = \frac{1}{2} \left[\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} + \frac{\mu_3^2}{\sigma_3^2} \right] = \frac{1}{2} \left[\frac{(-1)^2}{2} + \frac{(1)^2}{3} + \frac{(-2)^2}{2} \right] = \frac{17}{12},$$

and the number of degrees of freedom is $n = 3 = v$. Thus, $u \sim \chi_3^2(\lambda)$ or a real noncentral chisquare with $v = 3$ degrees of freedom and non-centrality parameter $\frac{17}{12}$. Then $E[u] = E[\chi_3^2(\lambda)] = v + 2\lambda = 3 + 2(\frac{17}{12}) = \frac{35}{6}$. $\text{Var}(u) = \text{Var}(\chi_3^2(\lambda)) = 2v + 8\lambda = (2)(3) + 8(\frac{17}{12}) = \frac{52}{3}$. Let the density of u be denoted by $g(u)$. Then

$$\begin{aligned}
 g(u) &= \frac{u^{\frac{n}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \frac{u^k}{(\frac{n}{2})_k} \\
 &= \frac{u^{\frac{1}{2}} e^{-\frac{u}{2}}}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(17/12)^k e^{-17/12}}{k!} \frac{u^k}{(\frac{3}{2})_k}, \quad 0 \leq u < \infty,
 \end{aligned}$$

and zero elsewhere.

2.3a.1. Noncentral chisquare having n degrees of freedom in the complex domain

Let us now consider independently Gaussian distributed variables in the complex domain. Let the complex scalar variables $\tilde{x}_j \sim \tilde{N}(\tilde{\mu}_j, \sigma_j^2)$, $j = 1, \dots, n$, be independently distributed. Then, we have already established that $\sum_{j=1}^n (\frac{\tilde{x}_j - \tilde{\mu}_j}{\sigma_j})^* (\frac{\tilde{x}_j - \tilde{\mu}_j}{\sigma_j}) \sim \tilde{\chi}_n^2$, which is a chisquare variable having n degrees of freedom in the complex domain. If we let $\tilde{u} = \sum_{j=1}^n \frac{\tilde{x}_j^* \tilde{x}_j}{\sigma_j^2}$, then this u will be said to have a noncentral chisquare distribution with n degrees of freedom and non-centrality parameter λ in the complex domain where \tilde{x}_j^* is only the conjugate since it is a scalar quantity. Since, in this case, \tilde{u} is real, we may associate the mgf of \tilde{u} with a real parameter t . Now, proceeding as in the real case, we obtain the mgf of \tilde{u} , denoted by $M_{\tilde{u}}(t)$, as follows:

$$\begin{aligned}
 M_{\tilde{u}}(t) &= E[e^{t\tilde{u}}] = (1-t)^{-n} e^{-\lambda + \frac{\lambda}{1-t}}, \quad 1-t > 0, \quad \lambda = \sum_{j=1}^n \frac{\tilde{\mu}_j^* \tilde{\mu}_j}{\sigma_j^2} = \tilde{\mu}^* \Sigma^{-1} \tilde{\mu} \\
 &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} (1-t)^{-(n+k)}. \tag{2.3a.2}
 \end{aligned}$$

Note that the inverse corresponding to $(1-t)^{-(n+k)}$ is a chisquare density in the complex domain with $n+k$ degrees of freedom, and that part of the density, denoted by $f_1(u)$, is

$$f_1(u) = \frac{1}{\Gamma(n+k)} u^{n+k-1} e^{-u} = \frac{1}{\Gamma(n)(n)_k} u^{n-1} u^k e^{-u}.$$

Thus, the noncentral chisquare density with n degrees of freedom in the complex domain, that is, $u = \tilde{\chi}_n^2(\lambda)$, denoted by $f_{u,\lambda}(u)$, is

$$f_{u,\lambda}(u) = \frac{u^{n-1}}{\Gamma(n)} e^{-u} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \frac{u^k}{(n)_k}, \tag{2.3a.3}$$

which, referring to Eqs. (2.3.5)–(2.3.9) in connection with a non-central chisquare in the real domain, can also be represented in various ways.

Example 2.3a.3. Let $\tilde{x}_1 \sim \tilde{N}_1(1 + i, 2)$, $\tilde{x}_2 \sim \tilde{N}_1(2 + i, 4)$ and $\tilde{x}_3 \sim \tilde{N}_1(1 - i, 2)$ be independently distributed univariate complex Gaussian random variables and

$$\tilde{u} = \frac{\tilde{x}_1^* \tilde{x}_1}{\sigma_1^2} + \frac{\tilde{x}_2^* \tilde{x}_2}{\sigma_2^2} + \frac{\tilde{x}_3^* \tilde{x}_3}{\sigma_3^2} = \frac{\tilde{x}_1^* \tilde{x}_1}{2} + \frac{\tilde{x}_2^* \tilde{x}_2}{4} + \frac{\tilde{x}_3^* \tilde{x}_3}{2}.$$

Compute $E[\tilde{u}]$ and $\text{Var}(\tilde{u})$ and provide an explicit representation of the density of \tilde{u} .

Solution 2.3a.3. In this case, \tilde{u} has a noncentral chisquare distribution with degrees of freedom $\nu = n = 3$ and non-centrality parameter λ given by

$$\lambda = \frac{\tilde{\mu}_1^* \tilde{\mu}_1}{\sigma_1^2} + \frac{\tilde{\mu}_2^* \tilde{\mu}_2}{\sigma_2^2} + \frac{\tilde{\mu}_3^* \tilde{\mu}_3}{\sigma_3^2} = \frac{[(1)^2 + (1)^2]}{2} + \frac{[(2)^2 + (1)^2]}{4} + \frac{[(1)^2 + (-1)^2]}{2} = 1 + \frac{5}{4} + 1 = \frac{13}{4}.$$

The density, denoted by $g_1(\tilde{u})$, is given in (i). In this case \tilde{u} will be a real gamma with the parameters $(\alpha = n, \beta = 1)$ to which a Poisson series is appended:

$$g_1(u) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \frac{u^{n+k-1} e^{-u}}{\Gamma(n+k)}, \quad 0 \leq u < \infty, \quad (i)$$

and zero elsewhere. Then, the expected value of u and the variance of u are available from (i) by direct integration.

$$E[u] = \int_0^{\infty} u g_1(u) du = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \frac{\Gamma(n+k+1)}{\Gamma(n+k)}.$$

But $\frac{\Gamma(n+k+1)}{\Gamma(n+k)} = n+k$ and the summation over k can be taken as the expected value of $n+k$ in a Poisson distribution. Thus, $E[\tilde{\chi}_n^2(\lambda)] = n + E[k] = n + \lambda$. Now, in the expected value of

$$[\tilde{\chi}_n^2(\lambda)][\tilde{\chi}_n^2(\lambda)]^* = [\tilde{\chi}_n^2(\lambda)]^2 = [u]^2,$$

which is real in this case, the integral part over u gives

$$\frac{\Gamma(n+k+2)}{\Gamma(n+k)} = (n+k+1)(n+k) = n^2 + 2nk + k^2 + n + k$$

with expected value $n^2 + 2n\lambda + n + \lambda + (\lambda^2 + \lambda)$. Hence,

$$\text{Var}(\tilde{\chi}_n^2(\lambda)) = E[u - E(u)][u - E(u)]^* = \text{Var}(u) = n^2 + 2n\lambda + n + \lambda + (\lambda^2 + \lambda) - (n + \lambda)^2,$$

which simplifies to $n + 2\lambda$. Accordingly,

$$E[\tilde{\chi}_n^2(\lambda)] = n + \lambda \text{ and } \text{Var}(\tilde{\chi}_n^2(\lambda)) = n + 2\lambda, \quad (ii)$$

so that

$$E[u] = n + \lambda = 3 + \frac{13}{4} = \frac{25}{4} \text{ and } \text{Var}(u) = n + 2\lambda = 3 + \frac{13}{2} = \frac{19}{2}. \quad (iii)$$

The explicit form of the density is then

$$g_1(u) = \frac{u^2 e^{-u}}{2} \sum_{k=0}^{\infty} \frac{(13/4)^k e^{-13/4}}{k!} \frac{u^k}{(3)_k}, \quad 0 \leq u < \infty, \quad (iv)$$

and zero elsewhere.

Exercises 2.3

2.3.1. Let x_1, \dots, x_n be iid variables with common density a real gamma density with the parameters α and β or with the mgf $(1 - \beta t)^{-\alpha}$, $1 - \beta t > 0$, $\alpha > 0$, $\beta > 0$. Let $u_1 = x_1 + \dots + x_n$, $u_2 = \frac{1}{n}(x_1 + \dots + x_n)$, $u_3 = u_2 - \alpha\beta$, $u_4 = \frac{\sqrt{nu_3}}{\beta\sqrt{\alpha}}$. Evaluate the mgfs and thereby the densities of u_1, u_2, u_3, u_4 . Show that they are all gamma densities for all finite n , may be relocated. Show that when $n \rightarrow \infty$, $u_4 \rightarrow N_1(0, 1)$ or u_4 goes to a real standard normal when n goes to infinity.

2.3.2. Let x_1, \dots, x_n be a simple random sample of size n from a real population with mean value μ and variance $\sigma^2 < \infty$, $\sigma > 0$. Then the central limit theorem says that $\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \rightarrow N_1(0, 1)$ as $n \rightarrow \infty$, where $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$. Translate this statement for

(1): binomial probability function $f_1(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \dots, n$, $0 < p < 1$

and $f_1(x) = 0$ elsewhere; (2): negative binomial probability law $f_2(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$, $x = k, k+1, \dots$, $0 < p < 1$ and zero elsewhere; (3): geometric probability law $f_3(x) = p(1-p)^{x-1}$, $x = 1, 2, \dots$, $0 < p < 1$ and $f_3(x) = 0$ elsewhere; (4): Poisson probability law $f_4(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \dots$, $\lambda > 0$ and $f_4(x) = 0$ elsewhere.

2.3.3. Repeat Exercise 2.3.2 if the population is (1): $g_1(x) = c_1 x^{\gamma-1} e^{-ax^\delta}$, $x \geq 0$, $\delta > 0$, $a > 0$, $\gamma > 0$ and $g_1(x) = 0$ elsewhere; (2): The real pathway model $g_2(x) = c_q x^\gamma [1 - a(1-q)x^\delta]^{1-q}$, $a > 0$, $\delta > 0$, $1 - a(1-q)x^\delta > 0$ and for the cases $q < 1$, $q > 1$, $q \rightarrow 1$, and $g_2(x) = 0$ elsewhere.

2.3.4. Let $x \sim N_1(\mu_1, \sigma_1^2)$, $y \sim N_1(\mu_2, \sigma_2^2)$ be real Gaussian and be independently distributed. Let $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ be simple random samples from x and y respectively. Let $u_1 = \sum_{j=1}^{n_1} (x_j - \mu_1)^2$, $u_2 = \sum_{j=1}^{n_2} (y_j - \mu_2)^2$, $u_3 = 2x_1 - 3x_2 + y_1 - y_2 + 2y_3$,

$$u_4 = \frac{\sum_{j=1}^{n_1} (x_j - \mu_1)^2 / \sigma_1^2}{\sum_{j=1}^{n_2} (y_j - \mu_2)^2 / \sigma_2^2}, u_5 = \frac{\sum_{j=1}^{n_1} (x_j - \bar{x})^2 / \sigma_1^2}{\sum_{j=1}^{n_2} (y_j - \bar{y})^2 / \sigma_2^2}.$$

Compute the densities of u_1, u_2, u_3, u_4, u_5 .

2.3.5. In Exercise 2.3.4 if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ compute the densities of u_3, u_4, u_5 there, and $u_6 = \sum_{j=1}^{n_1} (x_j - \bar{x})^2 + \sum_{j=1}^{n_2} (y_j - \bar{y})^2$ if (1): $n_1 = n_2$, (2): $n_1 \neq n_2$.

2.3.6. For the noncentral chisquare in the complex case, discussed in (2.3a.5) evaluate the mean value and the variance.

2.3.7. For the complex case, starting with the mgf, derive the noncentral chisquare density and show that it agrees with that given in (2.3a.3).

2.3.8. Give the detailed proofs of the independence of linear forms and sets of linear forms in the complex Gaussian case.

2.4. Distributions of Products and Ratios and Connection to Fractional Calculus

Distributions of products and ratios of real scalar random variables are connected to numerous topics including Krätzel integrals and transforms, reaction-rate probability integrals in nuclear reaction-rate theory, the inverse Gaussian distribution, integrals occurring in fractional calculus, Kobayashi integrals and Bayesian structures. Let $x_1 > 0$ and $x_2 > 0$ be real scalar positive random variables that are independently distributed with density functions $f_1(x_1)$ and $f_2(x_2)$, respectively. We respectively denote the product and ratio of these variables by $u_2 = x_1 x_2$ and $u_1 = \frac{x_2}{x_1}$. What are then the densities of u_1 and u_2 ? We first consider the density of the product. Let $u_2 = x_1 x_2$ and $v = x_2$. Then $x_1 = \frac{u_2}{v}$ and $x_2 = v$, $dx_1 \wedge dx_2 = \frac{1}{v} du \wedge dv$. Let the joint density of u_2 and v be denoted by $g(u_2, v)$ and the marginal density of u_2 by $g_2(u_2)$. Then

$$g(u_2, v) = \frac{1}{v} f_1\left(\frac{u_2}{v}\right) f_2(v) \text{ and } g_2(u_2) = \int_v \frac{1}{v} f_2\left(\frac{u_2}{v}\right) f_1(v) dv. \quad (2.4.1)$$

For example, let f_1 and f_2 be generalized gamma densities, in which case

$$f_j(x_j) = \frac{a_j^{\frac{\gamma_j}{\delta_j}}}{\Gamma\left(\frac{\gamma_j}{\delta_j}\right)} x_j^{\gamma_j - 1} e^{-a_j x_j^{\delta_j}}, \quad a_j > 0, \delta_j > 0, \gamma_j > 0, x_j \geq 0 \quad j = 1, 2,$$

and $f_j(x_j) = 0$ elsewhere. Then,

$$\begin{aligned}
 g_2(u_2) &= c \int_0^\infty \left(\frac{1}{v}\right) \left(\frac{u_2}{v}\right)^{\gamma_1-1} v^{\gamma_2-1} \\
 &\quad \times e^{-a_2 v^{\delta_2} - a_1 \left(\frac{u_2}{v}\right)^{\delta_1}} dv, \quad c = \prod_{j=1}^2 \frac{a_j^{\frac{\gamma_j}{\delta_j}}}{\Gamma\left(\frac{\gamma_j}{\delta_j}\right)} \\
 &= c u_2^{\gamma_1-1} \int_0^\infty v^{\gamma_2-\gamma_1-1} e^{-a_2 v^{\delta_2} - a_1 (u_2^{\delta_1}/v^{\delta_1})} dv. \tag{2.4.2}
 \end{aligned}$$

The integral in (2.4.2) is connected to several topics. For $\delta_1 = 1, \delta_2 = 1$, this integral is the basic Krätzel integral and Krätzel transform, see Mathai and Haubold (2020). When $\delta_2 = 1, \delta_1 = \frac{1}{2}$, the integral in (2.4.2) is the basic reaction-rate probability integral in nuclear reaction-rate theory, see Mathai and Haubold (1988). For $\delta_1 = 1, \delta_2 = 1$, the integrand in (2.4.2), once normalized, is the inverse Gaussian density for appropriate values of $\gamma_2 - \gamma_1 - 1$. Observe that (2.4.2) is also connected to the Bayesian structure of unconditional densities if the conditional and marginal densities belong to generalized gamma family of densities. When $\delta_2 = 1$, the integral is a mgf of the remaining part with a_2 as the mgf parameter (It is therefore the Laplace transform of the remaining part of the function).

Now, let us consider different f_1 and f_2 . Let $f_1(x_1)$ be a real type-1 beta density with the parameters $(\gamma + 1, \alpha)$, $\Re(\alpha) > 0, \Re(\gamma) > -1$ (in statistical problems, the parameters are real but in this case the results hold as well for complex parameters; accordingly, the conditions are stated for complex parameters), that is, the density of x_1 is

$$f_1(x_1) = \frac{\Gamma(\gamma + 1 + \alpha)}{\Gamma(\gamma + 1)\Gamma(\alpha)} x_1^\gamma (1 - x_1)^{\alpha-1}, \quad 0 \leq x_1 \leq 1, \quad \alpha > 0, \quad \gamma > -1,$$

and $f_1(x_1) = 0$ elsewhere. Let $f_2(x_2) = f(x_2)$ where f is an arbitrary density. Then, the density of u_2 is given by

$$\begin{aligned}
 g_2(u_2) &= c \frac{1}{\Gamma(\alpha)} \int_v \frac{1}{v} \left(\frac{u_2}{v}\right)^\gamma \left(1 - \frac{u_2}{v}\right)^{\alpha-1} f(v) dv, \quad c = \frac{\Gamma(\gamma + 1 + \alpha)}{\Gamma(\gamma + 1)} \\
 &= c \frac{u_2^\gamma}{\Gamma(\alpha)} \int_{v \geq u_2 > 0} v^{-\gamma-\alpha} (v - u_2)^{\alpha-1} f(v) dv \tag{2.4.3}
 \end{aligned}$$

$$= c K_{2,u_2,\gamma}^{-\alpha} f \tag{2.4.4}$$

where $K_{2,u_2,\gamma}^{-\alpha} f$ is the Erdélyi-Kober fractional integral of order α of the second kind, with parameter γ in the real scalar variable case. Hence, if f is an arbitrary density, then $K_{2,u_2,\gamma}^{-\alpha} f$ is $\frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1+\alpha)} g_2(u_2)$ or a constant multiple of the density of a product of independently distributed real scalar positive random variables where one of them has a real type-1 beta density and the other has an arbitrary density. When f_1 and f_2 are densities, then $g_2(u_2)$ has the structure

$$g_2(u_2) = \int_v^1 \frac{1}{v} f_1\left(\frac{u_2}{v}\right) f_2(v) dv. \quad (i)$$

Whether or not f_1 and f_2 are densities, the structure in (i) is called *the Mellin convolution of a product* in the sense that if we take the Mellin transform of g_2 , with Mellin parameter s , then

$$M_{g_2}(s) = M_{f_1}(s) M_{f_2}(s) \quad (2.4.5)$$

where $M_{g_2}(s) = \int_0^\infty u_2^{s-1} g_2(u_2) du_2$,

$$M_{f_1}(s) = \int_0^\infty x_1^{s-1} f_1(x_1) dx_1 \text{ and } M_{f_2}(s) = \int_0^\infty x_2^{s-1} f_2(x_2) dx_2,$$

whenever the Mellin transforms exist. Here (2.4.5) is the Mellin convolution of a product property. In statistical terms, when f_1 and f_2 are densities and when x_1 and x_2 are independently distributed, we have

$$E[u_2^{s-1}] = E[x_1^{s-1}] E[x_2^{s-1}] \quad (2.4.6)$$

whenever the expected values exist. Taking different forms of f_1 and f_2 , where f_1 has a factor $\frac{(1-x_1)^{\alpha-1}}{\Gamma(\alpha)}$ for $0 \leq x_1 \leq 1$, $\Re(\alpha) > 0$, it can be shown that the structure appearing in (i) produces all the various fractional integrals of the second kind of order α available in the literature for the real scalar variable case, such as the Riemann-Liouville fractional integral, Weyl fractional integral, etc. Connections of distributions of products and ratios to fractional integrals were established in a series of papers which appeared in *Linear Algebra and its Applications*, see Mathai (2013, 2014, 2015).

Now, let us consider the density of a ratio. Again, let $x_1 > 0$, $x_2 > 0$ be independently distributed real scalar random variables with density functions $f_1(x_1)$ and $f_2(x_2)$, respectively. Let $u_1 = \frac{x_2}{x_1}$ and let $v = x_2$. Then $dx_1 \wedge dx_2 = -\frac{v}{u_1^2} du_1 \wedge dv$. If we take $x_1 = v$, the Jacobian will be only v and not $-\frac{v}{u_1^2}$ and the final structure will be different. However, the first transformation is required in order to establish connections to fractional integral

of the first kind. If f_1 and f_2 are generalized gamma densities as described earlier and if $x_1 = v$, then the marginal density of u_1 , denoted by $g_1(u_1)$, will be as follows:

$$g_1(u_1) = c \int_v v v^{\gamma_1-1} (u_1 v)^{\gamma_2-1} e^{-a_1 v^{\delta_1} - a_2 (u_1 v)^{\delta_2}} dv, \quad c = \prod_{j=1}^2 \frac{a_j^{\frac{\gamma_j}{\delta_j}}}{\Gamma(\frac{\gamma_j}{\delta_j})},$$

$$= c u_1^{\gamma_2-1} \int_{v=0}^{\infty} v^{\gamma_1+\gamma_2-1} e^{-a_1 v^{\delta_1} - a_2 (u_1 v)^{\delta_2}} dv \quad (2.4.7)$$

$$= \frac{c}{\delta} \Gamma\left(\frac{\gamma_1 + \gamma_2}{\delta}\right) u_1^{\gamma_2-1} (a_1 + a_2 u_1^{\delta})^{-\frac{\gamma_1+\gamma_2}{\delta}}, \quad \text{for } \delta_1 = \delta_2 = \delta. \quad (2.4.8)$$

On the other hand, if $x_2 = v$, then the Jacobian is $-\frac{v}{u_1^2}$ and the marginal density, again denoted by $g_1(u_1)$, will be as follows when f_1 and f_2 are gamma densities:

$$g_1(u_1) = c \int_v \left(\frac{v}{u_1^2}\right) \left(\frac{v}{u_1}\right)^{\gamma_1-1} v^{\gamma_2-1} e^{-a_1 \left(\frac{v}{u_1}\right)^{\delta_1} - a_2 v^{\delta_2}} dv$$

$$= c u_1^{-\gamma_1-1} \int_{v=0}^{\infty} v^{\gamma_1+\gamma_2-1} e^{-a_2 v^{\delta_2} - a_1 \left(\frac{v}{u_1}\right)^{\delta_1}} dv.$$

This is one of the representations of the density of a product discussed earlier, which is also connected to Krätzel integral, reaction-rate probability integral, and so on. Now, let us consider a type-1 beta density for x_1 with the parameters (γ, α) having the following density:

$$f_1(x_1) = \frac{\Gamma(\gamma + \alpha)}{\Gamma(\gamma)\Gamma(\alpha)} x_1^{\gamma-1} (1 - x_1)^{\alpha-1}, \quad 0 \leq x_1 \leq 1,$$

for $\gamma > 0$, $\alpha > 0$ and $f_1(x_1) = 0$ elsewhere. Let $f_2(x_2) = f(x_2)$ where f is an arbitrary density. Letting $u_1 = \frac{x_2}{x_1}$ and $x_2 = v$, the density of u_1 , again denoted by $g_1(u_1)$, is

$$g_1(u_1) = \frac{\Gamma(\gamma + \alpha)}{\Gamma(\gamma)\Gamma(\alpha)} \int_v \frac{v}{u_1^2} \left(\frac{v}{u_1}\right)^{\gamma-1} \left(1 - \frac{v}{u_1}\right)^{\alpha-1} f(v) dv$$

$$= \frac{\Gamma(\gamma + \alpha)}{\Gamma(\gamma)} \frac{u_1^{-\gamma-\alpha}}{\Gamma(\alpha)} \int_{v \leq u_1} v^{\gamma} (u_1 - v)^{\alpha-1} f(v) dv \quad (2.4.9)$$

$$= \frac{\Gamma(\gamma + \alpha)}{\Gamma(\gamma)} K_{1,u_1,\gamma}^{-\alpha} f, \quad \Re(\alpha) > 0, \quad (2.4.10)$$

where $K_{1,u_1,\gamma}^{-\alpha} f$ is Erdélyi-Kober fractional integral of the first kind of order α and parameter γ . If f_1 and f_2 are densities, this Erdélyi-Kober fractional integral of the first kind is a constant multiple of the density of a ratio $g_1(u_1)$. In statistical terms,

$$u_1 = \frac{x_2}{x_1} \Rightarrow E[u_1^{s-1}] = E[x_2^{s-1}]E[x_1^{-s+1}] \text{ with } E[x_1^{-s+1}] = E[x_1^{(2-s)-1}] \Rightarrow$$

$$M_{g_1}(s) = M_{f_1}(2-s)M_{f_2}(s), \quad (2.4.11)$$

which is the Mellin convolution of a ratio. Whether or not f_1 and f_2 are densities, (2.4.11) is taken as the Mellin convolution of a ratio and it cannot be given statistical interpretations when f_1 and f_2 are not densities. For example, let $f_1(x_1) = x_1^{-\alpha} \frac{(1-x_1)^{\alpha-1}}{\Gamma(\alpha)}$ and $f_2(x_2) = x_2^\alpha f(x_2)$ where $f(x_2)$ is an arbitrary function. Then the Mellin convolution of a ratio, as in (2.4.11), again denoted by $g_1(u_1)$, is given by

$$g_1(u_1) = \int_{v \leq u_1} \frac{(u_1 - v)^{\alpha-1}}{\Gamma(\alpha)} f(v) dv, \quad \Re(\alpha) > 0. \quad (2.4.12)$$

This is *Riemann-Liouville fractional integral of the first kind of order α* if v is bounded below; when v is not bounded below, then it is *Weyl fractional integral of the first kind of order α* . An introduction to fractional calculus is presented in Mathai and Haubold (2018). The densities of u_1 and u_2 are connected to various problems in different areas for different functions f_1 and f_2 .

In the $p \times p$ matrix case in the complex domain, we will assume that the matrix is Hermitian positive definite. Note that when $p = 1$, Hermitian positive definite means a real positive variable. Hence in the scalar case, we will not discuss ratios and products in the complex domain since densities must be real-valued functions.

Exercises 2.4

2.4.1. Derive the density of (1): a real non-central F, where the numerator chisquare is non-central and the denominator chisquare is central, (2): a real doubly non-central F where both the chisquares are non-central with non-centrality parameters λ_1 and λ_2 respectively.

2.4.2. Let x_1 and x_2 be real gamma random variables with parameters (α_1, β) and (α_2, β) with the same β respectively and be independently distributed. Let $u_1 = \frac{x_1}{x_1+x_2}$, $u_2 = \frac{x_1}{x_2}$, $u_3 = x_1 + x_2$. Compute the densities of u_1, u_2, u_3 . Hint: Use the transformation $x_1 = r \cos^2 \theta$, $x_2 = r \sin^2 \theta$.

2.4.3. Let x_1 and x_2 be as defined as in Exercise 2.4.2. Let $u = x_1x_2$. Derive the density of u .

2.4.4. Let x_j have a real type-1 beta density with the parameters (α_j, β_j) , $j = 1, 2$ and be independently distributed. Let $u_1 = x_1x_2$, $u_2 = \frac{x_1}{x_2}$. Derive the densities of u_1 and u_2 . State the conditions under which these densities reduce to simpler known densities.

2.4.5. Evaluate (i): Weyl fractional integral of the second kind of order α if the arbitrary function is $f(v) = e^{-v}$; (ii) Riemann-Liouville fractional integral of the first kind of order α if the lower limit is 0 and the arbitrary function is $f(v) = v^\delta$.

2.4.6. In Exercise 2.4.2 show that (1): u_1 and u_3 are independently distributed; (2): u_2 and u_3 are independently distributed.

2.4.7. In Exercise 2.4.2 show that for arbitrary h , $E\left[\frac{x_1}{x_1+x_2}\right]^h = \frac{E(x_1^h)}{E(x_1+x_2)^h}$ and state the conditions for the existence of the moments. [Observe that, in general, $E\left(\frac{y_1}{y_2}\right)^h \neq \frac{E(y_1^h)}{E(y_2^h)}$ even if y_1 and y_2 are independently distributed.]

2.4.8. Derive the corresponding densities in Exercise 2.4.1 for the complex domain by taking the chisquares in the complex domain.

2.4.9. Extend the results in Exercise 2.4.2 to the complex domain by taking chisquare variables in the complex domain instead of gamma variables.

2.5. General Structures

2.5.1. Product of real scalar gamma variables

Let x_1, \dots, x_k be independently distributed real scalar gamma random variables with x_j having the density $f_j(x_j) = c_j x_j^{\alpha_j-1} e^{-\frac{x_j}{\beta_j}}$, $0 \leq x_j < \infty$, $\alpha_j > 0$, $\beta_j > 0$ and $f_j(x_j) = 0$ elsewhere. Consider the product $u = x_1x_2 \cdots x_k$. Such structures appear in many situations such as geometrical probability problems when we consider gamma distributed random points, see Mathai (1999). How can we determine the density of such a general structure? The transformation of variables technique is not a feasible procedure in this case. Since the x_j 's are positive, we may determine the Mellin transforms of the x_j 's with parameter s . Then, when $f_j(x_j)$ is a density, the Mellin transform $M_{f_j}(s)$, once

expressed in terms of an expected value, is $M_{f_j}(s) = E[x_j^{s-1}]$ whenever the expected value exists:

$$\begin{aligned} M_{f_j}(s) &= E[x_j^{s-1}] = \frac{1}{\beta_j^{\alpha_j} \Gamma(\beta_j)} \int_0^\infty x_j^{s-1} x_j^{\alpha_j-1} e^{-\frac{x_j}{\beta_j}} dx_j \\ &= \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j)} \beta_j^{s-1}, \Re(\alpha_j + s - 1) > 0. \end{aligned}$$

Hence,

$$E[u^{s-1}] = \prod_{j=1}^k E[x_j^{s-1}] = \prod_{j=1}^k \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j)} \beta_j^{s-1}, \Re(\alpha_j + s - 1) > 0, j = 1, \dots, k,$$

and the density of u is available from the inverse Mellin transform. If $g(u)$ is the density of u , then

$$g(u) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left\{ \prod_{j=1}^k \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j)} \beta_j^{s-1} \right\} u^{-s} dx, i = \sqrt{-1}. \quad (2.5.1)$$

This is a contour integral where c is any real number such that $c > -\Re(\alpha_j - 1)$, $j = 1, \dots, k$. The integral in (2.5.1) is available in terms of a known special function, namely Meijer's G-function. The G-function can be defined as follows:

$$\begin{aligned} G(z) &= G_{p,q}^{m,n}(z) = G_{p,q}^{m,n} \left[z \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right] \\ &= \frac{1}{2\pi i} \int_L \frac{\{\prod_{j=1}^m \Gamma(b_j + s)\} \{\prod_{j=1}^n \Gamma(1 - a_j - s)\}}{\{\prod_{j=m+1}^q \Gamma(1 - b_j - s)\} \{\prod_{j=n+1}^p \Gamma(a_j + s)\}} z^{-s} ds, i = \sqrt{-1}. \end{aligned} \quad (2.5.2)$$

The existence conditions, different possible contours L , as well as properties and applications are discussed in Mathai (1993), Mathai and Saxena (1973, 1978), and Mathai et al. (2010). With the help of (2.5.2), we may now express (2.5.1) as follows in terms of a G-function:

$$g(u) = \left\{ \prod_{j=1}^k \frac{1}{\beta_j \Gamma(\alpha_j)} \right\} G_{0,k}^{k,0} \left[\frac{u}{\beta_1 \dots \beta_k} \middle|_{\alpha_j-1, j=1, \dots, k} \right] \quad (2.5.3)$$

for $0 \leq u < \infty$. Series and computable forms of a general G-function are provided in Mathai (1993). They are built-in functions in the symbolic computational packages *Mathematica* and *MAPLE*.

2.5.2. Product of real scalar type-1 beta variables

Let y_1, \dots, y_k be independently distributed real scalar type-1 beta random variables with the parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$, $j = 1, \dots, k$. Consider the product $u_1 = y_1 \cdots y_k$. Such a structure occurs in several contexts. It appears for instance in geometrical probability problems in connection with type-1 beta distributed random points. As well, when testing certain hypotheses on the parameters of one or more multivariate normal populations, the resulting likelihood ratio criteria, also known as λ -criteria, or one- to-one functions thereof, have the structure of a product of independently distributed real type-1 beta variables under the null hypothesis. The density of u_1 can be obtained by proceeding as in the previous section. Since the moment of u_1 of order $s - 1$ is

$$\begin{aligned} E[u_1^{s-1}] &= \prod_{j=1}^k E[y_j^{s-1}] \\ &= \prod_{j=1}^k \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j + \beta_j + s - 1)} \end{aligned} \quad (2.5.4)$$

for $\Re(\alpha_j + s - 1) > 0$, $j = 1, \dots, k$, Then, the density of u_1 , denoted by $g_1(u_1)$, is given by

$$\begin{aligned} g_1(u_1) &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} [E(u_1^{s-1})] u_1^{-s} ds \\ &= \left\{ \prod_{j=1}^k \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)} \right\} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left\{ \prod_{j=1}^k \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j + \beta_j + s - 1)} \right\} u_1^{-s} ds \\ &= \left\{ \prod_{j=1}^k \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)} \right\} G_{k,k}^{k,0} \left[u_1 \middle| \begin{matrix} \alpha_j + \beta_j - 1, j=1, \dots, k \\ \alpha_j - 1, j=1, \dots, k \end{matrix} \right] \end{aligned} \quad (2.5.5)$$

for $0 \leq u_1 \leq 1$, $\Re(\alpha_j + s - 1) > 0$, $j = 1, \dots, k$.

2.5.3. Product of real scalar type-2 beta variables

Let $u_2 = z_1 z_2 \cdots z_k$ where the z_j 's are independently distributed real scalar type-2 beta random variables with the parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$, $j = 1, \dots, k$. Such products are encountered in several situations, including certain problems in geometrical probability that are discussed in Mathai (1999). Then,

$$E[z_j^{s-1}] = \frac{\Gamma(\alpha_j + s - 1) \Gamma(\beta_j - s + 1)}{\Gamma(\alpha_j) \Gamma(\beta_j)}, \quad -\Re(\alpha_j - 1) < \Re(s) < \Re(\beta_j + 1),$$

and

$$E[u_2^{s-1}] = \prod_{j=1}^k E[z_j^{s-1}] = \left\{ \prod_{j=1}^k [\Gamma(\alpha_j) \Gamma(\beta_j)]^{-1} \right\} \left\{ \prod_{j=1}^k \Gamma(\alpha_j + s - 1) \Gamma(\beta_j - s + 1) \right\}. \quad (2.5.6)$$

Hence, the density of u_2 , denoted by $g_2(u_2)$, is given by

$$\begin{aligned} g_2(u_2) &= \left\{ \prod_{j=1}^k [\Gamma(\alpha_j) \Gamma(\beta_j)]^{-1} \right\} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left\{ \prod_{j=1}^k \Gamma(\alpha_j + s - 1) \Gamma(\beta_j - s + 1) \right\} u_2^{-s} ds \\ &= \left\{ \prod_{j=1}^k [\Gamma(\alpha_j) \Gamma(\beta_j)]^{-1} \right\} G_{k,k}^{k,k} \left[u_2 \middle|_{\alpha_j-1, j=1, \dots, k}^{-\beta_j, j=1, \dots, k} \right], \quad u_2 \geq 0. \end{aligned} \quad (2.5.7)$$

2.5.4. General products and ratios

Let us consider a structure of the following form:

$$u_3 = \frac{t_1 \cdots t_r}{t_{r+1} \cdots t_k},$$

where the t_j 's are independently distributed real positive variables, such as real type-1 beta, real type-2 beta, and real gamma variables, where the expected values $E[t_j^{s-1}]$ for $j = 1, \dots, k$, will produce various types of gamma products, some containing $+s$ and others, $-s$, both in the numerator and in the denominator. Accordingly, we obtain a general structure such as that appearing in (2.5.2), and the density of u_3 , denoted by $g_3(u_3)$, will then be proportional to a general G-function.

2.5.5. The H-function

Let $u = v_1 v_2 \cdots v_k$ where the v_j 's are independently distributed generalized real gamma variables with densities

$$h_j(v_j) = \frac{a_j^{\frac{\gamma_j}{\delta_j}}}{\Gamma(\frac{\gamma_j}{\delta_j})} v_j^{\gamma_j-1} e^{-a_j v_j^{\delta_j}}, \quad v_j \geq 0,$$

for $a_j > 0$, $\delta_j > 0$, $\gamma_j > 0$, and $h_j(v_j) = 0$ elsewhere for $j = 1, \dots, k$. Then,

$$E[v_j^{s-1}] = \frac{\Gamma(\frac{\gamma_j+s-1}{\delta_j})}{\Gamma(\frac{\gamma_j}{\delta_j})} \frac{1}{a_j^{\frac{s-1}{\delta_j}}},$$

for $\Re(\gamma_j + s - 1) > 0$, $v_j \geq 0$, $\delta_j > 0$, $a_j > 0$, $\gamma_j > 0$, $j = 1, \dots, k$, and

$$E[u^{s-1}] = \left\{ \prod_{j=1}^k \frac{a_j^{\frac{1}{\delta_j}}}{\Gamma(\frac{\gamma_j}{\delta_j})} \right\} \left\{ \prod_{j=1}^k \Gamma(\frac{\gamma_j - 1}{\delta_j} + \frac{s}{\delta_j}) a_j^{-\frac{s}{\delta_j}} \right\}. \quad (2.5.8)$$

Thus, the density of u , denoted by $g_3(u)$, is given by

$$\begin{aligned} g_3(u) &= \left\{ \prod_{j=1}^k \frac{a_j^{\frac{1}{\delta_j}}}{\Gamma(\frac{\gamma_j}{\delta_j})} \right\} \frac{1}{2\pi i} \int_L \left\{ \prod_{j=1}^k \Gamma(\frac{\gamma_j - 1}{\delta_j} + \frac{s}{\delta_j}) a_j^{-\frac{s}{\delta_j}} \right\} u^{-s} ds \\ &= \left\{ \prod_{j=1}^k \frac{a_j^{\frac{1}{\delta_j}}}{\Gamma(\frac{\gamma_j}{\delta_j})} \right\} H_{0,k}^{k,0} \left[\left\{ \prod_{j=1}^k a_j^{\frac{1}{\delta_j}} \right\} u \middle|_{(\frac{\gamma_j-1}{\delta_j}, \frac{1}{\delta_j}), j=1,\dots,k} \right] \end{aligned} \quad (2.5.9)$$

for $u \geq 0$, $\Re(\gamma_j + s - 1) > 0$, $j = 1, \dots, k$, where L is a suitable contour and the general H-function is defined as follows:

$$\begin{aligned} H(z) &= H_{p,q}^{m,n}(z) = H_{p,q}^{m,n} \left[z \middle|_{(b_1, \beta_1), \dots, (b_q, \beta_q)}^{(a_1, \alpha_1), \dots, (a_p, \alpha_p)} \right] \\ &= \frac{1}{2\pi i} \int_L \frac{\{\prod_{j=1}^m \Gamma(b_j + \beta_j s)\} \{\prod_{j=1}^n \Gamma(1 - a_j - \alpha_j s)\}}{\{\prod_{j=m+1}^q \Gamma(1 - b_j - \beta_j s)\} \{\prod_{j=n+1}^p \Gamma(a_j + \alpha_j s)\}} z^{-s} ds \end{aligned} \quad (2.5.10)$$

where $\alpha_j > 0$, $j = 1, \dots, p$; $\beta_j > 0$, $j = 1, \dots, q$ are real and positive, b_j 's and a_j 's are complex numbers, the contour L separates the poles of $\Gamma(b_j + \beta_j s)$, $j = 1, \dots, m$, lying on one side of it and the poles of $\Gamma(1 - a_j - \alpha_j s)$, $j = 1, \dots, n$, which must lie on the other side. The existence conditions and the various types of possible contours are discussed in Mathai and Saxena (1978) and Mathai et al. (2010). Observe that we can consider arbitrary powers of the variables present in u , u_1 , u_2 and u_3 as introduced in Sects. 2.5.1–2.5.5; however, in this case, the densities of these various structures will be expressible in terms of H-functions rather than G-functions. In the G-function format as defined in (2.5.2), the complex variable s has ± 1 as its coefficients, whereas the coefficients of s in the H-function, that is, $\pm \alpha_j$, $\alpha_j > 0$ and $\pm \beta_j$, $\beta_j > 0$, are not restricted to unities.

We will give a simple illustrative example that requires the evaluation of an inverse Mellin transform. Let $f(x) = e^{-x}$, $x > 0$. Then, the Mellin transform is

$$M_f(s) = \int_0^{\infty} x^{s-1} e^{-x} dx = \Gamma(s), \quad \Re(s) > 0,$$

and it follows from the inversion formula that

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s) x^{-s} ds, \quad \Re(s) > 0, \quad i = \sqrt{-1}. \quad (2.5.11)$$

If $f(x)$ is unknown and we are told that the Mellin transform of a certain function is $\Gamma(s)$, then are we going to retrieve $f(x)$ as e^{-x} from the inversion formula? Let us explore this problem. The poles of $\Gamma(s)$ occur at $s = 0, -1, -2, \dots$. Thus, if we take c in the contour of integration as $c > 0$, this contour will enclose all the poles of $\Gamma(s)$. We may now apply Cauchy's residue theorem. By definition, the residue at $s = -\nu$, denoted by R_ν , is

$$R_\nu = \lim_{s \rightarrow -\nu} (s + \nu) \Gamma(s) x^{-s}.$$

We cannot substitute $s = -\nu$ to obtain the limit in this case. However, noting that

$$(s + \nu) \Gamma(s) x^{-s} = \frac{(s + \nu)(s + \nu - 1) \cdots s \Gamma(s) x^{-s}}{(s + \nu - 1) \cdots s} = \frac{\Gamma(s + \nu + 1) x^{-s}}{(s + \nu - 1) \cdots s},$$

which follows from the recursive relationship, $\alpha \Gamma(\alpha) = \Gamma(\alpha + 1)$, the limit can be taken:

$$\begin{aligned} \lim_{s \rightarrow -\nu} (s + \nu) \Gamma(s) x^{-s} &= \lim_{s \rightarrow -\nu} \frac{\Gamma(s + \nu + 1) x^{-s}}{(s + \nu - 1) \cdots s} \\ &= \frac{\Gamma(1) x^\nu}{(-1)(-2) \cdots (-\nu)} = \frac{(-1)^\nu x^\nu}{\nu!}. \end{aligned} \quad (2.5.12)$$

Hence, the sum of the residues is

$$\sum_{\nu=0}^{\infty} R_\nu = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu x^\nu}{\nu!} = e^{-x},$$

and the function is recovered.

Note 2.5.1. Distributions of products and ratios of random variables in the complex domain could as well be worked out. However, since they may not necessarily have practical applications, they will not be discussed herein. Certain product and ratio distributions for variables in the complex domain which reduce to real variables, such as a chisquare in the complex domain, have already been previously discussed.

Exercises 2.5

2.5.1. Evaluate the density of $u = x_1 x_2$ where the x_j 's are independently distributed real type-1 beta random variables with the parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$, $j = 1, 2$ by using Mellin and inverse Mellin transform technique. Evaluate the density for the case $\alpha_1 - \alpha_2 \neq \pm v$, $v = 0, 1, \dots$ so that the poles are simple.

2.5.2. Repeat Exercise 2.5.1 if x_j 's are (1): real type-2 beta random variables with parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$ and (2): real gamma random variables with the parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$, $j = 1, 2$.

2.5.3. Let $u = \frac{u_1}{u_2}$ where u_1 and u_2 are real positive random variables. Then the h -th moment, for arbitrary h , is $E[\frac{u_1}{u_2}]^h \neq \frac{E[u_1^h]}{E[u_2^h]}$ in general. Give two examples where $E[\frac{u_1}{u_2}]^h = \frac{E[u_1^h]}{E[u_2^h]}$.

2.5.4. $E[\frac{1}{u}]^h = E[u^{-h}] \neq \frac{1}{E[u^h]}$ in general. Give two examples where $E[\frac{1}{u}] = \frac{1}{E[u]}$.

2.5.5. Let $u = \frac{x_1 x_2}{x_3 x_4}$ where the x_j 's are independently distributed. Let x_1, x_3 be type-1 beta random variables, x_2 be a type-2 beta random variable, and x_4 be a gamma random variable with parameters (α_j, β_j) , $\alpha_j > 0$, $\beta_j > 0$, $j = 1, 2, 3, 4$. Determine the density of u .

2.6. A Collection of Random Variables

Let x_1, \dots, x_n be iid (independently and identically distributed) real scalar random variables with a common density denoted by $f(x)$, that is, assume that the sample comes from the population that is specified by $f(x)$. Let the common mean value be μ and the common variance be $\sigma^2 < \infty$, that is, $E(x_j) = \mu$ and $\text{Var}(x_j) = \sigma^2$, $j = 1, \dots, n$, where E denotes the expected value. Denoting the sample average by $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$, what can be said about \bar{x} when $n \rightarrow \infty$? This is the type of questions that will be investigated in this section.

2.6.1. Chebyshev's inequality

For some $k > 0$, let us examine the probability content of $|x - \mu|$ where $\mu = E(x)$ and the variance of x is $\sigma^2 < \infty$. Consider the probability that the random variable x lies outside the interval $\mu - k\sigma < x < \mu + k\sigma$, that is k times the standard deviation σ away from the mean value μ . From the definition of the variance σ^2 for a real scalar random variable x ,

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\
&= \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu - k\sigma}^{\mu + k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 f(x) dx \\
&\geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 f(x) dx
\end{aligned}$$

since the probability content over the interval $\mu - k\sigma < x < \mu + k\sigma$ is omitted. Over this interval, the probability is either positive or zero, and hence the inequality. However, the intervals $(-\infty, \mu - k\sigma]$ and $[\mu + k\sigma, \infty)$, that is, $-\infty < x \leq \mu - k\sigma$ and $\mu + k\sigma \leq x < \infty$ or $-\infty < x - \mu \leq -k\sigma$ and $k\sigma \leq x - \mu < \infty$, can thus be described as the intervals for which $|x - \mu| \geq k\sigma$. In these intervals, the smallest value that $|x - \mu|$ can take on is $k\sigma$, $k > 0$ or equivalently, the smallest value that $|x - \mu|^2$ can assume is $(k\sigma)^2 = k^2\sigma^2$. Accordingly, the above inequality can be further sharpened as follows:

$$\begin{aligned}
\sigma^2 &\geq \int_{|x - \mu| \geq k\sigma} (x - \mu)^2 f(x) dx \geq \int_{|x - \mu| \geq k\sigma} (k\sigma)^2 f(x) dx \Rightarrow \\
\frac{\sigma^2}{k^2\sigma^2} &\geq \int_{|x - \mu| \geq k\sigma} f(x) dx \Rightarrow \\
\frac{1}{k^2} &\geq \int_{|x - \mu| \geq k\sigma} f(x) dx = Pr\{|x - \mu| \geq k\sigma\}, \text{ that is,} \\
\frac{1}{k^2} &\geq Pr\{|x - \mu| \geq k\sigma\},
\end{aligned}$$

which can be written as

$$Pr\{|x - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \text{ or } Pr\{|x - \mu| < k\sigma\} \geq 1 - \frac{1}{k^2}. \quad (2.6.1)$$

If $k\sigma = k_1$, $k = \frac{k_1}{\sigma^2}$, and the above inequalities can be written as follows:

$$Pr\{|x - \mu| \geq k\} \leq \frac{\sigma^2}{k^2} \text{ or } Pr\{|x - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}. \quad (2.6.2)$$

The inequalities (2.6.1) and (2.6.2) are known as Chebyshev's inequalities (also referred to as Chebycheff's inequalities). For example, when $k = 2$, Chebyshev's inequality states

that $Pr\{|x - \mu| < 2\sigma\} \geq 1 - \frac{1}{4} = 0.75$, which is not a very sharp probability limit. If $x \sim N_1(\mu, \sigma^2)$, then we know that

$$Pr\{|x - \mu| < 1.96\sigma\} \approx 0.95 \text{ and } Pr\{|x - \mu| < 3\sigma\} \approx 0.99.$$

Note that the bound 0.75 resulting from Chebyshev's inequality seriously underestimate the actual probability for a Gaussian variable x . However, what is astonishing about this inequality, is that the given probability bound holds for any distribution, whether it be continuous, discrete or mixed. Sharper bounds can of course be obtained for the probability content of the interval $[\mu - k\sigma, \mu + k\sigma]$ when the exact distribution of x is known.

These inequalities can be expressed in terms of generalized moments. Let $\mu_r^{\frac{1}{r}} = \{E|x - \mu|^r\}^{\frac{1}{r}}$, $r = 1, 2, \dots$, which happens to be a measure of scatter in x from the mean value μ . Given that

$$\mu_r = \int_{-\infty}^{\infty} |x - \mu|^r f(x) dx,$$

consider the probability content of the intervals specified by $|x - \mu| \geq k\mu_r^{\frac{1}{r}}$ for $k > 0$. Paralleling the derivations of (2.6.1) and (2.6.2), we have

$$\begin{aligned} \mu_r &\geq \int_{|x-\mu| \geq k\mu_r^{\frac{1}{r}}} |x - \mu|^r f(x) dx \geq \int_{|x-\mu| \geq k\mu_r^{\frac{1}{r}}} |(k\mu_r^{\frac{1}{r}})|^r f(x) dx \Rightarrow \\ Pr\{|x - \mu| \geq k\mu_r^{\frac{1}{r}}\} &\leq \frac{1}{k^r} \text{ or } Pr\{|x - \mu| < k\mu_r^{\frac{1}{r}}\} \geq 1 - \frac{1}{k^r}, \end{aligned} \quad (2.6.3)$$

which can also be written as

$$Pr\{|x - \mu| \geq k\} \leq \frac{\mu_r}{k^r} \text{ or } Pr\{|x - \mu| < k\} \geq 1 - \frac{\mu_r}{k^r}, \quad r = 1, 2, \dots \quad (2.6.4)$$

Note that when $r = 2$, $\mu_r = \sigma^2$, and Chebyshev's inequalities as specified in (2.6.1) and (2.6.2) are obtained from (2.6.3) and (2.6.4), respectively. If x is a real scalar positive random variables with $f(x) = 0$ for $x \leq 0$, we can then obtain similar inequalities in terms of the first moment μ . For $k > 0$,

$$\begin{aligned} \mu = E(x) &= \int_0^{\infty} xf(x) dx \text{ since } f(x) = 0 \text{ for } x \leq 0 \\ &= \int_0^k xf(x) dx + \int_k^{\infty} xf(x) dx \geq \int_k^{\infty} xf(x) dx \geq \int_k^{\infty} kf(x) dx \Rightarrow \\ \frac{\mu}{k} &\geq \int_k^{\infty} f(x) dx = Pr\{x \geq k\}. \end{aligned}$$

Accordingly, we have the following inequality for any real positive random variable x :

$$\Pr\{x \geq k\} \leq \frac{\mu}{k} \text{ for } x > 0, k > 0. \quad (2.6.5)$$

Suppose that our variable is $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$, where x_1, \dots, x_n are iid variables with common mean value μ and the common variance $\sigma^2 < \infty$. Then, since $\text{Var}(\bar{x}) = \frac{\sigma^2}{n}$ and $E(\bar{x}) = \mu$, Chebyshev's inequality states that

$$\Pr\{|\bar{x} - \mu| < k\} \geq 1 - \frac{\sigma^2}{n} \rightarrow 1 \text{ as } n \rightarrow \infty \quad (2.6.6)$$

or $\Pr\{|\bar{x} - \mu| \geq k\} \rightarrow 0$ as $n \rightarrow \infty$. However, since a probability cannot be greater than 1, $\Pr\{|\bar{x} - \mu| < k\} \rightarrow 1$ as $n \rightarrow \infty$. In other words, \bar{x} tends to μ with probability 1 as $n \rightarrow \infty$. This is referred to as the *Weak Law of Large Numbers*.

The Weak Law of Large Numbers

Let x_1, \dots, x_n be iid with common mean value μ and common variance $\sigma^2 < \infty$. Then, as $n \rightarrow \infty$,

$$\Pr\{\bar{x} \rightarrow \mu\} \rightarrow 1. \quad (2.6.7)$$

Another limiting property is known as the *Central Limit Theorem*. Let x_1, \dots, x_n be iid real scalar random variables with common mean value μ and common variance $\sigma^2 < \infty$. Letting $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ denote the sample mean, the standardized sample mean is

$$u = \frac{\bar{x} - E(\bar{x})}{\sqrt{\text{Var}(\bar{x})}} = \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu) = \frac{1}{\sigma\sqrt{n}}[(x_1 - \mu) + \dots + (x_n - \mu)]. \quad (i)$$

Consider the characteristic function of $x - \mu$, that is,

$$\begin{aligned} \phi_{x-\mu}(t) &= E[e^{it(x-\mu)}] = 1 + \frac{it}{1!}E(x-\mu) + \frac{(it)^2}{2!}E(x-\mu)^2 + \dots \\ &= 1 + 0 - \frac{t^2}{2!}E(x-\mu)^2 + \dots = 1 + \frac{t}{1!}\phi^{(1)}(0) + \frac{t^2}{2!}\phi^{(2)}(0) + \dots \quad (ii) \end{aligned}$$

where $\phi^{(r)}(0)$ is the r -th derivative of $\phi(t)$ with respect to t , evaluated at $t = 0$. Let us consider the characteristic function of our standardized sample mean u .

Making use of the last representation of u in (i), we have $\phi_{\frac{\sum_{j=1}^n(x_j-\mu)}{\sigma\sqrt{n}}}(t) = [\phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}})]^n$ so that $\phi_u(t) = [\phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}})]^n$ or $\ln \phi_u(t) = n \ln \phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}})$. It then

follows from (ii) that

$$\begin{aligned} [\phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}})] &= 1 + 0 - \frac{t^2}{2!} \frac{\sigma^2}{n\sigma^2} - i \frac{t^3}{3!} \frac{E(x_j - \mu)^3}{(\sigma\sqrt{n})^3} + \dots \\ &= 1 - \frac{t^2}{2n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right). \end{aligned} \tag{iii}$$

Now noting that $\ln(1 - y) = -[y + \frac{y^2}{2} + \frac{y^3}{3} + \dots]$ whenever $|y| < 1$, we have

$$\ln \phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}}) = -\frac{t^2}{2n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \Rightarrow n \ln \phi_{x_j-\mu}(\frac{t}{\sigma\sqrt{n}}) = -\frac{t^2}{2} + O\left(\frac{1}{n^{\frac{1}{2}}}\right) \rightarrow -\frac{t^2}{2} \text{ as } n \rightarrow \infty.$$

Consequently, as $n \rightarrow \infty$,

$$\phi_u(t) = e^{-\frac{t^2}{2}} \Rightarrow u \rightarrow N_1(0, 1) \text{ as } n \rightarrow \infty.$$

This is known as the central limit theorem.

The Central Limit Theorem. Let x_1, \dots, x_n be iid real scalar random variables having common mean value μ and common variance $\sigma^2 < \infty$. Let the sample mean be $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ and u denote the standardized sample mean. Then

$$u = \frac{\bar{x} - E(\bar{x})}{\sqrt{\text{Var}(\bar{x})}} = \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu) \rightarrow N_1(0, 1) \text{ as } n \rightarrow \infty. \tag{2.6.8}$$

Generalizations, extensions and more rigorous statements of this theorem are available in the literature. We have focussed on the substance of the result, assuming that a simple random sample is available and that the variance of the population is finite.

Exercises 2.6

2.6.1. For a binomial random variable with the probability function $f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$, $0 < p < 1$, $x = 0, 1, \dots, n$, $n = 1, 2, \dots$ and zero elsewhere, show that the standardized binomial variable itself, namely $\frac{x-np}{\sqrt{np(1-p)}}$ goes to the standard normal when $n \rightarrow \infty$.

2.6.2. State the central limit theorem for the following real scalar populations by evaluating the mean value and variance there, assuming that a simple random sample is available: (1) Poisson random variable with parameter λ ; (2) Geometric random variable with parameter p ; (3) Negative binomial random variable with parameters (p, k) ; (4) Discrete

hypergeometric probability law with parameters (a, b, n) ; (5) Uniform density over $[a, b]$; (6) Exponential density with parameter θ ; (7) Gamma density with the parameters (α, β) ; (8) Type-1 beta random variable with the parameters (α, β) ; (9) Type-2 beta random variable with the parameters (α, β) .

2.6.3. State the central limit theorem for the following probability/density functions: (1):

$$f(x) = \begin{cases} 0.5, & x = 2 \\ 0.5, & x = 5 \end{cases}$$

and $f(x) = 0$ elsewhere; (2): $f(x) = 2e^{-2x}$, $0 \leq x < \infty$ and zero elsewhere; (3): $f(x) = 1$, $0 \leq x \leq 1$ and zero elsewhere. Assume that a simple random sample is available from each population.

2.6.4. Consider a real scalar gamma random variable x with the parameters (α, β) and show that $E(x) = \alpha\beta$ and variance of x is $\alpha\beta^2$. Assume a simple random sample x_1, \dots, x_n from this population. Derive the densities of (1): $x_1 + \dots + x_n$; (2): \bar{x} ; (3): $\bar{x} - \alpha\beta$; (4): Standardized sample mean \bar{x} . Show that the densities in all these cases are still gamma densities, may be relocated, for all finite values of n however large n may be.

2.6.5. Consider the density $f(x) = \frac{c}{x^\alpha}$, $1 \leq x < \infty$ and zero elsewhere, where c is the normalizing constant. Evaluate c stating the relevant conditions. State the central limit theorem for this population, stating the relevant conditions.

2.7. Parameter Estimation: Point Estimation

There exist several methods for estimating the parameters of a given density/probability function, based on a simple random sample of size n (iid variables from the population designated by the density/probability function). The most popular methods of point estimation are the method of maximum likelihood and the method of moments.

2.7.1. The method of moments and the method of maximum likelihood

The likelihood function $L(\theta)$ is the joint density/probability function of the sample values, at an observed sample point, x_1, \dots, x_n . As a function of θ , $L(\theta)$, or a one-to-one function thereof, is maximized in order to determine the most likely value of θ in terms of a function of the given sample. This estimation process is referred to as the method of maximum likelihood.

Let $m_r = \frac{\sum_{j=1}^n x_j^r}{n}$ denote the r -th integer moment of the sample, where x_1, \dots, x_n is the observed sample point, the corresponding population r -th moment being $E[x^r]$, where E denotes the expected value. According to the method of moments, the estimates of the parameters are obtained by solving $m_r = E[x^r]$, $r = 1, 2, \dots$.

For example, consider a $N_1(\mu, \sigma^2)$ population with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0, \quad (2.7.1)$$

where μ and σ^2 are the parameters here. Let x_1, \dots, x_n be a simple random sample from this population. Then, the joint density of x_1, \dots, x_n , denoted by $L = L(x_1, \dots, x_n; \mu, \sigma^2)$, is

$$\begin{aligned} L &= \frac{1}{[\sqrt{2\pi}\sigma]^n} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2} = \frac{1}{[\sqrt{2\pi}\sigma]^n} e^{-\frac{1}{2\sigma^2} [\sum_{j=1}^n (x_j - \bar{x})^2 + n(\bar{x} - \mu)^2]} \Rightarrow \\ \ln L &= -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \left[\sum_{j=1}^n (x_j - \bar{x})^2 + n(\bar{x} - \mu)^2 \right], \quad \bar{x} = \frac{1}{n}(x_1 + \dots + x_n). \end{aligned} \quad (2.7.2)$$

Maximizing L or $\ln L$, since L and $\ln L$ are one-to-one functions, with respect to μ and $\theta = \sigma^2$, and solving for μ and σ^2 produces the maximum likelihood estimators (MLE's). An observed value of the estimator is the corresponding estimate. It follows from a basic result in Calculus that the extrema of L can be determined by solving the equations

$$\frac{\partial}{\partial \mu} \ln L = 0 \quad (i)$$

and

$$\frac{\partial}{\partial \theta} \ln L = 0, \quad \theta = \sigma^2. \quad (ii)$$

Equation (i) produces the solution $\mu = \bar{x}$ so that \bar{x} is the MLE of μ . Note that \bar{x} is a random variable and that \bar{x} evaluated at a sample point or at a set of observations on x_1, \dots, x_n produces the corresponding estimate. We will denote both the estimator and estimate of μ by $\hat{\mu}$. As well, we will utilize the same abbreviation, namely, MLE for the maximum likelihood estimator and the corresponding estimate. Solving (ii) and substituting $\hat{\mu}$ to μ , we have $\hat{\theta} = \hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 = s^2 =$ the sample variance as an estimate of $\theta = \sigma^2$. Does the point (\bar{x}, s^2) correspond to a local maximum or a local minimum or a saddle point? Since the matrix of second order partial derivatives at the point (\bar{x}, s^2) is

negative definite, the critical point (\bar{x}, s^2) corresponds to a maximum. Thus, in this case, $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = s^2$ are the maximum likelihood estimators/estimates of the parameters μ and σ^2 , respectively. If we were to differentiate with respect to σ instead of $\theta = \sigma^2$ in (ii), we would obtain the same estimators, since for any differentiable function $g(t)$, $\frac{d}{dt}g(t) = 0 \Rightarrow \frac{d}{d\phi(t)}g(t) = 0$ if $\frac{d}{dt}\phi(t) \neq 0$. In this instance, $\phi(\sigma) = \sigma^2$ and $\frac{d}{d\sigma}\sigma^2 \neq 0$.

For obtaining the moment estimates, we equate the sample integer moments to the corresponding population moments, that is, we let $m_r = E[x^r]$, $r = 1, 2$, two equations being required to estimate μ and σ^2 . Note that $m_1 = \bar{x}$ and $m_2 = \frac{1}{n} \sum_{j=1}^n x_j^2$. Then, consider the equations

$$\bar{x} = E[x] = \mu \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^n x_j^2 = E[x^2] \Rightarrow s^2 = \sigma^2.$$

Thus, the moment estimators/estimates of μ and σ^2 , which are $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = s^2$, happen to be identical to the MLE's in this case.

Let us consider the type-1 beta population with parameters (α, β) whose density is

$$f_1(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \quad \alpha > 0, \quad \beta > 0, \quad (2.7.3)$$

and zero otherwise. In this case, the likelihood function contains gamma functions and the derivatives of gamma functions involve psi and zeta functions. Accordingly, the maximum likelihood approach is not very convenient here. However, we can determine moment estimates without much difficulty from (2.7.3). The first two population integer moments are obtained directly from a representation of the h -th moment:

$$E[x^h] = \frac{\Gamma(\alpha + h)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + h)}, \quad \Re(\alpha + h) > 0 \Rightarrow$$

$$E[x] = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1)} = \frac{\alpha}{\alpha + \beta} \quad (2.7.4)$$

$$E[x^2] = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} = E[x] \frac{\alpha + 1}{\alpha + \beta + 1}. \quad (2.7.5)$$

Equating the sample moments to the corresponding population moments, that is, letting $m_1 = E[x]$ and $m_2 = E[x^2]$, it follows from (2.7.4) that

$$\bar{x} = \frac{\alpha}{\alpha + \beta} \Rightarrow \frac{\beta}{\alpha} = \frac{1 - \bar{x}}{\bar{x}}; \quad (iii)$$

Then, from (2.7.5), we have

$$\frac{\frac{1}{n} \sum_{j=1}^n x_j^2}{\bar{x}} = \frac{\alpha + 1}{\alpha + \beta + 1} = \frac{1}{1 + \frac{\beta}{\alpha + 1}} \Rightarrow \frac{\bar{x} - \sum_{j=1}^n x_j^2/n}{\sum_{j=1}^n x_j^2/n} = \frac{\beta}{\alpha + 1}. \quad (iv)$$

The parameter β can be eliminated from (iii) and (iv), which yields an estimate of α ; $\hat{\beta}$ is then obtained from (iii). Thus, the moment estimates are available from the equations, $m_r = E[x^r]$, $r = 1, 2$, even though these equations are nonlinear in the parameters α and β . The method of maximum likelihood or the method of moments can similarly yield parameters estimates for populations that are otherwise distributed.

2.7.2. Bayes' estimates

This procedure is more relevant when the parameters in a given statistical density/probability function have their own distributions. For example, let the real scalar variable x be discrete having a binomial probability law for the fixed (given) parameter p , that is, let $f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$, $0 < p < 1$, $x = 0, 1, \dots, n$, $n = 1, 2, \dots$, and $f(x|p) = 0$ elsewhere be the conditional probability function. Let p have a prior type-1 beta density with known parameters α and β , that is, let the prior density of p be

$$g(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad 0 < p < 1, \quad \alpha > 0, \quad \beta > 0$$

and $g(p) = 0$ elsewhere. Then, the joint probability function $f(x, p) = f(x|p)g(p)$ and the unconditional probability function of x , denoted by $f_1(x)$, is as follows:

$$\begin{aligned} f_1(x) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \binom{n}{x} \int_0^1 p^{\alpha+x-1} (1-p)^{\beta+n-x-1} dp \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \binom{n}{x} \frac{\Gamma(\alpha + x)\Gamma(\beta + n - x)}{\Gamma(\alpha + \beta + n)}. \end{aligned}$$

Thus, the posterior density of p , given x , denoted by $g_1(p|x)$, is

$$g_1(p|x) = \frac{f(x, p)}{f_1(x)} = \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} p^{\alpha+x-1} (1-p)^{\beta+n-x-1}.$$

Accordingly, the expected value of p in this conditional distribution of p given x , which is called the posterior density of p , is known as the Bayes estimate of p :

$$\begin{aligned}
E[p|x] &= \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} \int_0^1 p p^{\alpha+x-1} (1-p)^{\beta+n-x-1} dp \\
&= \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x)\Gamma(\beta + n - x)} \frac{\Gamma(\alpha + x + 1)\Gamma(\beta + n - x)}{\Gamma(\alpha + \beta + n + 1)} \\
&= \frac{\Gamma(\alpha + x + 1)}{\Gamma(\alpha + x)} \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + \beta + n + 1)} = \frac{\alpha + x}{\alpha + \beta + n}.
\end{aligned}$$

The prior estimate/estimator of p as obtained from the binomial distribution is $\frac{x}{n}$ and the posterior estimate or the Bayes estimate of p is

$$E[p|x] = \frac{\alpha + x}{\alpha + \beta + n},$$

so that $\frac{x}{n}$ is revised to $\frac{\alpha+x}{\alpha+\beta+n}$. In general, if the conditional density/probability function of x given θ is $f(x|\theta)$ and the prior density/probability function of θ is $g(\theta)$, then the posterior density of θ is $g_1(\theta|x)$ and $E[\theta|x]$ or *the expected value of θ in the conditional distribution of θ given x is the Bayes estimate of θ .*

2.7.3. Interval estimation

Before concluding this section, the concept of confidence intervals or interval estimation of a parameter will be briefly touched upon. For example, let x_1, \dots, x_n be iid $N_1(\mu, \sigma^2)$ and let $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$. Then, $\bar{x} \sim N_1(\mu, \frac{\sigma^2}{n})$, $(\bar{x} - \mu) \sim N_1(0, \frac{\sigma^2}{n})$ and $z = \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu) \sim N_1(0, 1)$. Since the standard normal density $N_1(0, 1)$ is free of any parameter, one can select two percentiles, say, a and b , from a standard normal table and make a probability statement such as $Pr\{a < z < b\} = 1 - \alpha$ for every given α ; for instance, $Pr\{-1.96 \leq z \leq 1.96\} \approx 0.95$ for $\alpha = 0.05$. Let $Pr\{-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}}\} = 1 - \alpha$ where $z_{\frac{\alpha}{2}}$ is such that $Pr(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. The following inequalities are mathematically equivalent and hence the probabilities associated with the corresponding intervals are equal:

$$\begin{aligned}
-z_{\frac{\alpha}{2}} \leq z \leq z_{\frac{\alpha}{2}} &\Leftrightarrow -z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu) \leq z_{\frac{\alpha}{2}} \\
&\Leftrightarrow \mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\
&\Leftrightarrow \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \tag{i}
\end{aligned}$$

Accordingly,

$$Pr\{\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\} = 1 - \alpha \tag{ii}$$

⇔

$$Pr\{\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\} = 1 - \alpha. \quad (iii)$$

Note that (iii) is not a usual probability statement as opposed to (ii), which is a probability statement on a random variable. In (iii), the interval $[\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$ is random and μ is a constant. This can be given the interpretation that the random interval covers the parameter μ with probability $1 - \alpha$, which means that we are $100(1 - \alpha)\%$ confident that the random interval will cover the unknown parameter μ or that the random interval is an interval estimator and an observed value of the interval is the interval estimate of μ . We could construct such an interval only because the distribution of z is parameter-free, which enabled us to make a probability statement on μ .

In general, if $u = u(x_1, \dots, x_n, \theta)$ is a function of the sample values and the parameter θ (which may be a vector of parameters) and if the distribution of u is free of all parameter, then such a quantity is referred to as a *pivotal quantity*. Since the distribution of the pivotal quantity is parameter-free, we can find two numbers a and b such that $Pr\{a \leq u \leq b\} = 1 - \alpha$ for every given α . If it is possible to convert the statement $a \leq u \leq b$ into a mathematically equivalent statement of the type $u_1 \leq \theta \leq u_2$, so that $Pr\{u_1 \leq \theta \leq u_2\} = 1 - \alpha$ for every given α , then $[u_1, u_2]$ is called a $100(1 - \alpha)\%$ *confidence interval or interval estimate for θ* , u_1 and u_2 being referred to as the *lower confidence limit* and the *upper confidence limit*, and $1 - \alpha$ being called the *confidence coefficient*. Additional results on interval estimation and the construction of confidence intervals are, for instance, presented in Mathai and Haubold (2017b).

Exercises 2.7

2.7.1. Obtain the method of moments estimators for the parameters (α, β) in a real type-2 beta population. Assume that a simple random sample of size n is available.

2.7.2. Obtain the estimate/estimator of the parameters by the method of moments and the method of maximum likelihood in the real (1): exponential population with parameter θ , (2): Poisson population with parameter λ . Assume that a simple random sample of size n is available.

2.7.3. Let x_1, \dots, x_n be a simple random sample of size n from a point Bernoulli population $f_2(x) = p^x(1 - p)^{1-x}$, $x = 0, 1$, $0 < p < 1$ and zero elsewhere. Obtain the MLE as well as moment estimator for p . [Note: These will be the same estimators for p in all the populations based on Bernoulli trials, such as binomial population, geometric population, negative binomial population].

2.7.4. If possible, obtain moment estimators for the parameters of a real generalized gamma population, $f_3(x) = c x^{\alpha-1} e^{-bx^\delta}$, $\alpha > 0$, $b > 0$, $\delta > 0$, $x \geq 0$ and zero elsewhere, where c is the normalizing constant.

2.7.5. If possible, obtain the MLE of the parameters a, b in the following real uniform population $f_4(x) = \frac{1}{b-a}$, $b > a$, $a \leq x < b$ and zero elsewhere. What are the MLE if $a < x < b$? What are the moment estimators in these two situations?

2.7.6. Construct the Bayes' estimate/estimator of the parameter λ in a Poisson probability law if the prior density for λ is a gamma density with known parameters (α, β) .

2.7.7. By selecting the appropriate pivotal quantities, construct a 95% confidence interval for (1): Poisson parameter λ ; (2): Exponential parameter θ ; (3): Normal parameter σ^2 ; (4): θ in a uniform density $f(x) = \frac{1}{\theta}$, $0 \leq x \leq \theta$ and zero elsewhere.

References

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford, 1993.

A.M. Mathai (1999): *Introduction to Geometrical Probability: Distributional Aspects and Applications*, Gordon and Breach, Amsterdam, 1999.

A.M. Mathai (2005): A pathway to matrix-variate gamma and normal densities, *Linear Algebra and its Applications*, **410**, 198-216.

A.M. Mathai (2013): Fractional integral operators in the complex matrix-variate case, *Linear Algebra and its Applications*, **439**, 2901-2913.

A.M. Mathai (2014): Fractional integral operators involving many matrix variables, *Linear Algebra and its Applications*, **446**, 196-215.

A.M. Mathai (2015): Fractional differential operators in the complex matrix-variate case, *Linear Algebra and its Applications*, **478**, 200-217.

A.M. Mathai and Hans J. Haubold (1988): *Modern Problems in Nuclear and Neutrino Astrophysics*, Akademie-Verlag, Berlin.

A.M. Mathai and Hans J. Haubold (2017a): *Linear Algebra, A course for Physicists and Engineers*, De Gruyter, Germany, 2017.

A.M. Mathai and Hans J. Haubold (2017b): *Probability and Statistics, A course for Physicists and Engineers*, De Gruyter, Germany, 2017.

A.M. Mathai and Hans J. Haubold (2018): *Erdélyi-Kober Fractional Calculus from a Statistical Perspective, Inspired by Solar Neutrino Physics*, Springer Briefs in Mathematical Physics, 31, Springer Nature, Singapore, 2018.

A.M. Mathai and Hans J. Haubold (2020): Mathematical aspects of Krätzel integral and Krätzel transform, *Mathematics*, **8(526)**; doi:10.3390/math8040526.

A.M. Mathai and Serge B. Provost (1992): *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, New York, 1992.

A.M. Mathai and Serge B. Provost (2006): On q-logistic and related distributions, *IEEE Transactions on Reliability*, **55(2)**, 237-244.

A.M. Mathai and R.K. Saxena (1973): *Generalized Hypergeometric Functions*, Springer Lecture Notes No. 348, Heidelberg, 1973.

A.M. Mathai and R. K. Saxena (1978): *The H-function with Applications in Statistics and Other Disciplines*, Wiley Halsted, New York, 1978.

A.M. Mathai, R.K. Saxena and Hans J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York, 2010.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

The Multivariate Gaussian and Related Distributions



3.1. Introduction

Real scalar mathematical as well as random variables will be denoted by lower-case letters such as x, y, z , and vector/matrix variables, whether mathematical or random, will be denoted by capital letters such as X, Y, Z , in the real case. Complex variables will be denoted with a tilde: $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$, for instance. Constant matrices will be denoted by A, B, C , and so on. A tilde will be placed above constant matrices *only* if one wishes to stress the point that the matrix is in the complex domain. Equations will be numbered chapter and section-wise. Local numbering will be done subsection-wise. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value of the determinant of A will be denoted as $|\det(A)|$. Observe that in the complex domain, $\det(A) = a + ib$ where a and b are real scalar quantities, and then, $|\det(A)|^2 = a^2 + b^2$.

Multivariate usually refers to a collection of scalar variables. Vector/matrix variable situations are also of the multivariate type but, in addition, the positions of the variables must also be taken into account. In a function involving a matrix, one cannot permute its elements since each permutation will produce a different matrix. For example,

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix}, \tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{12} & \tilde{x}_{22} \end{bmatrix}$$

are all multivariate cases but the elements or the individual variables must remain at the set positions in the matrices.

The definiteness of matrices will be needed in our discussion. Definiteness is defined and discussed only for symmetric matrices in the real domain and Hermitian matrices in the complex domain. Let $A = A'$ be a real $p \times p$ matrix and Y be a $p \times 1$ real vector, Y' denoting its transpose. Consider the quadratic form $Y'AY$, $A = A'$, for all possible Y excluding the null vector, that is, $Y \neq O$. We say that the real quadratic form $Y'AY$ as well

as the real matrix $A = A'$ are positive definite, which is denoted $A > O$, if $Y'AY > 0$, for all possible non-null Y . Letting $A = A'$ be a real $p \times p$ matrix, if for all real $p \times 1$ vector $Y \neq O$,

$$\begin{aligned} Y'AY > 0, A > O & \text{ (positive definite)} \\ Y'AY \geq 0, A \geq O & \text{ (positive semi-definite)} \\ Y'AY < 0, A < O & \text{ (negative definite)} \\ Y'AY \leq 0, A \leq O & \text{ (negative semi-definite)}. \end{aligned} \tag{3.1.1}$$

All the matrices that do not belong to any one of the above categories are said to be *indefinite matrices*, in which case A will have both positive and negative eigenvalues. For example, for some Y , $Y'AY$ may be positive and for some other values of Y , $Y'AY$ may be negative. The definiteness of Hermitian matrices can be defined in a similar manner. A square matrix A in the complex domain is called Hermitian if $A = A^*$ where A^* means the conjugate transpose of A . Either the conjugates of all the elements of A are taken and the matrix is then transposed or the matrix A is first transposed and the conjugate of each of its elements is then taken. If $\tilde{z} = a + ib$, $i = \sqrt{-1}$ and a, b real scalar, then the conjugate of \tilde{z} , conjugate being denoted by a bar, is $\bar{\tilde{z}} = a - ib$, that is, i is replaced by $-i$. For instance, since

$$B = \begin{bmatrix} 2 & 1+i \\ 1-i & 5 \end{bmatrix} \Rightarrow \bar{B} = \begin{bmatrix} 2 & 1-i \\ 1+i & 5 \end{bmatrix} \Rightarrow (\bar{B})' = \bar{B}' = \begin{bmatrix} 2 & 1+i \\ 1-i & 5 \end{bmatrix} = B^*,$$

$B = B^*$, and thus the matrix B is Hermitian. In general, if \tilde{X} is a $p \times p$ matrix, then, \tilde{X} can be written as $\tilde{X} = X_1 + iX_2$ where X_1 and X_2 are real matrices and $i = \sqrt{-1}$. And if $\tilde{X} = X^*$ then $\tilde{X} = X_1 + iX_2 = X^* = X_1' - iX_2'$ or X_1 is symmetric and X_2 is skew symmetric so that all the diagonal elements of a Hermitian matrix are real. The definiteness of a Hermitian matrix can be defined parallel to that in the real case. Let $A = A^*$ be a Hermitian matrix. In the complex domain, definiteness is defined only for Hermitian matrices. Let $Y \neq O$ be a $p \times 1$ non-null vector and let Y^* be its conjugate transpose. Then, consider the Hermitian form Y^*AY , $A = A^*$. If $Y^*AY > 0$ for all possible non-null $Y \neq O$, the Hermitian form Y^*AY , $A = A^*$ as well as the Hermitian matrix A are said to be positive definite, which is denoted $A > O$. Letting $A = A^*$, if for all non-null Y ,

$$\begin{aligned} Y^*AY > 0, A > O & \text{ (Hermitian positive definite)} \\ Y^*AY \geq 0, A \geq O & \text{ (Hermitian positive semi-definite)} \\ Y^*AY < 0, A < O & \text{ (Hermitian negative definite)} \\ Y^*AY \leq 0, A \leq O & \text{ (Hermitian negative semi-definite)}, \end{aligned} \tag{3.1.2}$$

and when none of the above cases applies, we have *indefinite matrices or indefinite Hermitian forms*.

We will also make use of properties of the square root of matrices. If we were to define the square root of A as B such as $B^2 = A$, there would then be several candidates for B . Since a multiplication of A with A is involved, A has to be a square matrix. Consider the following matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$A_4 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, A_5 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

whose squares are all equal to I_2 . Thus, there are clearly several candidates for the square root of this identity matrix. However, if we restrict ourselves to the class of positive definite matrices in the real domain and Hermitian positive definite matrices in the complex domain, then we can define a unique square root, denoted by $A^{\frac{1}{2}} > O$.

For the various Jacobians used in this chapter, the reader may refer to Chap. 1, further details being available from Mathai (1997).

3.1a. The Multivariate Gaussian Density in the Complex Domain

Consider the complex scalar random variables $\tilde{x}_1, \dots, \tilde{x}_p$. Let $\tilde{x}_j = x_{j1} + ix_{j2}$ where x_{j1}, x_{j2} are real and $i = \sqrt{-1}$. Let $E[x_{j1}] = \mu_{j1}$, $E[x_{j2}] = \mu_{j2}$ and $E[\tilde{x}_j] = \mu_{j1} + i\mu_{j2} \equiv \tilde{\mu}_j$. Let the variances be as follows: $\text{Var}(x_{j1}) = \sigma_{j1}^2$, $\text{Var}(x_{j2}) = \sigma_{j2}^2$. For a complex variable, the variance is defined as follows:

$$\begin{aligned} \text{Var}(\tilde{x}_j) &= E[\tilde{x}_j - E(\tilde{x}_j)][\tilde{x}_j - E(\tilde{x}_j)]^* \\ &= E[(x_{j1} - \mu_{j1}) + i(x_{j2} - \mu_{j2})][(x_{j1} - \mu_{j1}) - i(x_{j2} - \mu_{j2})] \\ &= E[(x_{j1} - \mu_{j1})^2 + (x_{j2} - \mu_{j2})^2] = \text{Var}(x_{j1}) + \text{Var}(x_{j2}) = \sigma_{j1}^2 + \sigma_{j2}^2 \\ &\equiv \sigma_j^2. \end{aligned}$$

A covariance matrix associated with the $p \times 1$ vector $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_p)'$ in the complex domain is defined as $\text{Cov}(\tilde{X}) = E[\tilde{X} - E(\tilde{X})][\tilde{X} - E(\tilde{X})]^* \equiv \Sigma$ with $E(\tilde{X}) \equiv \tilde{\mu} = (\tilde{\mu}_1, \dots, \tilde{\mu}_p)'$. Then we have

$$\Sigma = \text{Cov}(\tilde{X}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix}$$

where the covariance between \tilde{x}_r and \tilde{x}_s , two distinct elements in \tilde{X} , requires explanation. Let $\tilde{x}_r = x_{r1} + ix_{r2}$ and $\tilde{x}_s = x_{s1} + ix_{s2}$ where x_{r1} , x_{r2} , x_{s1} , x_{s2} are all real. Then, the covariance between \tilde{x}_r and \tilde{x}_s is

$$\begin{aligned} \text{Cov}(\tilde{x}_r, \tilde{x}_s) &= E[\tilde{x}_r - E(\tilde{x}_r)][\tilde{x}_s - E(\tilde{x}_s)]^* = \text{Cov}[(x_{r1} + ix_{r2}), (x_{s1} - ix_{s2})] \\ &= \text{Cov}(x_{r1}, x_{s1}) + \text{Cov}(x_{r2}, x_{s2}) + i[\text{Cov}(x_{r2}, x_{s1}) - \text{Cov}(x_{r1}, x_{s2}) = \sigma_{rs}]. \end{aligned}$$

Note that none of the individual covariances on the right-hand side need be equal to each other. Hence, σ_{rs} need not be equal to σ_{sr} . In terms of vectors, we have the following: Let $\tilde{X} = X_1 + iX_2$ where X_1 and X_2 are real vectors. The covariance matrix associated with \tilde{X} , which is denoted by $\text{Cov}(\tilde{X})$, is

$$\begin{aligned} \text{Cov}(\tilde{X}) &= E([\tilde{X} - E(\tilde{X})][\tilde{X} - E(\tilde{X})]^*) \\ &= E([(X_1 - E(X_1)) + i(X_2 - E(X_2))][(X_1' - E(X_1')) - i(X_2' - E(X_2'))]) \\ &= \text{Cov}(X_1, X_1) + \text{Cov}(X_2, X_2) + i[\text{Cov}(X_2, X_1) - \text{Cov}(X_1, X_2)] \\ &\equiv \Sigma_{11} + \Sigma_{22} + i[\Sigma_{21} - \Sigma_{12}] \end{aligned}$$

where Σ_{12} need not be equal to Σ_{21} . Hence, in general, $\text{Cov}(X_1, X_2)$ need not be equal to $\text{Cov}(X_2, X_1)$. We will denote the whole configuration as $\text{Cov}(\tilde{X}) = \Sigma$ and assume it to be Hermitian positive definite. We will define the p -variate Gaussian density in the complex domain as the following real-valued function:

$$f(\tilde{X}) = \frac{1}{\pi^p |\det(\Sigma)|} e^{-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1} (\tilde{X} - \tilde{\mu})} \quad (3.1a.1)$$

where $|\det(\Sigma)|$ denotes the absolute value of the determinant of Σ . Let us verify that the normalizing constant is indeed $\frac{1}{\pi^p |\det(\Sigma)|}$. Consider the transformation $\tilde{Y} = \Sigma^{-\frac{1}{2}} (\tilde{X} - \tilde{\mu})$ which gives $d\tilde{X} = [\det(\Sigma \Sigma^*)]^{\frac{1}{2}} d\tilde{Y} = |\det(\Sigma)| d\tilde{Y}$ in light of (1.6a.1). Then $|\det(\Sigma)|$ is canceled and the exponent becomes $-\tilde{Y}^* \tilde{Y} = -[|\tilde{y}_1|^2 + \dots + |\tilde{y}_p|^2]$. But

$$\int_{\tilde{y}_j} e^{-|\tilde{y}_j|^2} d\tilde{y}_j = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y_{j1}^2 + y_{j2}^2)} dy_{j1} \wedge dy_{j2} = \pi, \quad \tilde{y}_j = y_{j1} + iy_{j2}, \quad (i)$$

which establishes the normalizing constant. Let us examine the mean value and the covariance matrix of \tilde{X} in the complex case. Let us utilize the same transformation, $\Sigma^{-\frac{1}{2}} (\tilde{X} - \tilde{\mu})$. Accordingly,

$$E[\tilde{X}] = \tilde{\mu} + E[(\tilde{X} - \tilde{\mu})] = \tilde{\mu} + \Sigma^{\frac{1}{2}} E[\tilde{Y}].$$

However,

$$E[\tilde{Y}] = \frac{1}{\pi^p} \int_{\tilde{Y}} \tilde{Y} e^{-\tilde{Y}^* \tilde{Y}} d\tilde{Y},$$

and the integrand has each element in \tilde{Y} producing an odd function whose integral converges, so that the integral over \tilde{Y} is null. Thus, $E[\tilde{X}] = \tilde{\mu}$, the first parameter appearing in the exponent of the density (3.1a.1). Now the covariance matrix in \tilde{X} is the following:

$$\text{Cov}(\tilde{X}) = E[(\tilde{X} - E(\tilde{X}))(\tilde{X} - E(\tilde{X}))^*] = \Sigma^{\frac{1}{2}} E[\tilde{Y}\tilde{Y}^*] \Sigma^{\frac{1}{2}}.$$

We consider the integrand in $E[\tilde{Y}\tilde{Y}^*]$ and follow steps parallel to those used in the real case. It is a $p \times p$ matrix where the non-diagonal elements are odd functions whose integrals converge and hence each of these elements will integrate out to zero. The first diagonal element in $\tilde{Y}\tilde{Y}^*$ is $|\tilde{y}_1|^2$. Its associated integral is

$$\begin{aligned} & \int \dots \int |\tilde{y}_1|^2 e^{-(|\tilde{y}_1|^2 + \dots + |\tilde{y}_p|^2)} d\tilde{y}_1 \wedge \dots \wedge d\tilde{y}_p \\ &= \left\{ \prod_{j=2}^p \int_{\tilde{y}_j} e^{-|\tilde{y}_j|^2} d\tilde{y}_j \right\} \int_{\tilde{y}_1} |\tilde{y}_1|^2 e^{-|\tilde{y}_1|^2} d\tilde{y}_1. \end{aligned}$$

From (i),

$$\int_{\tilde{y}_1} |\tilde{y}_1|^2 e^{-|\tilde{y}_1|^2} d\tilde{y}_1 = \pi; \quad \prod_{j=2}^p \int_{\tilde{y}_j} e^{-|\tilde{y}_j|^2} d\tilde{y}_j = \pi^{p-1},$$

where $|\tilde{y}_1|^2 = y_{11}^2 + y_{12}^2$, $\tilde{y}_1 = y_{11} + iy_{12}$, $i = \sqrt{-1}$, and y_{11} , y_{12} real. Let $y_{11} = r \cos \theta$, $y_{12} = r \sin \theta \Rightarrow dy_{11} \wedge dy_{12} = r dr \wedge d\theta$ and

$$\begin{aligned} \int_{\tilde{y}_1} |\tilde{y}_1|^2 e^{-|\tilde{y}_1|^2} d\tilde{y}_1 &= \left(\int_{r=0}^{\infty} r(r^2) e^{-r^2} dr \right) \left(\int_{\theta=0}^{2\pi} d\theta \right), \quad (\text{letting } u = r^2) \\ &= (2\pi) \left(\frac{1}{2} \int_0^{\infty} u e^{-u} du \right) = (2\pi) \left(\frac{1}{2} \right) = \pi. \end{aligned}$$

Thus the first diagonal element in $\tilde{Y}\tilde{Y}^*$ integrates out to π^p and, similarly, each diagonal element will integrate out to π^p , which is canceled by the term π^p present in the normalizing constant. Hence the integral over $\tilde{Y}\tilde{Y}^*$ gives an identity matrix and the covariance matrix of \tilde{X} is Σ , the other parameter appearing in the density (3.1a.1). Hence the two parameters therein are the mean value vector and the covariance matrix of \tilde{X} .

Example 3.1a.1. Consider the matrix Σ and the vector \tilde{X} with expected value $E[\tilde{X}] = \tilde{\mu}$ as follows:

$$\Sigma = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}, \quad E[\tilde{X}] = \begin{bmatrix} 1+2i \\ 2-i \end{bmatrix} = \tilde{\mu}.$$

Show that Σ is Hermitian positive definite so that it can be a covariance matrix of \tilde{X} , that is, $\text{Cov}(\tilde{X}) = \Sigma$. If \tilde{X} has a bivariate Gaussian distribution in the complex domain; $\tilde{X} \sim \tilde{N}_2(\tilde{\mu}, \Sigma)$, $\Sigma > O$, then write down (1) the exponent in the density explicitly; (2) the density explicitly.

Solution 3.1a.1. The transpose and conjugate transpose of Σ are

$$\Sigma' = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}, \quad \Sigma^* = \bar{\Sigma}' = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} = \Sigma$$

and hence Σ is Hermitian. The eigenvalues of Σ are available from the equation

$$(2-\lambda)(3-\lambda) - (1-i)(1+i) = 0 \Rightarrow \lambda^2 - 5\lambda + 4 = 0 \\ \Rightarrow (\lambda-4)(\lambda-1) \text{ or } \lambda_1 = 4, \lambda_2 = 1.$$

Thus, the eigenvalues are positive [the eigenvalues of a Hermitian matrix will always be real]. This property of eigenvalues being positive, combined with the property that Σ is Hermitian proves that Σ is Hermitian positive definite. This can also be established from the leading minors of Σ . The leading minors are $\det((2)) = 2 > 0$ and $\det(\Sigma) = (2)(3) - (1-i)(1+i) = 4 > 0$. Since Σ is Hermitian and its leading minors are all positive, Σ is positive definite. Let us evaluate the inverse by making use of the formula $\Sigma^{-1} = \frac{1}{\det(\Sigma)}(\text{Cof}(\Sigma))'$ where $\text{Cof}(\Sigma)$ represents the matrix of cofactors of the elements in Σ . [These formulae hold whether the elements in the matrix are real or complex]. That is,

$$\Sigma^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix}, \quad \Sigma^{-1}\Sigma = I. \quad (ii)$$

The exponent in a bivariate complex Gaussian density being $-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu})$, we have

$$-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu}) = -\frac{1}{4} \{ 3 [\tilde{x}_1 - (1+2i)]^* [\tilde{x}_1 - (1+2i)] \\ - (1+i) [\tilde{x}_1 - (1+2i)]^* [\tilde{x}_2 - (2-i)] \\ - (1-i) [\tilde{x}_2 - (2-i)]^* [\tilde{x}_1 - (1+2i)] \\ + 2 [\tilde{x}_2 - (2-i)]^* [\tilde{x}_2 - (2-i)] \}. \quad (iii)$$

Thus, the density of the $\tilde{N}_2(\tilde{\mu}, \Sigma)$ vector whose components can assume any complex value is

$$f(\tilde{X}) = \frac{e^{-(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})}}{4\pi^2} \quad (3.1a.2)$$

where Σ^{-1} is given in (ii) and the exponent, in (iii).

Exercises 3.1

3.1.1. Construct a 2×2 Hermitian positive definite matrix A and write down a Hermitian form with this A as its matrix.

3.1.2. Construct a 2×2 Hermitian matrix B where the determinant is 4, the trace is 5, and first row is $2, 1 + i$. Then write down explicitly the Hermitian form X^*BX .

3.1.3. Is B in Exercise 3.1.2 positive definite? Is the Hermitian form X^*BX positive definite? Establish the results.

3.1.4. Construct two 2×2 Hermitian matrices A and B such that $AB = O$ (null), if that is possible.

3.1.5. Specify the eigenvalues of the matrix B in Exercise 3.1.2, obtain a unitary matrix Q , $QQ^* = I$, $Q^*Q = I$ such that Q^*BQ is diagonal and write down the canonical form for a Hermitian form $X^*BX = \lambda_1|y_1|^2 + \lambda_2|y_2|^2$.

3.2. The Multivariate Normal or Gaussian Distribution, Real Case

We may define a real p -variate Gaussian density via the following characterization: Let x_1, \dots, x_p be real scalar variables and X be a $p \times 1$ vector with x_1, \dots, x_p as its elements, that is, $X' = (x_1, \dots, x_p)$. Let $L' = (a_1, \dots, a_p)$ where a_1, \dots, a_p are arbitrary real scalar constants. Consider the linear function $u = L'X = X'L = a_1x_1 + \dots + a_px_p$. If, for all possible L , $u = L'X$ has a real univariate Gaussian distribution, then the vector X is said to have a multivariate Gaussian distribution. For any linear function $u = L'X$, $E[u] = L'E[X] = L'\mu$, $\mu' = (\mu_1, \dots, \mu_p)$, $\mu_j = E[x_j]$, $j = 1, \dots, p$, and $\text{Var}(u) = L'\Sigma L$, $\Sigma = \text{Cov}(X) = E[X - E(X)][X - E(X)]'$ in the real case. If u is univariate normal then its mgf, with parameter t , is the following:

$$M_u(t) = E[e^{tu}] = e^{tE(u) + \frac{t^2}{2}\text{Var}(u)} = e^{tL'\mu + \frac{t^2}{2}L'\Sigma L}.$$

Note that $tL'\mu + \frac{t^2}{2}L'\Sigma L = (tL)'\mu + \frac{1}{2}(tL)'\Sigma(tL)$ where there are p parameters a_1, \dots, a_p when the a_j 's are arbitrary. As well, tL contains only p parameters as, for example, ta_j is a single parameter when both t and a_j are arbitrary. Then,

$$M_u(t) = M_X(tL) = e^{(tL)'\mu + \frac{1}{2}(tL)'\Sigma(tL)} = e^{T'\mu + \frac{1}{2}T'\Sigma T} = M_X(T), \quad T = tL. \quad (3.2.1)$$

Thus, when L is arbitrary, the mgf of u qualifies to be the mgf of a p -vector X . The density corresponding to (3.2.1) is the following, when $\Sigma > O$:

$$f(X) = c e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}, \quad -\infty < x_j < \infty, \quad -\infty < \mu_j < \infty, \quad \Sigma > O$$

for $j = 1, \dots, p$. We can evaluate the normalizing constant c when $f(X)$ is a density, in which case the total integral is unity. That is,

$$1 = \int_X f(X) dX = \int_X c e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} dX.$$

Let $\Sigma^{-\frac{1}{2}}(X - \mu) = Y \Rightarrow dY = |\Sigma|^{-\frac{1}{2}}d(X - \mu) = |\Sigma|^{-\frac{1}{2}}dX$ since μ is a constant. The Jacobian of the transformation may be obtained from Theorem 1.6.1. Now,

$$1 = c|\Sigma|^{\frac{1}{2}} \int_Y e^{-\frac{1}{2}Y'Y} dY.$$

But $Y'Y = y_1^2 + \dots + y_p^2$ where y_1, \dots, y_p are the real elements in Y and $\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_j^2} dy_j = \sqrt{2\pi}$. Hence $\int_Y e^{-\frac{1}{2}Y'Y} dY = (\sqrt{2\pi})^p$. Then $c = [|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}]^{-1}$ and the p -variate real Gaussian or normal density is given by

$$f(X) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} \quad (3.2.2)$$

for $\Sigma > O$, $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $j = 1, \dots, p$. The density (3.2.2) is called the nonsingular normal density in the real case—nonsingular in the sense that Σ is nonsingular. In fact, Σ is also real positive definite in the nonsingular case. When Σ is singular, we have a singular normal distribution which does not have a density function. However, in the singular case, all the properties can be studied with the help of the associated mgf which is of the form in (3.2.1), as the mgf exists whether Σ is nonsingular or singular.

We will use the standard notation $X \sim N_p(\mu, \Sigma)$ to denote a p -variate real normal or Gaussian distribution with mean value vector μ and covariance matrix Σ . If it is nonsingular real Gaussian, we write $\Sigma > O$; if it is singular normal, then we specify $|\Sigma| = 0$. If we wish to combine the singular and nonsingular cases, we write $X \sim N_p(\mu, \Sigma)$, $\Sigma \geq O$.

What are the mean value vector and the covariance matrix of a real p -Gaussian vector X ?

$$\begin{aligned} E[X] &= E[X - \mu] + E[\mu] = \mu + \int_X (X - \mu) f(X) dX \\ &= \mu + \frac{1}{|\Sigma|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \int_X (X - \mu) e^{-\frac{1}{2}(X-\mu)' \Sigma^{-1} (X-\mu)} dX \\ &= \mu + \frac{\Sigma^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \int_Y Y e^{-\frac{1}{2}Y'Y} dY, \quad Y = \Sigma^{-\frac{1}{2}}(X - \mu). \end{aligned}$$

The expected value of a matrix is the matrix of the expected value of every element in the matrix. The expected value of the component y_j of $Y' = (y_1, \dots, y_p)$ is

$$E[y_j] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_j e^{-\frac{1}{2}y_j^2} dy_j \left\{ \prod_{i \neq j=1}^p \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i \right\}.$$

The product is equal to 1 and the first integrand being an odd function of y_j , it is equal to 0 since integral is convergent. Thus, $E[Y] = O$ (a null vector) and $E[X] = \mu$, the first parameter appearing in the exponent of the density. Now, consider the covariance matrix of X . For a vector real X ,

$$\begin{aligned} \text{Cov}(X) &= E[X - E(X)][X - E(X)]' = E[(X - \mu)(X - \mu)'] \\ &= \frac{1}{|\Sigma|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \int_X (X - \mu)(X - \mu)' e^{-\frac{1}{2}(X-\mu)' \Sigma^{-1} (X-\mu)} dX \\ &= \frac{1}{(2\pi)^{\frac{p}{2}}} \Sigma^{\frac{1}{2}} \left[\int_Y Y Y' e^{-\frac{1}{2}Y'Y} dY \right] \Sigma^{\frac{1}{2}}, \quad Y = \Sigma^{-\frac{1}{2}}(X - \mu). \end{aligned}$$

But

$$Y Y' = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} [y_1, \dots, y_p] = \begin{bmatrix} y_1^2 & y_1 y_2 & \cdots & y_1 y_p \\ y_2 y_1 & y_2^2 & \cdots & y_2 y_p \\ \vdots & \vdots & \ddots & \vdots \\ y_p y_1 & y_p y_2 & \cdots & y_p^2 \end{bmatrix}.$$

The non-diagonal elements are linear in each variable y_i and y_j , $i \neq j$ and hence the integrals over the non-diagonal elements will be equal to zero due to a property of convergent integrals over odd functions. Hence we only need to consider the diagonal elements. When considering y_1 , the integrals over y_2, \dots, y_p will give the following:

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_j^2} dy_j = \sqrt{2\pi}, \quad j = 2, \dots, p \Rightarrow (2\pi)^{\frac{p-1}{2}}$$

and hence we are left with

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_1^2 e^{-\frac{1}{2}y_1^2} dy_1 = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} y_1^2 e^{-\frac{1}{2}y_1^2} dy_1$$

due to evenness of the integrand, the integral being convergent. Let $u = y_1^2$ so that $y_1 = u^{\frac{1}{2}}$ since $y_1 > 0$. Then $dy_1 = \frac{1}{2}u^{\frac{1}{2}-1}du$. The integral is available as $\Gamma(\frac{3}{2})2^{\frac{3}{2}} = \frac{1}{2}\Gamma(\frac{1}{2})2^{\frac{3}{2}} = \sqrt{2\pi}$ since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and the constant is canceled leaving 1. This shows that each diagonal element integrates out to 1 and hence the integral over YY' is the identity matrix after absorbing $(2\pi)^{-\frac{p}{2}}$. Thus $\text{Cov}(X) = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$ the inverse of which is the other parameter appearing in the exponent of the density. Hence the two parameters are

$$\mu = E[X] \text{ and } \Sigma = \text{Cov}(X). \quad (3.2.3)$$

The bivariate case

When $p = 2$, we obtain the bivariate real normal density from (3.2.2), which is denoted by $f(x_1, x_2)$. Note that when $p = 2$,

$$(X - \mu)' \Sigma^{-1} (X - \mu) = (x_1 - \mu_1, x_2 - \mu_2) \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

where $\sigma_1^2 = \text{Var}(x_1) = \sigma_{11}$, $\sigma_2^2 = \text{Ver}(x_2) = \sigma_{22}$, $\sigma_{12} = \text{Cov}(x_1, x_2) = \sigma_1\sigma_2\rho$ where ρ is the correlation between x_1 and x_2 , and ρ , in general, is defined as

$$\rho = \frac{\text{Cov}(x_1, x_2)}{\sqrt{\text{Var}(x_1)\text{Var}(x_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}, \quad \sigma_1 \neq 0, \quad \sigma_2 \neq 0,$$

which means that ρ is defined only for non-degenerate random variables, or equivalently, that the probability mass of either variable should not lie at a single point. This ρ is a scale-free covariance, the covariance measuring the joint variation in (x_1, x_2) corresponding to the square of scatter, $\text{Var}(x)$, in a real scalar random variable x . The covariance, in general, depends upon the units of measurements of x_1 and x_2 , whereas ρ is a scale-free pure coefficient. This ρ does not measure relationship between x_1 and x_2 for $-1 < \rho < 1$. But for $\rho = \pm 1$ it can measure linear relationship. Oftentimes, ρ is misinterpreted as measuring any relationship between x_1 and x_2 , which is not the case as can be seen from

the counterexamples pointed out in Mathai and Haubold (2017). If $\rho_{x,y}$ is the correlation between two real scalar random variables x and y and if $u = a_1x + b_1$ and $v = a_2y + b_2$ where $a_1 \neq 0, a_2 \neq 0$ and b_1, b_2 are constants, then $\rho_{u,v} = \pm\rho_{x,y}$. It is positive when $a_1 > 0, a_2 > 0$ or $a_1 < 0, a_2 < 0$ and negative otherwise. Thus, ρ is both location and scale invariant.

The determinant of Σ in the bivariate case is

$$|\Sigma| = \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2(1 - \rho^2), \quad -1 < \rho < 1, \quad \sigma_1 > 0, \quad \sigma_2 > 0.$$

The inverse is as follows, taking the inverse as the transpose of the matrix of cofactors divided by the determinant:

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}. \quad (3.2.4)$$

Then,

$$(X - \mu)' \Sigma^{-1} (X - \mu) = \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \equiv Q. \quad (3.2.5)$$

Hence, the real bivariate normal density is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1 - \rho^2)}} e^{-\frac{Q}{2(1 - \rho^2)}} \quad (3.2.6)$$

where Q is given in (3.2.5). Observe that Q is a positive definite quadratic form and hence $Q > 0$ for all X and μ . We can also obtain an interesting result on the standardized variables of x_1 and x_2 . Let the standardized x_j be $y_j = \frac{x_j - \mu_j}{\sigma_j}$, $j = 1, 2$ and $u = y_1 - y_2$. Then

$$\text{Var}(u) = \text{Var}(y_1) + \text{Var}(y_2) - 2\text{Cov}(y_1, y_2) = 1 + 1 - 2\rho = 2(1 - \rho). \quad (3.2.7)$$

This shows that the smaller the absolute value of ρ is, the larger the variance of u , and vice versa, noting that $-1 < \rho < 1$ in the bivariate real normal case but in general, $-1 \leq \rho \leq 1$. Observe that if $\rho = 0$ in the bivariate normal density given in (3.2.6), this joint density factorizes into the product of the marginal densities of x_1 and x_2 , which implies that x_1 and x_2 are independently distributed when $\rho = 0$. In general, for real scalar random variables x and y , $\rho = 0$ need not imply independence; however, in the bivariate

normal case, $\rho = 0$ if and only if x_1 and x_2 are independently distributed. As well, the exponent in (3.2.6) has the following feature:

$$Q = (X - \mu)' \Sigma^{-1} (X - \mu) = \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) = c \quad (3.2.8)$$

where c is positive describes an ellipse in two-dimensional Euclidean space, and for a general p ,

$$(X - \mu)' \Sigma^{-1} (X - \mu) = c > 0, \quad \Sigma > O, \quad (3.2.9)$$

describes the surface of an ellipsoid in the p -dimensional Euclidean space, observing that $\Sigma^{-1} > O$ when $\Sigma > O$.

Example 3.2.1. Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 1 \\ -1 & 1 & 2 \end{bmatrix}.$$

Show that $\Sigma > O$ and that Σ can be a covariance matrix for X . Taking $E[X] = \mu$ and $\text{Cov}(X) = \Sigma$, construct the exponent of a trivariate real Gaussian density explicitly and write down the density.

Solution 3.2.1. Let us verify the definiteness of Σ . Note that $\Sigma = \Sigma'$ (symmetric). The leading minors are $|3| = 3 > 0$, $\begin{vmatrix} 3 & 0 \\ 0 & 3 \end{vmatrix} = 9 > 0$, $|\Sigma| = 12 > 0$, and hence $\Sigma > O$. The matrix of cofactors of Σ , that is, $\text{Cof}(\Sigma)$ and the inverse of Σ are the following:

$$\text{Cof}(\Sigma) = \begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix}, \quad \Sigma^{-1} = \frac{1}{12} \begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix}. \quad (i)$$

Thus the exponent of the trivariate real Gaussian density is $-\frac{1}{2}Q$ where

$$\begin{aligned} Q &= \frac{1}{12} [x_1 - 1, x_2 + 1, x_3 + 2] \begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 + 1 \\ x_3 + 2 \end{bmatrix} \\ &= \frac{1}{12} \{5(x_1 - 1)^2 + 5(x_2 + 1)^2 + 9(x_3 + 2)^2 - 2(x_1 - 1)(x_2 + 1) \\ &\quad + 6(x_1 - 1)(x_3 + 2) - 6(x_2 + 1)(x_3 + 2)\}. \end{aligned} \quad (ii)$$

The normalizing constant of the density being

$$(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}} = (2\pi)^{\frac{3}{2}} [12]^{\frac{1}{2}} = 2^{\frac{5}{2}} \sqrt{3} \pi^{\frac{3}{2}},$$

the resulting trivariate Gaussian density is

$$f(X) = [2^{\frac{5}{2}} \sqrt{3} \pi^{\frac{3}{2}}]^{-1} e^{-\frac{1}{2}Q}$$

for $-\infty < x_j < \infty$, $j = 1, 2, 3$, where Q is specified in (ii).

3.2.1. The moment generating function in the real case

We have defined the multivariate Gaussian distribution via the following characterization whose proof relies on its moment generating function: if all the possible linear combinations of the components of a random vector are real univariate normal, then this vector must follow a real multivariate Gaussian distribution. We are now looking into the derivation of the mgf given the density. For a parameter vector T , with $T' = (t_1, \dots, t_p)$, we have

$$\begin{aligned} M_X(T) &= E[e^{T'X}] = \int_X e^{T'X} f(X) dX = e^{T'\mu} E[e^{T'(X-\mu)}] \\ &= \frac{e^{T'\mu}}{|\Sigma|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \int_X e^{T'(X-\mu) - \frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} dX. \end{aligned}$$

Observe that the moment generating function (mgf) in the real multivariate case is the expected value of e raised to a linear function of the real scalar variables. Making the transformation $Y = \Sigma^{-\frac{1}{2}}(X - \mu) \Rightarrow dY = |\Sigma|^{-\frac{1}{2}} dX$. The exponent can be simplified as follows:

$$\begin{aligned} T'(X - \mu) - \frac{1}{2}(X - \mu)'\Sigma^{-1}(X - \mu) &= -\frac{1}{2}\{-2T'\Sigma^{\frac{1}{2}}Y + Y'Y\} \\ &= -\frac{1}{2}\{(Y - \Sigma^{\frac{1}{2}}T)'\Sigma^{-\frac{1}{2}}(Y - \Sigma^{\frac{1}{2}}T) - T'\Sigma T\}. \end{aligned}$$

Hence

$$M_X(T) = e^{T'\mu + \frac{1}{2}T'\Sigma T} \frac{1}{(2\pi)^{\frac{p}{2}}} \int_Y e^{-\frac{1}{2}(Y - \Sigma^{\frac{1}{2}}T)'\Sigma^{-\frac{1}{2}}(Y - \Sigma^{\frac{1}{2}}T)} dY.$$

The integral over Y is 1 since this is the total integral of a multivariate normal density whose mean value vector is $\Sigma^{\frac{1}{2}}T$ and covariance matrix is the identity matrix. Thus the mgf of a multivariate real Gaussian vector is

$$M_X(T) = e^{T'\mu + \frac{1}{2}T'\Sigma T}. \tag{3.2.10}$$

In the singular normal case, we can still take (3.2.10) as the mgf for $\Sigma \geq O$ (non-negative definite), which encompasses the singular and nonsingular cases. Then, one can study properties of the normal distribution whether singular or nonsingular via (3.2.10).

We will now apply the differential operator $\frac{\partial}{\partial T}$ defined in Sect. 1.7 on the moment generating function of a $p \times 1$ real normal random vector X and evaluate the result at $T = O$ to obtain the mean value vector of this distribution, that is, $\mu = E[X]$. As well, $E[XX']$ is available by applying the operator $\frac{\partial}{\partial T} \frac{\partial}{\partial T'}$ on the mgf, and so on. From the mgf in (3.2.10), we have

$$\begin{aligned} \frac{\partial}{\partial T} M_X(T)|_{T=O} &= \frac{\partial}{\partial T} e^{T'\mu + \frac{1}{2}T'\Sigma T}|_{T=O} \\ &= [e^{T'\mu + \frac{1}{2}T'\Sigma T} [\mu + \Sigma T]]|_{T=O} \Rightarrow \mu = E[X]. \end{aligned} \quad (i)$$

Then,

$$\frac{\partial}{\partial T'} M_X(T) = e^{T'\mu + \frac{1}{2}T'\Sigma T} [\mu' + T'\Sigma]. \quad (ii)$$

Remember to write the scalar quantity, $M_X(T)$, on the left for scalar multiplication of matrices. Now,

$$\begin{aligned} \frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T) &= \frac{\partial}{\partial T} e^{T'\mu + \frac{1}{2}T'\Sigma T} [\mu' + T'\Sigma] \\ &= M_X(T) [\mu + \Sigma T] [\mu' + T'\Sigma] + M_X(T) [\Sigma]. \end{aligned}$$

Hence,

$$E[XX'] = \left[\frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T)|_{T=O} \right] = \Sigma + \mu\mu'. \quad (iii)$$

But

$$\text{Cov}(X) = E[XX'] - E[X]E[X'] = (\Sigma + \mu\mu') - \mu\mu' = \Sigma. \quad (iv)$$

In the multivariate real Gaussian case, we have only two parameters μ and Σ and both of these are available from the above equations. In the general case, we can evaluate higher moments as follows:

$$E[\cdots X'XX'] = \cdots \frac{\partial}{\partial T'} \frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T)|_{T=O}. \quad (v)$$

If the characteristic function $\phi_X(T)$, which is available from the mgf by replacing T by iT , $i = \sqrt{-1}$, is utilized, then multiply the left-hand side of (v) by $i = \sqrt{-1}$ with each operator operating on $\phi_X(T)$ because $\phi_X(T) = M_X(iT)$. The corresponding differential operators can also be developed for the complex case.

Given a real p -vector $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$, what will be the distribution of a linear function of X ? Let $u = L'X$, $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $L' = (a_1, \dots, a_p)$ where a_1, \dots, a_p are real scalar constants. Let us examine its mgf whose argument is a real scalar parameter t . The mgf of u is available by integrating out over the density of X . We have

$$M_u(t) = E[e^{t'u}] = E[e^{tL'X}] = E[e^{(tL')X}].$$

This is of the same form as in (3.2.10) and hence, $M_u(t)$ is available from (3.2.10) by replacing T' by (tL') , that is,

$$M_u(t) = e^{t(L'\mu) + \frac{t^2}{2}L'\Sigma L} \Rightarrow u \sim N_1(L'\mu, L'\Sigma L). \quad (3.2.11)$$

This means that u is a univariate normal with mean value $L'\mu = E[u]$ and the variance of $L'\Sigma L = \text{Var}(u)$. Now, let us consider a set of linearly independent linear functions of X . Let A be a real $q \times p$, $q \leq p$ matrix of full rank q and let the linear functions $U = AX$ where U is $q \times 1$. Then $E[U] = AE[X] = A\mu$ and the covariance matrix in U is

$$\begin{aligned} \text{Cov}(U) &= E[U - E(U)][U - E(U)]' = E[A(X - \mu)(X - \mu)'A'] \\ &= AE[(X - \mu)(X - \mu)']A' = A \Sigma A'. \end{aligned}$$

Observe that since $\Sigma > O$, we can write $\Sigma = \Sigma_1 \Sigma_1'$ so that $A \Sigma A' = (A \Sigma_1)(A \Sigma_1)'$ and $A \Sigma_1$ is of full rank which means that $A \Sigma A' > O$. Therefore, letting T be a $q \times 1$ parameter vector, we have

$$M_U(T) = E[e^{T'U}] = E[e^{T'AX}] = E[e^{(T'A)X}],$$

which is available from (3.2.10). That is,

$$M_U(T) = e^{T'A\mu + \frac{1}{2}(T'A\Sigma A'T)} \Rightarrow U \sim N_q(A\mu, A\Sigma A').$$

Thus U is a q -variate multivariate normal with parameters $A\mu$ and $A\Sigma A'$ and we have the following result:

Theorem 3.2.1. *Let the vector random variable X have a real p -variate nonsingular $N_p(\mu, \Sigma)$ distribution and the $q \times p$ matrix A with $q \leq p$, be a full rank constant matrix. Then*

$$U = AX \sim N_q(A\mu, A\Sigma A'), \quad A\Sigma A' > O. \quad (3.2.12)$$

Corollary 3.2.1. Let the vector random variable X have a real p -variate nonsingular $N_p(\mu, \Sigma)$ distribution and B be a $1 \times p$ constant vector. Then $U_1 = BX$ has a univariate normal distribution with parameters $B\mu$ and $B\Sigma B'$.

Example 3.2.2. Let X , $\mu = E[X]$, $\Sigma = \text{Cov}(X)$, Y , and A be as follows:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \mu = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Let $y_1 = x_1 + x_2 + x_3$ and $y_2 = x_1 - x_2 + x_3$ and write $Y = AX$. If $\Sigma > O$ and if $X \sim N_3(\mu, \Sigma)$, derive the density of (1) Y ; (2) y_1 directly as well as from (1).

Solution 3.2.2. The leading minors of Σ are $|(4)| = 4 > 0$, $\begin{vmatrix} 4 & -2 \\ -2 & 3 \end{vmatrix} = 8 > 0$, $|\Sigma| = 12 > 0$ and $\Sigma = \Sigma'$. Being symmetric and positive definite, Σ is a *bona fide* covariance matrix. Now, $Y = AX$ where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix};$$

$$E[Y] = AE[X] = A \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad (i)$$

$$\text{Cov}(Y) = A \text{Cov}(X) A' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 3 & 11 \end{bmatrix}. \quad (ii)$$

Since A is of full rank (rank 2) and y_1 and y_2 are linear functions of the real Gaussian vector X , Y has a bivariate nonsingular real Gaussian distribution with parameters $E(Y)$ and $\text{Cov}(Y)$. Since

$$\begin{bmatrix} 7 & 3 \\ 3 & 11 \end{bmatrix}^{-1} = \frac{1}{68} \begin{bmatrix} 11 & -3 \\ -3 & 7 \end{bmatrix},$$

the density of Y has the exponent $-\frac{1}{2}Q$ where

$$\begin{aligned} Q &= \frac{1}{68} \left\{ [y_1 - 1, y_2 - 1] \begin{bmatrix} 11 & -3 \\ -3 & 7 \end{bmatrix} \begin{bmatrix} y_1 - 1 \\ y_2 - 1 \end{bmatrix} \right\} \\ &= \frac{1}{68} \{ 11(y_1 - 1)^2 + 7(y_2 - 1)^2 - 6(y_1 - 1)(y_2 - 1) \}. \quad (iii) \end{aligned}$$

The normalizing constant being $(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}} = 2\pi\sqrt{68} = 4\sqrt{17}\pi$, the density of Y , denoted by $f(Y)$, is given by

$$f(Y) = \frac{1}{4\sqrt{17}\pi} e^{-\frac{1}{2}Q} \quad (iv)$$

where Q is specified in (iii). This establishes (1). For establishing (2), we first start with the

formula. Let $y_1 = A_1 X \Rightarrow A_1 = [1, 1, 1]$, $E[y_1] = A_1 E[X] = [1, 1, 1] \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = 1$

and

$$\text{Var}(y_1) = A_1 \text{Cov}(X) A_1' = [1, 1, 1] \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 7.$$

Hence $y_1 \sim N_1(1, 7)$. For establishing this result directly, observe that y_1 is a linear function of real normal variables and hence, it is univariate real normal with the parameters $E[y_1]$ and $\text{Var}(y_1)$. We may also obtain the marginal distribution of y_1 directly from the parameters of the joint density of y_1 and y_2 , which are given in (i) and (ii). Thus, (2) is also established.

The marginal distributions can also be determined from the mgf. Let us partition T , μ and Σ as follows:

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad (v)$$

where T_1 , $\mu_{(1)}$, X_1 are $r \times 1$ and Σ_{11} is $r \times r$. Letting $T_2 = O$ (the null vector), we have

$$\begin{aligned} T' \mu + \frac{1}{2} T' \Sigma T &= [T_1', O'] \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix} + \frac{1}{2} [T_1', O'] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} T_1 \\ O \end{bmatrix} \\ &= T_1' \mu_{(1)} + \frac{1}{2} T_1' \Sigma_{11} T_1, \end{aligned}$$

which is the structure of the mgf of a real Gaussian distribution with mean value vector $E[X_1] = \mu_{(1)}$ and covariance matrix $\text{Cov}(X_1) = \Sigma_{11}$. Therefore X_1 is an r -variate real Gaussian vector and similarly, X_2 is $(p - r)$ -variate real Gaussian vector. The standard notation used for a p -variate normal distribution is $X \sim N_p(\mu, \Sigma)$, $\Sigma \geq O$, which includes the nonsingular and singular cases. In the nonsingular case, $\Sigma > O$, whereas $|\Sigma| = 0$ in the singular case.

From the mgf in (3.2.10) and (i) above, if we have $\Sigma_{12} = O$ with $\Sigma_{21} = \Sigma'_{12}$, then the mgf of $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ becomes $e^{T'_1 \mu_{(1)} + T'_2 \mu_{(2)} + \frac{1}{2} T'_1 \Sigma_{11} T_1 + \frac{1}{2} T'_2 \Sigma_{22} T_2}$. That is,

$$M_X(T) = M_{X_1}(T_1)M_{X_2}(T_2),$$

which implies that X_1 and X_2 are independently distributed. Hence the following result:

Theorem 3.2.2. *Let the real $p \times 1$ vector $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$, and let X be partitioned into subvectors X_1 and X_2 , with the corresponding partitioning of μ and Σ , that is,*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then, X_1 and X_2 are independently distributed if and only if $\Sigma_{12} = \Sigma'_{21} = O$.

Observe that a covariance matrix being null need not imply independence of the subvectors; however, in the case of subvectors having a joint normal distribution, it suffices to have a null covariance matrix to conclude that the subvectors are independently distributed.

3.2a. The Moment Generating Function in the Complex Case

The determination of the mgf in the complex case is somewhat different. Take a p -variate complex Gaussian $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$. Let $\tilde{T}' = (\tilde{t}_1, \dots, \tilde{t}_p)$ be a parameter vector. Let $\tilde{T} = T_1 + iT_2$, where T_1 and T_2 are $p \times 1$ real vectors and $i = \sqrt{-1}$. Let $\tilde{X} = X_1 + iX_2$ with X_1 and X_2 being real. Then consider $\tilde{T}^* \tilde{X} = (T'_1 - iT'_2)(X_1 + iX_2) = T'_1 X_1 + T'_2 X_2 + i(T'_1 X_2 - T'_2 X_1)$. But $T'_1 X_1 + T'_2 X_2$ already contains the necessary number of parameters and all the corresponding real variables and hence to be consistent with the definition of the mgf in the real case one must take only the real part in $\tilde{T}^* \tilde{X}$. Hence the mgf in the complex case, denoted by $M_{\tilde{X}}(\tilde{T})$, is defined as $E[e^{\Re(\tilde{T}^* \tilde{X})}]$. For convenience, we may take $\tilde{X} = \tilde{X} - \tilde{\mu} + \tilde{\mu}$. Then $E[e^{\Re(\tilde{T}^* \tilde{X})}] = e^{\Re(\tilde{T}^* \tilde{\mu})} E[e^{\Re(\tilde{T}^* (\tilde{X} - \tilde{\mu}))}]$. On making the transformation $\tilde{Y} = \Sigma^{-\frac{1}{2}}(\tilde{X} - \tilde{\mu})$, $|\det(\Sigma)|$ appearing in the denominator of the density of \tilde{X} is canceled due to the Jacobian of the transformation and we have $(\tilde{X} - \tilde{\mu}) = \Sigma^{\frac{1}{2}} \tilde{Y}$. Thus,

$$E[e^{\Re(\tilde{T}^* \tilde{Y})}] = \frac{1}{\pi^p} \int_{\tilde{Y}} e^{\Re(\tilde{T}^* \Sigma^{\frac{1}{2}} \tilde{Y}) - \tilde{Y}^* \tilde{Y}} d\tilde{Y}. \quad (i)$$

For evaluating the integral in (i), we can utilize the following result which will be stated here as a lemma.

Lemma 3.2a.1. Let \tilde{U} and \tilde{V} be two $p \times 1$ vectors in the complex domain. Then

$$2 \Re(\tilde{U}^* \tilde{V}) = \tilde{U}^* \tilde{V} + \tilde{V}^* \tilde{U} = 2 \Re(\tilde{V}^* \tilde{U}).$$

Proof: Let $\tilde{U} = U_1 + iU_2$, $\tilde{V} = V_1 + iV_2$ where U_1, U_2, V_1, V_2 are real vectors and $i = \sqrt{-1}$. Then $\tilde{U}^* \tilde{V} = [U_1' - iU_2'] [V_1 + iV_2] = U_1' V_1 + U_2' V_2 + i[U_1' V_2 - U_2' V_1]$. Similarly $\tilde{V}^* \tilde{U} = V_1' U_1 + V_2' U_2 + i[V_1' U_2 - V_2' U_1]$. Observe that since U_1, U_2, V_1, V_2 are real, we have $U_i' V_j = V_j' U_i$ for all i and j . Hence, the sum $\tilde{U}^* \tilde{V} + \tilde{V}^* \tilde{U} = 2[U_1' V_1 + U_2' V_2] = 2 \Re(\tilde{V}^* \tilde{U})$. This completes the proof.

Now, the exponent in (i) can be written as

$$\Re(\tilde{T}^* \Sigma^{\frac{1}{2}} \tilde{Y}) = \frac{1}{2} \tilde{T}^* \Sigma^{\frac{1}{2}} \tilde{Y} + \frac{1}{2} \tilde{Y}^* \Sigma^{\frac{1}{2}} \tilde{T}$$

by using Lemma 3.2a.1, observing that $\Sigma = \Sigma^*$. Let us expand $(\tilde{Y} - C)^* (\tilde{Y} - C)$ as $\tilde{Y}^* \tilde{Y} - \tilde{Y}^* C - C^* \tilde{Y} + C^* C$ for some C . Comparing with the exponent in (i), we may take $C^* = \frac{1}{2} \tilde{T}^* \Sigma^{\frac{1}{2}}$ so that $C^* C = \frac{1}{4} \tilde{T}^* \Sigma \tilde{T}$. Therefore in the complex Gaussian case, the mgf is

$$M_{\tilde{X}}(\tilde{T}) = e^{\Re(\tilde{T}^* \tilde{\mu}) + \frac{1}{4} \tilde{T}^* \Sigma \tilde{T}}. \quad (3.2a.1)$$

Example 3.2a.1. Let \tilde{X} , $E[\tilde{X}] = \tilde{\mu}$, $\text{Cov}(\tilde{X}) = \Sigma$ be the following where $\tilde{X} \sim \tilde{N}_2(\tilde{\mu}, \Sigma)$, $\Sigma > O$,

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} 1 - i \\ 2 - 3i \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3 & 1 + i \\ 1 - i & 2 \end{bmatrix}.$$

Compute the mgf of \tilde{X} explicitly.

Solution 3.2a.1. Let $\tilde{T} = \begin{bmatrix} \tilde{t}_1 \\ \tilde{t}_2 \end{bmatrix}$ where let $\tilde{t}_1 = t_{11} + it_{12}$, $\tilde{t}_2 = t_{21} + it_{22}$ with $t_{11}, t_{12}, t_{21}, t_{22}$ being real scalar parameters. The mgf of \tilde{X} is

$$M_{\tilde{X}}(\tilde{T}) = e^{\Re(\tilde{T}^* \tilde{\mu}) + \frac{1}{4} \tilde{T}^* \Sigma \tilde{T}}.$$

Consider the first term in the exponent of the mgf:

$$\begin{aligned} \Re(\tilde{T}^* \tilde{\mu}) &= \Re \left\{ [t_{11} - it_{12}, t_{21} - it_{22}] \begin{bmatrix} 1 - i \\ 2 - 3i \end{bmatrix} \right\} \\ &= \Re \{ (t_{11} - it_{12})(1 - i) + (t_{21} - it_{22})(2 - 3i) \} \\ &= t_{11} - t_{12} + 2t_{21} - 3t_{22}. \end{aligned}$$

The second term in the exponent is the following:

$$\begin{aligned} \frac{1}{4} \tilde{T}^* \Sigma \tilde{T} &= \frac{1}{4} \{ [\tilde{t}_1^*, \tilde{t}_2^*] \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} \begin{bmatrix} \tilde{t}_1 \\ \tilde{t}_2 \end{bmatrix} \} \\ &= \frac{1}{4} \{ 3\tilde{t}_1^* \tilde{t}_1 + 2\tilde{t}_2^* \tilde{t}_2 + (1+i)\tilde{t}_1^* \tilde{t}_2 + (1-i)\tilde{t}_2^* \tilde{t}_1 \}. \end{aligned}$$

Note that since the parameters are scalar quantities, the conjugate transpose means only the conjugate or $\tilde{t}_j^* = \bar{\tilde{t}}_j$, $j = 1, 2$. Let us look at the non-diagonal terms. Note that $[(1+i)\tilde{t}_1^* \tilde{t}_2] + [(1-i)\tilde{t}_2^* \tilde{t}_1]$ gives $2(t_{11}t_{21} + t_{12}t_{22} + t_{12}t_{21} - t_{11}t_{22})$. However, $\tilde{t}_1^* \tilde{t}_1 = t_{11}^2 + t_{12}^2$, $\tilde{t}_2^* \tilde{t}_2 = t_{21}^2 + t_{22}^2$. Hence if the exponent of $M_{\tilde{X}}(\tilde{t})$ is denoted by ϕ ,

$$\begin{aligned} \phi &= [t_{11} - t_{12} + 2t_{21} - 3t_{22}] + \frac{1}{4} \{ 3(t_{11}^2 + t_{12}^2) + 2(t_{21}^2 + t_{22}^2) \\ &\quad + 2(t_{11}t_{21} + t_{12}t_{22} + t_{12}t_{21} - t_{11}t_{22}) \}. \end{aligned} \quad (i)$$

Thus the mgf is

$$M_{\tilde{X}}(\tilde{T}) = e^\phi$$

where ϕ is given in (i).

3.2a.1. Moments from the moment generating function

We can also derive the moments from the mgf of (3.2a.1) by operating with the differential operator of Sect. 1.7 of Chap. 1. For the complex case, the operator $\frac{\partial}{\partial X_1}$ in the real case has to be modified. Let $\tilde{X} = X_1 + iX_2$ be a $p \times 1$ vector in the complex domain where X_1 and X_2 are real and $p \times 1$ and $i = \sqrt{-1}$. Then in the complex domain the differential operator is

$$\frac{\partial}{\partial \tilde{X}} = \frac{\partial}{\partial X_1} + i \frac{\partial}{\partial X_2}. \quad (ii)$$

Let $\tilde{T} = T_1 + iT_2$, $\tilde{\mu} = \mu_{(1)} + i\mu_{(2)}$, $\Sigma = \Sigma_1 + i\Sigma_2$ where $T_1, T_2, \mu_{(1)}, \mu_{(2)}, \Sigma_1, \Sigma_2$ are all real and $i = \sqrt{-1}$, $\Sigma_1 = \Sigma_1'$, and $\Sigma_2' = -\Sigma_2$ because Σ is Hermitian. Note that $\tilde{T}^* \Sigma \tilde{T} = (T_1' - iT_2') \Sigma (T_1 + iT_2) = T_1' \Sigma T_1 + T_2' \Sigma T_2 + i(T_1' \Sigma T_2 - T_2' \Sigma T_1)$, and observe that

$$T_j' \Sigma T_j = T_j' (\Sigma_1 + i\Sigma_2) T_j = T_j' \Sigma_1 T_j + 0 \quad (iii)$$

for $j = 1, 2$ since Σ_2 is skew symmetric. The exponent in the mgf in (3.2a.1) can be simplified as follows: Letting u denote the exponent in the mgf and observing that $[\tilde{T}^* \Sigma \tilde{T}]^* = \tilde{T}^* \Sigma \tilde{T}$ is real,

$$\begin{aligned}
u &= \Re(\tilde{T}^* \tilde{\mu}) + \frac{1}{4} \tilde{T}^* \Sigma \tilde{T} = \Re(T'_1 - iT'_2)(\mu_{(1)} + i\mu_{(2)}) + \frac{1}{4}(T'_1 - iT'_2) \Sigma (T_1 + iT_2) \\
&= T'_1 \mu_{(1)} + T'_2 \mu_{(2)} + \frac{1}{4}[T'_1 \Sigma T_1 + T'_2 \Sigma T_2] + \frac{1}{4}u_1, \quad u_1 = i(T'_1 \Sigma T_2 - T'_2 \Sigma T_1) \\
&= T'_1 \mu_{(1)} + T'_2 \mu_{(2)} + \frac{1}{4}[T'_1 \Sigma_1 T_1 + T'_2 \Sigma_1 T_2] + \frac{1}{4}u_1. \tag{iv}
\end{aligned}$$

In this last line, we have made use of the result in (iii). The following lemma will enable us to simplify u_1 .

Lemma 3.2a.2. *Let T_1 and T_2 be real $p \times 1$ vectors. Let the $p \times p$ matrix Σ be Hermitian, $\Sigma = \Sigma^* = \Sigma_1 + i\Sigma_2$, with $\Sigma_1 = \Sigma'_1$ and $\Sigma_2 = -\Sigma'_2$. Then*

$$\begin{aligned}
u_1 &= i(T'_1 \Sigma T_2 - T'_2 \Sigma T_1) = -2T'_1 \Sigma_2 T_2 = 2T'_2 \Sigma_2 T_1 \\
&\Rightarrow \frac{\partial}{\partial T_1} u_1 = -2\Sigma_2 T_2 \quad \text{and} \quad \frac{\partial}{\partial T_2} u_1 = 2\Sigma_2 T_1. \tag{v}
\end{aligned}$$

Proof: This result will be established by making use of the following general properties: For a 1×1 matrix, the transpose is itself whereas the conjugate transpose is the conjugate of the same quantity. That is, $(a + ib)' = a + ib$, $(a + ib)^* = a - ib$ and if the conjugate transpose is equal to itself then the quantity is real or equivalently, if $(a + ib) = (a + ib)^* = a - ib$ then $b = 0$ and the quantity is real. Thus,

$$\begin{aligned}
u_1 &= i(T'_1 \Sigma T_2 - T'_2 \Sigma T_1) = i[T'_1(\Sigma_1 + i\Sigma_2)T_2 - T'_2(\Sigma_1 + i\Sigma_2)T_1], \\
&= iT'_1 \Sigma_1 T_2 - T'_1 \Sigma_2 T_2 - iT'_2 \Sigma_1 T_1 + T'_2 \Sigma_2 T_1 = -T'_1 \Sigma_2 T_2 + T'_2 \Sigma_2 T_1 \\
&= -2T'_1 \Sigma_2 T_2 = 2T'_2 \Sigma_2 T_1. \tag{vi}
\end{aligned}$$

The following properties were utilized: $T'_i \Sigma_1 T_j = T'_j \Sigma_1 T_i$ for all i and j since Σ_1 is a symmetric matrix, the quantity is 1×1 and real and hence, the transpose is itself; $T'_i \Sigma_2 T_j = -T'_j \Sigma_2 T_i$ for all i and j because the quantities are 1×1 and then, the transpose is itself, but the transpose of $\Sigma'_2 = -\Sigma_2$. This completes the proof.

Now, let us apply the operator $(\frac{\partial}{\partial T_1} + i\frac{\partial}{\partial T_2})$ to the mgf in (3.2a.1) and determine the various quantities. Note that in light of results stated in Chap. 1, we have

$$\begin{aligned}
\frac{\partial}{\partial T_1} (T'_1 \Sigma_1 T_1) &= 2\Sigma_1 T_1, \quad \frac{\partial}{\partial T_1} (-2T'_1 \Sigma_2 T_2) = -2\Sigma_2 T_2, \quad \frac{\partial}{\partial T_1} \Re(\tilde{T}^* \tilde{\mu}) = \mu_{(1)}, \\
\frac{\partial}{\partial T_2} (T'_2 \Sigma_1 T_2) &= 2\Sigma_1 T_2, \quad \frac{\partial}{\partial T_2} (2T'_2 \Sigma_2 T_1) = 2\Sigma_2 T_1, \quad \frac{\partial}{\partial T_2} \Re(\tilde{T}^* \tilde{\mu}) = \mu_{(2)}.
\end{aligned}$$

Thus, given (ii)–(vi), the operator applied to the exponent of the mgf gives the following result:

$$\begin{aligned} \left(\frac{\partial}{\partial T_1} + i \frac{\partial}{\partial T_2} \right) u &= \mu_{(1)} + i \mu_{(2)} + \frac{1}{4} [2\Sigma_1 T_1 - 2\Sigma_2 T_2 + 2\Sigma_1 i T_2 + 2\Sigma_2 i T_1] \\ &= \tilde{\mu} + \frac{1}{4} [2(\Sigma_1 + i \Sigma_2) T_1 + 2(\Sigma_1 + i \Sigma_2) i T_2] = \tilde{\mu} + \frac{1}{4} [2\Sigma \tilde{T}] = \tilde{\mu} + \frac{1}{2} \Sigma \tilde{T}, \end{aligned}$$

so that

$$\begin{aligned} \frac{\partial}{\partial \tilde{T}} M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} &= \left(\frac{\partial}{\partial T_1} + i \frac{\partial}{\partial T_2} \right) M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} \\ &= [M_{\tilde{X}}(\tilde{T})[\tilde{\mu} + \frac{1}{2} \tilde{\Sigma} \tilde{T}]|_{T_1=O, T_2=O} = \tilde{\mu}, \end{aligned} \quad (vii)$$

noting that $\tilde{T} = O$ implies that $T_1 = O$ and $T_2 = O$. For convenience, let us denote the operator by

$$\frac{\partial}{\partial \tilde{T}} = \left(\frac{\partial}{\partial T_1} + i \frac{\partial}{\partial T_2} \right).$$

From (vii), we have

$$\begin{aligned} \frac{\partial}{\partial \tilde{T}} M_{\tilde{X}}(\tilde{T}) &= M_{\tilde{X}}(\tilde{T})[\tilde{\mu} + \frac{1}{2} \Sigma \tilde{T}], \\ \frac{\partial}{\partial \tilde{T}^*} M_{\tilde{X}}(\tilde{T}) &= [\tilde{\mu}^* + \frac{1}{2} \tilde{T}^* \tilde{\Sigma}] M_{\tilde{X}}(\tilde{T}). \end{aligned}$$

Now, observe that

$$\begin{aligned} \tilde{T}^* \Sigma &= (T'_1 - iT'_2) \Sigma = T'_1 \Sigma - iT'_2 \Sigma \Rightarrow \\ \frac{\partial}{\partial T_1} (\tilde{T}^* \Sigma) &= \Sigma, \quad \frac{\partial}{\partial T_2} (\tilde{T}^* \Sigma) = -i \Sigma, \\ \left(\frac{\partial}{\partial T_1} + i \frac{\partial}{\partial T_2} \right) (\tilde{T}^* \Sigma) &= \Sigma - i(i) \Sigma = 2\Sigma, \end{aligned}$$

and

$$\frac{\partial}{\partial \tilde{T}} \frac{\partial}{\partial \tilde{T}^*} M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \tilde{\mu} \tilde{\mu}^* + \tilde{\Sigma}.$$

Thus,

$$\frac{\partial}{\partial \tilde{T}} M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \tilde{\mu} \quad \text{and} \quad \frac{\partial}{\partial \tilde{T}} \frac{\partial}{\partial \tilde{T}^*} M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \tilde{\Sigma} + \tilde{\mu} \tilde{\mu}^*,$$

and then $\text{Cov}(\tilde{X}) = \tilde{\Sigma}$. In general, for higher order moments, one would have

$$E[\dots \tilde{X}^* \tilde{X} \tilde{X}^*] = \dots \frac{\partial}{\partial \tilde{T}^*} \frac{\partial}{\partial \tilde{T}} \frac{\partial}{\partial \tilde{T}^*} M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=0}.$$

3.2a.2. Linear functions

Let $\tilde{w} = L^* \tilde{X}$ where $L^* = (a_1, \dots, a_p)$ and a_1, \dots, a_p are scalar constants, real or complex. Then the mgf of \tilde{w} can be evaluated by integrating out over the p -variate complex Gaussian density of \tilde{X} . That is,

$$M_{\tilde{w}}(\tilde{t}) = E[e^{\Re(\tilde{t}\tilde{w})}] = E[e^{\Re(\tilde{t}L^*\tilde{X})}]. \quad (3.2a.2)$$

Note that this expected value is available from (3.2a.1) by replacing \tilde{T}^* by $\tilde{t}L^*$. Hence

$$M_{\tilde{w}}(\tilde{t}) = e^{\Re(\tilde{t}(L^*\tilde{\mu})) + \frac{1}{4}\tilde{t}\tilde{t}^*(L^*\Sigma L)}. \quad (3.2a.3)$$

Then from (2.1a.1), $\tilde{w} = L^* \tilde{X}$ is univariate complex Gaussian with the parameters $L^* \tilde{\mu}$ and $L^* \Sigma L$. We now consider several such linear functions: Let $\tilde{Y} = A \tilde{X}$ where A is $q \times p$, $q \leq p$ and of full rank q . The distribution of \tilde{Y} can be determined as follows. Since \tilde{Y} is a function of \tilde{X} , we can evaluate the mgf of \tilde{Y} by integrating out over the density of \tilde{X} . Since \tilde{Y} is $q \times 1$, let us take a $q \times 1$ parameter vector \tilde{U} . Then,

$$M_{\tilde{Y}}(\tilde{U}) = E[e^{\Re(\tilde{U}^*\tilde{Y})}] = E[e^{\Re(\tilde{U}^*A\tilde{X})}] = E[e^{\Re[(\tilde{U}^*A)\tilde{X}]}]. \quad (3.2a.4)$$

On comparing this expected value with (3.2a.1), we can write down the mgf of \tilde{Y} as the following:

$$M_{\tilde{Y}}(\tilde{U}) = e^{\Re(\tilde{U}^*A\tilde{\mu}) + \frac{1}{4}(\tilde{U}^*A)\Sigma(A^*\tilde{U})} = e^{\Re(\tilde{U}^*(A\tilde{\mu})) + \frac{1}{4}\tilde{U}^*(A\Sigma A^*)\tilde{U}}, \quad (3.2a.5)$$

which means that \tilde{Y} has a q -variate complex Gaussian distribution with the parameters $A \tilde{\mu}$ and $A \Sigma A^*$. Thus, we have the following result:

Theorem 3.2a.1. *Let $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$ be a p -variate nonsingular complex normal vector. Let A be a $q \times p$, $q \leq p$, constant real or complex matrix of full rank q . Let $\tilde{Y} = A \tilde{X}$. Then,*

$$\tilde{Y} \sim \tilde{N}_q(A \tilde{\mu}, A \Sigma A^*), \quad A \Sigma A^* > O. \quad (3.2a.6)$$

Let us consider the following partitioning of \tilde{T} , \tilde{X} , Σ where \tilde{T} is $p \times 1$, \tilde{T}_1 is $r \times 1$, $r \leq p$, \tilde{X}_1 is $r \times 1$, Σ_{11} is $r \times r$, $\tilde{\mu}_{(1)}$ is $r \times 1$:

$$\tilde{T} = \begin{bmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{bmatrix}, \tilde{T}_1 = \begin{bmatrix} \tilde{t}_1 \\ \vdots \\ \tilde{t}_r \end{bmatrix}, \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{bmatrix}.$$

Let $\tilde{T}_2 = O$. Then the mgf of \tilde{X} becomes that of \tilde{X}_1 as

$$[\tilde{T}_1^*, O] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \tilde{T}_1 \\ O \end{bmatrix} = \tilde{T}_1^* \Sigma_{11} \tilde{T}_1.$$

Thus the mgf of \tilde{X}_1 becomes

$$M_{\tilde{X}_1}(\tilde{T}_1) = e^{\Re(\tilde{T}_1^* \tilde{\mu}_{(1)}) + \frac{1}{4} \tilde{T}_1^* \Sigma_{11} \tilde{T}_1}. \quad (3.2a.7)$$

This is the mgf of the $r \times 1$ subvector \tilde{X}_1 and hence \tilde{X}_1 has an r -variate complex Gaussian density with the mean value vector $\tilde{\mu}_{(1)}$ and the covariance matrix Σ_{11} . In a real or complex Gaussian vector, the individual variables can be permuted among themselves with the corresponding permutations in the mean value vector and the covariance matrix. Hence, all subsets of components of \tilde{X} are Gaussian distributed. Thus, any set of r components of \tilde{X} is again a complex Gaussian for $r = 1, 2, \dots, p$ when \tilde{X} is a p -variate complex Gaussian.

Suppose that, in the mgf of (3.2a.1), $\Sigma_{12} = O$ where $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$ and

$$\tilde{X} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix}, \tilde{\mu} = \begin{pmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tilde{T} = \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix}.$$

When Σ_{12} is null, so is Σ_{21} since $\Sigma_{21} = \Sigma_{12}^*$. Then $\Sigma = \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix}$ is block-diagonal.

As well, $\Re(\tilde{T}^* \tilde{\mu}) = \Re(\tilde{T}_1^* \tilde{\mu}_{(1)}) + \Re(\tilde{T}_2^* \tilde{\mu}_{(2)})$ and

$$\tilde{T}^* \Sigma \tilde{T} = (\tilde{T}_1^*, \tilde{T}_2^*) \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix} = \tilde{T}_1^* \Sigma_{11} \tilde{T}_1 + \tilde{T}_2^* \Sigma_{22} \tilde{T}_2. \quad (i)$$

In other words, $M_{\tilde{X}}(\tilde{T})$ becomes the product of the the mgf of \tilde{X}_1 and the mgf of \tilde{X}_2 , that is, \tilde{X}_1 and \tilde{X}_2 are independently distributed whenever $\Sigma_{12} = O$.

Theorem 3.2a.2. *Let $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$, be a nonsingular complex Gaussian vector. Consider the partitioning of \tilde{X} , $\tilde{\mu}$, \tilde{T} , Σ as in (i) above. Then, the subvectors \tilde{X}_1 and \tilde{X}_2 are independently distributed as complex Gaussian vectors if and only if $\Sigma_{12} = O$ or equivalently, $\Sigma_{21} = O$.*

Exercises 3.2

3.2.1. Construct a 2×2 real positive definite matrix A . Then write down a bivariate real Gaussian density where the covariance matrix is this A .

3.2.2. Construct a 2×2 Hermitian positive definite matrix B and then construct a complex bivariate Gaussian density. Write the exponent and normalizing constant explicitly.

3.2.3. Construct a 3×3 real positive definite matrix A . Then create a real trivariate Gaussian density with this A being the covariance matrix. Write down the exponent and the normalizing constant explicitly.

3.2.4. Repeat Exercise 3.2.3 for the complex Gaussian case.

3.2.5. Let the $p \times 1$ real vector random variable have a p -variate real nonsingular Gaussian density $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let L be a $p \times 1$ constant vector. Let $u = L'X = X'L =$ a linear function of X . Show that $E[u] = L'\mu$, $\text{Var}(u) = L'\Sigma L$ and that u is a univariate Gaussian with the parameters $L'\mu$ and $L'\Sigma L$.

3.2.6. Show that the mgf of u in Exercise 3.2.5 is

$$M_u(t) = e^{t(L'\mu) + \frac{t^2}{2}L'\Sigma L}.$$

3.2.7. What are the corresponding results in Exercises 3.2.5 and 3.2.6 for the nonsingular complex Gaussian case?

3.2.8. Let $X \sim N_p(O, \Sigma)$, $\Sigma > O$, be a real p -variate nonsingular Gaussian vector. Let $u_1 = X'\Sigma^{-1}X$, and $u_2 = X'X$. Derive the densities of u_1 and u_2 .

3.2.9. Establish Theorem 3.2.1 by using transformation of variables [Hint: Augment the matrix A with a matrix B such that $C = \begin{pmatrix} A \\ B \end{pmatrix}$ is $p \times p$ and nonsingular. Derive the density of $Y = CX$, and therefrom, the marginal density of AX .]

3.2.10. By constructing counter examples or otherwise, show the following: Let the real scalar random variables x_1 and x_2 be such that $x_1 \sim N_1(\mu_1, \sigma_1^2)$, $\sigma_1 > 0$, $x_2 \sim N_1(\mu_2, \sigma_2^2)$, $\sigma_2 > 0$ and $\text{Cov}(x_1, x_2) = 0$. Then, the joint density need not be bivariate normal.

3.2.11. Generalize Exercise 3.2.10 to p -vectors X_1 and X_2 .

3.2.12. Extend Exercises 3.2.10 and 3.2.11 to the complex domain.

3.3. Marginal and Conditional Densities, Real Case

Let the $p \times 1$ vector have a real p -variate Gaussian distribution $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let X , μ and Σ be partitioned as the following:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

where X_1 and $\mu_{(1)}$ are $r \times 1$, X_2 and $\mu_{(2)}$ are $(p - r) \times 1$, Σ^{11} is $r \times r$, and so on. Then

$$\begin{aligned} (X - \mu)' \Sigma^{-1} (X - \mu) &= [(X_1 - \mu_{(1)})', (X_2 - \mu_{(2)})'] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} X_1 - \mu_{(1)} \\ X_2 - \mu_{(2)} \end{bmatrix} \\ &= (X_1 - \mu_{(1)})' \Sigma^{11} (X_1 - \mu_{(1)}) + (X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) \\ &\quad + (X_1 - \mu_{(1)})' \Sigma^{12} (X_2 - \mu_{(2)}) + (X_2 - \mu_{(2)})' \Sigma^{21} (X_1 - \mu_{(1)}). \end{aligned} \quad (i)$$

But

$$[(X_1 - \mu_{(1)})' \Sigma^{12} (X_2 - \mu_{(2)})]' = (X_2 - \mu_{(2)})' \Sigma^{21} (X_1 - \mu_{(1)})$$

and both are real 1×1 . Thus they are equal and we may write their sum as twice either one of them. Collecting the terms containing $X_2 - \mu_{(2)}$, we have

$$(X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) + 2(X_2 - \mu_{(2)})' \Sigma^{21} (X_1 - \mu_{(1)}). \quad (ii)$$

If we expand a quadratic form of the type $(X_2 - \mu_{(2)} + C)' \Sigma^{22} (X_2 - \mu_{(2)} + C)$, we have

$$\begin{aligned} (X_2 - \mu_{(2)} + C)' \Sigma^{22} (X_2 - \mu_{(2)} + C) &= (X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) \\ &\quad + (X_2 - \mu_{(2)})' \Sigma^{22} C + C' \Sigma^{22} (X_2 - \mu_{(2)}) + C' \Sigma^{22} C. \end{aligned} \quad (iii)$$

Comparing (ii) and (iii), let

$$\Sigma^{22} C = \Sigma^{21} (X_1 - \mu_{(1)}) \Rightarrow C = (\Sigma^{22})^{-1} \Sigma^{21} (X_1 - \mu_{(1)}).$$

Then,

$$C' \Sigma^{22} C = (X_1 - \mu_{(1)})' \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} (X_1 - \mu_{(1)}).$$

Hence,

$$\begin{aligned} (X - \mu)' \Sigma^{-1} (X - \mu) &= (X_1 - \mu_{(1)})' [\Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}] (X_1 - \mu_{(1)}) \\ &\quad + (X_2 - \mu_{(2)} + C)' \Sigma^{22} (X_2 - \mu_{(2)} + C), \end{aligned}$$

and after integrating out X_2 , the balance of the exponent is $(X_1 - \mu_{(1)})' \Sigma_{11}^{-1} (X_1 - \mu_{(1)})$, where Σ_{11} is the $r \times r$ leading submatrix in Σ ; the reader may refer to Sect. 1.3 for results on the inversion of partitioned matrices. Observe that $\Sigma_{11}^{-1} = \Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1} \Sigma^{21}$. The integral over X_2 only gives a constant and hence the marginal density of X_1 is

$$f_1(X_1) = c_1 e^{-\frac{1}{2}(X_1 - \mu_{(1)})' \Sigma_{11}^{-1} (X_1 - \mu_{(1)})}.$$

On noting that it has the same structure as the real multivariate Gaussian density, its normalizing constant can easily be determined and the resulting density is as follows:

$$f_1(X_1) = \frac{1}{|\Sigma_{11}|^{\frac{1}{2}} (2\pi)^{\frac{r}{2}}} e^{-\frac{1}{2}(X_1 - \mu_{(1)})' \Sigma_{11}^{-1} (X_1 - \mu_{(1)})}, \quad \Sigma_{11} > O, \quad (3.3.1)$$

for $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $j = 1, \dots, r$, and where Σ_{11} is the covariance matrix in X_1 and $\mu_{(1)} = E[X_1]$ and $\Sigma_{11} = \text{Cov}(X_1)$. From symmetry, we obtain the following marginal density of X_2 in the real Gaussian case:

$$f_2(X_2) = \frac{1}{|\Sigma_{22}|^{\frac{1}{2}} (2\pi)^{\frac{p-r}{2}}} e^{-\frac{1}{2}(X_2 - \mu_{(2)})' \Sigma_{22}^{-1} (X_2 - \mu_{(2)})}, \quad \Sigma_{22} > O, \quad (3.3.2)$$

for $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $j = r + 1, \dots, p$.

Observe that we can permute the elements in X as we please with the corresponding permutations in μ and the covariance matrix Σ . Hence the real Gaussian density in the p -variate case is a multivariate density and not a vector/matrix-variate density. From this property, it follows that every subset of the elements from X has a real multivariate Gaussian distribution and the individual variables have univariate real normal or Gaussian distribution. Hence our derivation of the marginal density of X_1 is a general density for a subset of r elements in X because those r elements can be brought to the first r positions through permutations of the elements in X with the corresponding permutations in μ and Σ .

The bivariate case

Let us look at the explicit form of the real Gaussian density for $p = 2$. In the bivariate case,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \quad \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{vmatrix} = \sigma_{11}\sigma_{22} - (\sigma_{12})^2.$$

For convenience, let us denote σ_{11} by σ_1^2 and σ_{22} by σ_2^2 . Then $\sigma_{12} = \sigma_1\sigma_2\rho$ where ρ is the correlation between x_1 and x_2 , and for $p = 2$,

$$|\Sigma| = \sigma_1^2\sigma_2^2 - (\sigma_1\sigma_2\rho)^2 = \sigma_1^2\sigma_2^2(1 - \rho^2).$$

Thus, in that case,

$$\begin{aligned}\Sigma^{-1} &= \frac{1}{|\Sigma|} [\text{cof}(\Sigma)]' = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{bmatrix} \\ &= \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}, \quad -1 < \rho < 1.\end{aligned}$$

Hence, substituting these into the general expression for the real Gaussian density and denoting the real bivariate density as $f(x_1, x_2)$, we have the following:

$$f(x_1, x_2) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}Q\right\} \quad (3.3.3)$$

where Q is the real positive definite quadratic form

$$Q = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2$$

for $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$, $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $j = 1, 2$.

The conditional density of X_1 given X_2 , denoted by $g_1(X_1|X_2)$, is the following:

$$\begin{aligned}g_1(X_1|X_2) &= \frac{f(X)}{f_2(X_2)} = \frac{|\Sigma_{22}|^{\frac{1}{2}}}{(2\pi)^{\frac{r}{2}}|\Sigma|^{\frac{1}{2}}} \\ &\quad \times \exp\left\{-\frac{1}{2}[(X - \mu)' \Sigma^{-1}(X - \mu) - (X_2 - \mu_{(2)})' \Sigma_{22}^{-1}(X_2 - \mu_{(2)})]\right\}.\end{aligned}$$

We can simplify the exponent, excluding $-\frac{1}{2}$, as follows:

$$\begin{aligned}(X - \mu)' \Sigma^{-1}(X - \mu) - (X_2 - \mu_{(2)})' \Sigma_{22}^{-1}(X_2 - \mu_{(2)}) \\ &= (X_1 - \mu_{(1)})' \Sigma^{11}(X_1 - \mu_{(1)}) + 2(X_1 - \mu_{(1)})' \Sigma^{12}(X_2 - \mu_{(2)}) \\ &\quad + (X_2 - \mu_{(2)})' \Sigma^{22}(X_2 - \mu_{(2)}) - (X_2 - \mu_{(2)})' \Sigma_{22}^{-1}(X_2 - \mu_{(2)}).\end{aligned}$$

But $\Sigma_{22}^{-1} = \Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}$. Hence the terms containing Σ^{22} are canceled. The remaining terms containing $X_2 - \mu_{(2)}$ are

$$2(X_1 - \mu_{(1)})' \Sigma^{12}(X_2 - \mu_{(2)}) + (X_2 - \mu_{(2)})' \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}(X_2 - \mu_{(2)}).$$

Combining these two terms with $(X_1 - \mu_{(1)})' \Sigma^{11}(X_1 - \mu_{(1)})$ results in the quadratic form $(X_1 - \mu_{(1)} + C)' \Sigma^{11}(X_1 - \mu_{(1)} + C)$ where $C = (\Sigma^{11})^{-1}\Sigma^{12}(X_2 - \mu_{(2)})$. Now, noting that

$$\frac{|\Sigma_{22}|^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} = \left[\frac{|\Sigma_{22}|}{|\Sigma_{22}| |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|} \right]^{\frac{1}{2}} = \frac{1}{|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{\frac{1}{2}}},$$

the conditional density of X_1 given X_2 , which is denoted by $g_1(X_1|X_2)$, can be expressed as follows:

$$g_1(X_1|X_2) = \frac{1}{(2\pi)^{\frac{r}{2}} |\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(X_1 - \mu_{(1)} + C)'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(X_1 - \mu_{(1)} + C)\right\} \quad (3.3.4)$$

where $C = (\Sigma^{11})^{-1}\Sigma^{12}(X_2 - \mu_{(2)})$. Hence, the conditional expectation and covariance of X_1 given X_2 are

$$\begin{aligned} E[X_1|X_2] &= \mu_{(1)} - C = \mu_{(1)} - (\Sigma^{11})^{-1}\Sigma^{12}(X_2 - \mu_{(2)}) \\ &= \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_{(2)}), \text{ which is linear in } X_2. \\ \text{Cov}(X_1|X_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \text{ which is free of } X_2. \end{aligned} \quad (3.3.5)$$

From the inverses of partitioned matrices obtained in Sect. 1.3, we have $-(\Sigma^{11})^{-1}\Sigma^{12} = \Sigma_{12}\Sigma_{22}^{-1}$, which yields the representation of the conditional expectation appearing in Eq. (3.3.5). The matrix $\Sigma_{12}\Sigma_{22}^{-1}$ is often called *the matrix of regression coefficients*. From symmetry, it follows that the conditional density of X_2 , given X_1 , denoted by $g_2(X_2|X_1)$, is given by

$$g_2(X_2|X_1) = \frac{1}{(2\pi)^{\frac{p-r}{2}} |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(X_2 - \mu_{(2)} + C_1)'(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}(X_2 - \mu_{(2)} + C_1)\right\} \quad (3.3.6)$$

where $C_1 = (\Sigma^{22})^{-1}\Sigma^{21}(X_1 - \mu_{(1)})$, and the conditional expectation and conditional variance of X_2 given X_1 are

$$\begin{aligned} E[X_2|X_1] &= \mu_{(2)} - C_1 = \mu_{(2)} - (\Sigma^{22})^{-1}\Sigma^{21}(X_1 - \mu_{(1)}) \\ &= \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_{(1)}), \text{ which linear in } X_1 \\ \text{Cov}(X_2|X_1) &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \text{ which is free of } X_1, \end{aligned} \quad (3.3.7)$$

the matrix $\Sigma_{21}\Sigma_{11}^{-1}$ being often called *the matrix of regression coefficients*.

What is then the conditional expectation of x_1 given x_2 in the bivariate normal case? From formula (3.3.5) for $p = 2$, we have

$$\begin{aligned} E[X_1|X_2] &= \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_{(2)}) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2) \\ &= \mu_1 + \frac{\sigma_1\sigma_2\rho}{\sigma_2^2}(x_2 - \mu_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2) = E[x_1|x_2], \end{aligned} \quad (3.3.8)$$

which is linear in x_2 . The coefficient $\frac{\sigma_1}{\sigma_2}\rho$ is often referred to as *the regression coefficient*. Then, from (3.3.7) we have

$$E[x_2|x_1] = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x_1 - \mu_1), \text{ which is linear in } x_1 \quad (3.3.9)$$

and $\frac{\sigma_2}{\sigma_1}\rho$ is *the regression coefficient*. Thus, (3.3.8) gives the best predictor of x_1 based on x_2 and (3.3.9), the best predictor of x_2 based on x_1 , both being linear in the case of a multivariate real normal distribution; in this case, we have a bivariate normal distribution.

Example 3.3.1. Let $X, x_1, x_2, x_3, E[X] = \mu, \text{Cov}(X) = \Sigma$ be specified as follows where $X \sim N_3(\mu, \Sigma), \Sigma > O$:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Compute (1) the marginal densities of x_1 and $X_2 = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}$; (2) the conditional density of x_1 given X_2 and the conditional density of X_2 given x_1 ; (3) conditional expectations or regressions of x_1 on X_2 and X_2 on x_1 .

Solution 3.3.1. Let us partition Σ accordingly, that is,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Sigma_{11} = \sigma_{11} = (3), \Sigma_{12} = [-2, 0], \Sigma_{21} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \Sigma_{22} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}.$$

Let us compute the following quantities:

$$\begin{aligned}\Sigma_{22}^{-1} &= \frac{1}{5} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} &= 3 - [-2, 0] \left(\frac{1}{5}\right) \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \frac{3}{5} \\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} &= \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} - \left(\frac{1}{3}\right) \begin{bmatrix} -2 \\ 0 \end{bmatrix} [-2, 0] = \begin{bmatrix} \frac{2}{3} & 1 \\ 1 & 3 \end{bmatrix} \\ [\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]^{-1} &= \frac{5}{3} \\ [\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]^{-1} &= \begin{bmatrix} 3 & -1 \\ -1 & \frac{2}{3} \end{bmatrix}.\end{aligned}$$

As well,

$$\Sigma_{12}\Sigma_{22}^{-1} = [-2, 0] \left(\frac{1}{5}\right) \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} = \left[-\frac{6}{5}, \frac{2}{5}\right] \text{ and } \Sigma_{21}\Sigma_{11}^{-1} = \left(\frac{1}{3}\right) \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 0 \end{bmatrix}.$$

Then we have the following:

$$\begin{aligned}E[X_1|X_2] &= \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_{(2)}) \\ &= -1 + \left[-\frac{6}{5}, \frac{2}{5}\right] \begin{bmatrix} x_2 - 0 \\ x_3 + 2 \end{bmatrix} = -1 - \frac{6}{5}x_2 + \frac{2}{5}(x_3 + 2) \end{aligned} \quad (i)$$

and

$$\text{Cov}(X_1|X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \frac{3}{5}; \quad (ii)$$

$$\begin{aligned}E[X_2|X_1] &= \mu_{(2)} + \Sigma_{12}\Sigma_{22}^{-1}(X_1 - \mu_{(1)}) \\ &= \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \begin{bmatrix} -2/3 \\ 0 \end{bmatrix} (x_1 + 1) = \begin{bmatrix} -\frac{2}{3}(x_1 + 1) \\ -2 \end{bmatrix} \end{aligned} \quad (iii)$$

and

$$\text{Cov}(X_2|X_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \begin{bmatrix} 2/3 & 1 \\ 1 & 3 \end{bmatrix}. \quad (iv)$$

The distributions of x_1 and X_2 are respectively $x_1 \sim N_1(-1, 3)$ and $X_2 \sim N_2(\mu_{(2)}, \Sigma_{22})$, the corresponding densities denoted by $f_1(x_1)$ and $f_2(X_2)$ being

$$\begin{aligned}f_1(x_1) &= \frac{1}{\sqrt{(2\pi)}\sqrt{3}} e^{-\frac{1}{6}(x_1+1)^2}, \quad -\infty < x_1 < \infty, \\ f_2(X_2) &= \frac{1}{(2\pi)\sqrt{5}} e^{-\frac{1}{2}Q_1}, \quad Q_1 = \frac{1}{5}[3(x_2)^2 - 2(x_2)(x_3 + 2) + 2(x_3 + 2)^2]\end{aligned}$$

for $-\infty < x_j < \infty$, $j = 2, 3$. The conditional distributions are $X_1|X_2 \sim N_1(E(X_1|X_2), \text{Var}(X_1|X_2))$ and $X_2|X_1 \sim N_2(E(X_2|X_1), \text{Cov}(X_2|X_1))$, the associated densities denoted by $g_1(X_1|X_2)$ and $g_2(X_2|X_1)$ being given by

$$g_1(X_1|X_2) = \frac{1}{\sqrt{(2\pi)(3/5)^{\frac{1}{2}}}} e^{-\frac{5}{6}[x_1+1+\frac{6}{5}x_2-\frac{2}{5}(x_3+2)]^2},$$

$$g_2(X_2|X_1) = \frac{1}{2\pi \times 1} e^{-\frac{1}{2}Q_2},$$

$$Q_2 = 3\left[x_2 + \frac{2}{3}(x_1 + 1)\right]^2 - 2\left[x_2 + \frac{2}{3}(x_1 + 1)\right](x_3 + 2) + \frac{2}{3}(x_3 + 2)^2$$

for $-\infty < x_j < \infty$, $j = 1, 2, 3$. This completes the computations.

3.3a. Conditional and Marginal Densities in the Complex Case

Let the $p \times 1$ complex vector \tilde{X} have the p -variate complex normal distribution, $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} > O$. As can be seen from the corresponding mgf which was derived in Sect. 3.2a, all subsets of the variables $\tilde{x}_1, \dots, \tilde{x}_p$ are again complex Gaussian distributed. This result can be obtained by integrating out the remaining variables from the p -variate complex Gaussian density. Let $\tilde{U} = \tilde{X} - \tilde{\mu}$ for convenience. Partition \tilde{X} , $\tilde{\mu}$, \tilde{U} into subvectors and Σ into submatrices as follows:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}, \quad \tilde{U} = \begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \end{bmatrix}$$

where \tilde{X}_1 , $\tilde{\mu}_{(1)}$, \tilde{U}_1 are $r \times 1$ and Σ^{11} is $r \times r$. Consider

$$\begin{aligned} \tilde{U}^* \Sigma^{-1} \tilde{U} &= [\tilde{U}_1^*, \tilde{U}_2^*] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \end{bmatrix} \\ &= \tilde{U}_1^* \Sigma^{11} \tilde{U}_1 + \tilde{U}_2^* \Sigma^{22} \tilde{U}_2 + \tilde{U}_1^* \Sigma^{12} \tilde{U}_2 + \tilde{U}_2^* \Sigma^{21} \tilde{U}_1 \end{aligned} \quad (i)$$

and suppose that we wish to integrate out \tilde{U}_2 to obtain the marginal density of \tilde{U}_1 . The terms containing \tilde{U}_2 are $\tilde{U}_2^* \Sigma^{22} \tilde{U}_2 + \tilde{U}_1^* \Sigma^{12} \tilde{U}_2 + \tilde{U}_2^* \Sigma^{21} \tilde{U}_1$. On expanding the Hermitian form

$$\begin{aligned} (\tilde{U}_2 + C)^* \Sigma^{22} (\tilde{U}_2 + C) &= \tilde{U}_2^* \Sigma^{22} \tilde{U}_2 + \tilde{U}_2^* \Sigma^{22} C \\ &\quad + C^* \Sigma^{22} \tilde{U}_2 + C^* \Sigma^{22} C, \end{aligned} \quad (ii)$$

for some C and comparing (i) and (ii), we may let $\Sigma^{21} \tilde{U}_1 = \Sigma^{22} C \Rightarrow C = (\Sigma^{22})^{-1} \Sigma^{21} \tilde{U}_1$. Then $C^* \Sigma^{22} C = \tilde{U}_1^* \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} \tilde{U}_1$ and (i) may thus be written as

$$\tilde{U}_1^* (\Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}) \tilde{U}_1 + (\tilde{U}_2 + C)^* \Sigma^{22} (\tilde{U}_2 + C), \quad C = (\Sigma^{22})^{-1} \Sigma^{21} \tilde{U}_1.$$

However, from Sect. 1.3 on partitioned matrices, we have

$$\Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} = \Sigma_{11}^{-1}.$$

As well,

$$\begin{aligned} \det(\Sigma) &= [\det(\Sigma_{11})][\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})] \\ &= [\det(\Sigma_{11})][\det((\Sigma^{22})^{-1})]. \end{aligned}$$

Note that the integral of $\exp\{-(\tilde{U}_2 + C)^* \Sigma^{22}(\tilde{U}_2 + C)\}$ over \tilde{U}_2 gives $\pi^{p-r} |\det(\Sigma^{22})^{-1}| = \pi^{p-r} |\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})|$. Hence the marginal density of \tilde{X}_1 is

$$\tilde{f}_1(\tilde{X}_1) = \frac{1}{\pi^r |\det(\Sigma_{11})|} e^{-(\tilde{X}_1 - \tilde{\mu}_{(1)})^* \Sigma_{11}^{-1}(\tilde{X}_1 - \tilde{\mu}_{(1)})}, \quad \Sigma_{11} > \mathbf{O}. \quad (3.3a.1)$$

It is an r -variate complex Gaussian density. Similarly \tilde{X}_2 has the $(p-r)$ -variate complex Gaussian density

$$\tilde{f}_2(\tilde{X}_2) = \frac{1}{\pi^{p-r} |\det(\Sigma_{22})|} e^{-(\tilde{X}_2 - \tilde{\mu}_{(2)})^* \Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)})}, \quad \Sigma_{22} > \mathbf{O}. \quad (3.3a.2)$$

Hence, the conditional density of \tilde{X}_1 given \tilde{X}_2 , is

$$\begin{aligned} \tilde{g}_1(\tilde{X}_1 | \tilde{X}_2) &= \frac{\tilde{f}(\tilde{X}_1, \tilde{X}_2)}{\tilde{f}_2(\tilde{X}_2)} = \frac{\pi^{p-r} |\det(\Sigma_{22})|}{\pi^p |\det(\Sigma)|} \\ &\quad \times e^{-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu}) + (\tilde{X}_2 - \tilde{\mu}_{(2)})^* \Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)})}. \end{aligned}$$

From Sect. 1.3, we have

$$|\det(\Sigma)| = |\det(\Sigma_{22})| |\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})|$$

and then the normalizing constant is $[\pi^r |\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})|]^{-1}$. The exponential part reduces to the following by taking $\tilde{U} = \tilde{X} - \tilde{\mu}$, $\tilde{U}_1 = \tilde{X}_1 - \tilde{\mu}_{(1)}$, $\tilde{U}_2 = \tilde{X}_2 - \tilde{\mu}_{(2)}$:

$$\begin{aligned} &(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu}) - (\tilde{X}_2 - \tilde{\mu}_{(2)})^* \Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \\ &= \tilde{U}_1^* \Sigma^{11} \tilde{U}_1 + \tilde{U}_2^* \Sigma^{22} \tilde{U}_2 + \tilde{U}_1^* \Sigma^{12} \tilde{U}_2 \\ &\quad + \tilde{U}_2^* \Sigma^{21} \tilde{U}_1 - \tilde{U}_2^* (\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}) \tilde{U}_2 \\ &= \tilde{U}_1^* \Sigma^{11} \tilde{U}_1 + \tilde{U}_2^* \Sigma^{21} (\Sigma^{11})^{-1} \Sigma^{12} \tilde{U}_2 + 2 \tilde{U}_1^* \Sigma^{12} \tilde{U}_2 \\ &= [\tilde{U}_1 + (\Sigma^{11})^{-1} \Sigma^{12} \tilde{U}_2]^* \Sigma^{11} [\tilde{U}_1 + (\Sigma^{11})^{-1} \Sigma^{12} \tilde{U}_2]. \quad (3.3a.3) \end{aligned}$$

This exponent has the same structure as that of a complex Gaussian density with $E[\tilde{U}_1|\tilde{U}_2] = -(\Sigma^{11})^{-1}\Sigma^{12}\tilde{U}_2$ and $\text{Cov}(\tilde{X}_1|\tilde{X}_2) = (\Sigma^{11})^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Therefore the conditional density of \tilde{X}_1 given \tilde{X}_2 is given by

$$\tilde{g}_1(\tilde{X}_1|\tilde{X}_2) = \frac{1}{\pi^r |\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})|} e^{-(\tilde{X}_1 - \tilde{\mu}_{(1)} + C)^* \Sigma^{11} (\tilde{X}_1 - \tilde{\mu}_{(1)} + C)},$$

$$C = -(\Sigma^{11})^{-1}\Sigma^{12}(\tilde{X}_2 - \tilde{\mu}_{(2)}). \quad (3.3a.4)$$

The conditional expectation of \tilde{X}_1 given \tilde{X}_2 is then

$$\begin{aligned} E[\tilde{X}_1|\tilde{X}_2] &= \tilde{\mu}_{(1)} - (\Sigma^{11})^{-1}\Sigma^{12}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \\ &= \tilde{\mu}_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \quad (\text{linear in } \tilde{X}_2) \end{aligned} \quad (3.3a.5)$$

which follows from a result on partitioning of matrices obtained in Sect. 1.3. The matrix $\Sigma_{12}\Sigma_{22}^{-1}$ is referred to as *the matrix of regression coefficients*. The conditional covariance matrix is

$$\text{Cov}(\tilde{X}_1|\tilde{X}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (\text{free of } \tilde{X}_2).$$

From symmetry, the conditional density of \tilde{X}_2 given \tilde{X}_1 is given by

$$\begin{aligned} \tilde{g}_2(\tilde{X}_2|\tilde{X}_1) &= \frac{1}{\pi^{p-r} |\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})|} \\ &\times e^{-(\tilde{X}_2 - \tilde{\mu}_{(2)} + C_1)^* \Sigma^{22} (\tilde{X}_2 - \tilde{\mu}_{(2)} + C_1)}, \end{aligned} \quad (3.3a.6)$$

$$C_1 = -(\Sigma^{22})^{-1}\Sigma^{21}(\tilde{X}_1 - \tilde{\mu}_{(1)}), \quad \Sigma^{22} > O.$$

Then the conditional expectation and the conditional covariance of \tilde{X}_2 given \tilde{X}_1 are the following:

$$\begin{aligned} E[\tilde{X}_2|\tilde{X}_1] &= \tilde{\mu}_{(2)} - (\Sigma^{22})^{-1}\Sigma^{21}(\tilde{X}_1 - \tilde{\mu}_{(1)}) \\ &= \tilde{\mu}_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\tilde{X}_1 - \tilde{\mu}_{(1)}) \quad (\text{linear in } \tilde{X}_1) \end{aligned} \quad (3.3a.7)$$

$$\text{Cov}(\tilde{X}_2|\tilde{X}_1) = (\Sigma^{22})^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad (\text{free of } \tilde{X}_1),$$

where, in this case, the matrix $\Sigma_{21}\Sigma_{11}^{-1}$ is referred to as *the matrix of regression coefficients*.

Example 3.3a.1. Let \tilde{X} , $\tilde{\mu} = E[\tilde{X}]$, $\Sigma = \text{Cov}(\tilde{X})$ be as follows:

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \tilde{\mu} = \begin{bmatrix} 1+i \\ 2-i \\ 3i \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 1+i & 0 \\ 1-i & 2 & i \\ 0 & -i & 3 \end{bmatrix}.$$

Consider the partitioning

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Sigma_{11} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}, \Sigma_{12} = \begin{bmatrix} 0 \\ i \end{bmatrix}, \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}, \tilde{X}_1 = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix},$$

$$\Sigma_{21} = [0, -i], \Sigma_{22} = (3), \tilde{X}_2 = (\tilde{x}_3),$$

where \tilde{x}_j , $j = 1, 2, 3$ are scalar complex variables and $\tilde{X} \sim \tilde{N}_3(\tilde{\mu}, \Sigma)$. Determine (1) the marginal densities of \tilde{X}_1 and \tilde{X}_2 ; (2) the conditional expectation of $\tilde{X}_1|\tilde{X}_2$ or $E[\tilde{X}_1|\tilde{X}_2]$ and the conditional expectation of $\tilde{X}_2|\tilde{X}_1$ or $E[\tilde{X}_2|\tilde{X}_1]$; (3) the conditional densities of $\tilde{X}_1|\tilde{X}_2$ and $\tilde{X}_2|\tilde{X}_1$.

Solution 3.3a.1. Note that $\Sigma = \Sigma^*$ and hence Σ is Hermitian. Let us compute the leading minors of Σ : $\det((3)) = 3 > 0$, $\det\left(\begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}\right) = 6 - (1+1) = 4 > 0$,

$$\det(\Sigma) = 3\left[\det\left(\begin{bmatrix} 2 & i \\ -i & 3 \end{bmatrix}\right)\right] - (1+i)\left[\det\left(\begin{bmatrix} 1-i & i \\ 0 & 3 \end{bmatrix}\right)\right] + 0$$

$$= (3)(5) - 3(1+1) = 9 > 0.$$

Hence Σ is Hermitian positive definite. Note that the cofactor expansion for determinants holds whether the elements present in the determinant are real or complex. Let us compute the inverses of the submatrices by taking the transpose of the matrix of cofactors divided by the determinant. This formula applies whether the elements comprising the matrix are real or complex. Then

$$\Sigma_{22}^{-1} = \frac{1}{3}, \Sigma_{11}^{-1} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}^{-1} = \frac{1}{4} \begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}; \quad (i)$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} - \begin{bmatrix} 0 \\ i \end{bmatrix} \left(\frac{1}{3}\right) [0, -i]$$

$$= \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1+i \\ 1-i & \frac{5}{3} \end{bmatrix} \Rightarrow \quad (ii)$$

$$[\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]^{-1} = \begin{bmatrix} 3+1+i & \\ 1-i & \frac{5}{3} \end{bmatrix}^{-1} = \frac{1}{3} \begin{bmatrix} \frac{5}{3} & -(1+i) \\ -(1-i) & 3 \end{bmatrix}; \quad (iii)$$

$$\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 3 - [0, -i] \left(\frac{1}{4} \begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}\right) \begin{bmatrix} 0 \\ i \end{bmatrix}$$

$$= 3 - \frac{3}{4} = \frac{9}{4} \Rightarrow \quad (iv)$$

$$[\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]^{-1} = \frac{4}{9}. \quad (v)$$

As well,

$$\Sigma_{12}\Sigma_{22}^{-1} = \frac{1}{3} \begin{bmatrix} 0 \\ i \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{i}{3} \end{bmatrix}; \quad (vi)$$

$$\Sigma_{21}\Sigma_{11}^{-1} = [0, -i] \left(\frac{1}{4} \begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix} \right) = \frac{1}{4} [1+i, -3i]. \quad (vii)$$

With these computations, all the questions can be answered. We have

$$\tilde{X}_1 \sim \tilde{N}_2(\tilde{\mu}_{(1)}, \Sigma_{11}), \quad \tilde{\mu}_{(1)} = \begin{bmatrix} 1+i \\ 2-i \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}$$

and $\tilde{X}_2 = \tilde{x}_3 \sim \tilde{N}_1(3i, 3)$. Let the densities of \tilde{X}_1 and $\tilde{X}_2 = \tilde{x}_3$ be denoted by $\tilde{f}_1(\tilde{X}_1)$ and $\tilde{f}_2(\tilde{x}_3)$, respectively. Then

$$\begin{aligned} \tilde{f}_2(\tilde{x}_3) &= \frac{1}{(\pi)(3)} e^{-\frac{1}{3}(\tilde{x}_3-3i)^*(\tilde{x}_3-3i)}, \\ \tilde{f}_1(\tilde{X}_1) &= \frac{1}{(\pi^2)(4)} e^{-Q_1}, \\ Q_1 &= \frac{1}{4} [2(\tilde{x}_1 - (1+i))^*(\tilde{x}_1 - (1+i)) \\ &\quad - (1+i)(\tilde{x}_1 - (1+i))^*(\tilde{x}_2 - (2-i)) - (1-i)(\tilde{x}_2 \\ &\quad - (2-i))^*(\tilde{x}_1 - (1+i)) + 3(\tilde{x}_2 - (2-i))^*(\tilde{x}_2 - (2-i))]. \end{aligned}$$

The conditional densities, denoted by $\tilde{g}_1(\tilde{X}_1|\tilde{X}_2)$ and $\tilde{g}_2(\tilde{X}_2|\tilde{X}_1)$, are the following:

$$\begin{aligned} \tilde{g}_1(\tilde{X}_1|\tilde{X}_2) &= \frac{1}{(\pi^2)(3)} e^{-Q_2}, \\ Q_2 &= \frac{1}{3} \left[\frac{5}{3}(\tilde{x}_1 - (1+i))^*(\tilde{x}_1 - (1+i)) \right. \\ &\quad - (1+i)(\tilde{x}_1 - (1+i))^*(\tilde{x}_2 - (2-i)) - \frac{i}{3}(\tilde{x}_3 - 3i) \\ &\quad - (1-i)(\tilde{x}_2 - (2-i)) - \frac{i}{3}(\tilde{x}_3 - 3i)^*(\tilde{x}_1 - (1+i)) \\ &\quad \left. + 3(\tilde{x}_2 - (2-i)) - \frac{i}{3}(\tilde{x}_3 - 3i)^*(\tilde{x}_2 - (2-i)) - \frac{i}{3}(\tilde{x}_3 - 3i) \right]; \end{aligned}$$

$$\begin{aligned}\tilde{g}_2(\tilde{X}_2|\tilde{X}_1) &= \frac{1}{(\pi)(9/4)}e^{-Q_3}, \\ Q_3 &= \frac{4}{9}[(\tilde{x}_3 - M_3)^*(\tilde{x}_3 - M_3)] \text{ where} \\ M_3 &= 3i + \frac{1}{4}\{(1+i)[\tilde{x}_1 - (1+i)] - 3i[\tilde{x}_2 - (2-i)]\} \\ &= 3i + \frac{1}{4}\{(1+i)\tilde{x}_1 - 3i\tilde{x}_2 + 3 + 4i\}.\end{aligned}$$

The bivariate complex Gaussian case

Letting ρ denote the correlation between \tilde{x}_1 and \tilde{x}_2 , it is seen from (3.3a.5) that for $p = 2$,

$$\begin{aligned}E[\tilde{X}_1|\tilde{X}_2] &= \tilde{\mu}_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) = \tilde{\mu}_1 + \frac{\sigma_{12}}{\sigma_2^2}(\tilde{x}_2 - \tilde{\mu}_2) \\ &= \tilde{\mu}_1 + \frac{\sigma_1\sigma_2\rho}{\sigma_2^2}(\tilde{x}_2 - \tilde{\mu}_2) = \tilde{\mu}_1 + \frac{\sigma_1}{\sigma_2}\rho(\tilde{x}_2 - \tilde{\mu}_2) = E[\tilde{x}_1|\tilde{x}_2] \quad (\text{linear in } \tilde{x}_2).\end{aligned}\tag{3.3a.8}$$

Similarly,

$$E[\tilde{x}_2|\tilde{x}_1] = \tilde{\mu}_2 + \frac{\sigma_2}{\sigma_1}\rho(\tilde{x}_1 - \tilde{\mu}_1) \quad (\text{linear in } \tilde{x}_1).\tag{3.3a.9}$$

Incidentally, σ_{12}/σ_2^2 and σ_{12}/σ_1^2 are referred to as *the regression coefficients*.

Exercises 3.3

3.3.1. Let the real $p \times 1$ vector X have a p -variate nonsingular normal density $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let $u = X'\Sigma^{-1}X$. Make use of the mgf to derive the density of u for (1) $\mu = O$, (2) $\mu \neq O$.

3.3.2. Repeat Exercise 3.3.1 for $\mu \neq O$ for the complex nonsingular Gaussian case.

3.3.3. Observing that the density coming from Exercise 3.3.1 is a noncentral chi-square density, coming from the real p -variate Gaussian, derive the non-central F (the numerator chisquare is noncentral and the denominator chisquare is central) density with m and n degrees of freedom and the two chisquares are independently distributed.

3.3.4. Repeat Exercise 3.3.3 for the complex Gaussian case.

3.3.5. Taking the density of u in Exercise 3.3.1 as a real noncentral chisquare density, derive the density of a real doubly noncentral F.

3.3.6. Repeat Exercise 3.3.5 for the corresponding complex case.

3.3.7. Construct a 3×3 Hermitian positive definite matrix V . Let this be the covariance matrix of a 3×1 vector variable \tilde{X} . Compute V^{-1} . Then construct a Gaussian density for this \tilde{X} . Derive the marginal joint densities of (1) \tilde{x}_1 and \tilde{x}_2 , (2) \tilde{x}_1 and \tilde{x}_3 , (3) \tilde{x}_2 and \tilde{x}_3 , where $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ are the components of \tilde{X} . Take $E[\tilde{X}] = O$.

3.3.8. In Exercise 3.3.7, compute (1) $E[\tilde{x}_1|\tilde{x}_2]$, (2) the conditional joint density of \tilde{x}_1, \tilde{x}_2 , given \tilde{x}_3 . Take $E[\tilde{X}] = \tilde{\mu} \neq O$.

3.3.9. In Exercise 3.3.8, compute the mgf in the conditional space of \tilde{x}_1 given \tilde{x}_2, \tilde{x}_3 , that is, $E[e^{t_1\tilde{x}_1}|\tilde{x}_2, \tilde{x}_3]$.

3.3.10. In Exercise 3.3.9, compute the mgf in the marginal space of \tilde{x}_2, \tilde{x}_3 . What is the connection of the results obtained in Exercises 3.3.9 and 3.3.10 with the mgf of \tilde{X} ?

3.4. Chisquaredness and Independence of Quadratic Forms in the Real Case

Let the $p \times 1$ vector X have a p -variate real Gaussian density with a null vector as its mean value and the identity matrix as its covariance matrix, that is, $X \sim N_p(O, I)$, that is, the components of X are mutually independently distributed real scalar standard normal variables. Let $u = X'AX$, $A = A'$ be a real quadratic form in this X . The chisquaredness of a quadratic form such as u has already been discussed in Chap. 2. In this section, we will start with such a u and then consider its generalizations. When $A = A'$, there exists an orthonormal matrix P , that is, $PP' = I$, $P'P = I$, such that $P'AP = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A . Letting $Y = P'X$, $E[Y] = P'O = O$ and $\text{Cov}(Y) = P'IP = I$. But Y is a linear function of X and hence, Y is also real Gaussian distributed; thus, $Y \sim N_p(O, I)$. Then, $y_j^2 \stackrel{iid}{\sim} \chi_1^2$, $j = 1, \dots, p$, or the y_j^2 's are independently distributed chisquares, each having one degree of freedom. Note that

$$u = X'AX = Y'P'APY = \lambda_1 y_1^2 + \dots + \lambda_p y_p^2. \quad (3.4.1)$$

We have the following result on the chisquaredness of quadratic forms in the real p -variate Gaussian case, which corresponds to Theorem 2.2.1.

Theorem 3.4.1. *Let the $p \times 1$ vector be real Gaussian with the parameters $\mu = O$ and $\Sigma = I$ or $X \sim N_p(O, I)$. Let $u = X'AX$, $A = A'$ be a quadratic form in this X . Then $u = X'AX \sim \chi_r^2$, that is, a real chisquare with r degrees of freedom, if and only if $A = A^2$ and the rank of A is r .*

Proof: When $A = A'$, we have the representation of the quadratic form given in (3.4.1). When $A = A^2$, all the eigenvalues of A are 1's and 0's. Then r of the λ_j 's are unities and the remaining ones are zeros and then (3.4.1) becomes the sum of r independently distributed real chisquares having one degree of freedom each, and hence the sum is a real chisquare with r degrees of freedom. For proving the second part, we will assume that $u = X'AX \sim \chi_r^2$. Then the mgf of u is $M_u(t) = (1 - 2t)^{-\frac{r}{2}}$ for $1 - 2t > 0$. The representation in (3.4.1) holds in general. The mgf of y_j^2 , $\lambda_j y_j^2$ and the sum of $\lambda_j y_j^2$ are the following:

$$M_{y_j^2}(t) = (1 - 2t)^{-\frac{1}{2}}, \quad M_{\lambda_j y_j^2}(t) = (1 - 2\lambda_j t)^{-\frac{1}{2}}, \quad M_u(t) = \prod_{j=1}^p (1 - 2\lambda_j t)^{-\frac{1}{2}}$$

for $1 - \lambda_j t > 0$, $j = 1, \dots, p$. Hence, we have the following identity:

$$(1 - 2t)^{-\frac{r}{2}} = \prod_{j=1}^p (1 - 2\lambda_j t)^{-\frac{1}{2}}, \quad 1 - 2t > 0, \quad 1 - 2\lambda_j t > 0, \quad j = 1, \dots, p. \quad (3.4.2)$$

Taking natural logarithm on both sides of (3.4.2), expanding and then comparing the coefficients of $2t$, $\frac{(2t)^2}{2}$, \dots , we have

$$r = \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \lambda_j^2 = \sum_{j=1}^p \lambda_j^3 = \dots \quad (3.4.3)$$

The only solution (3.4.3) can have is that r of the λ_j 's are unities and the remaining ones zeros. This property alone will not guarantee that A is idempotent. However, having eigenvalues that are equal to zero or one combined with the property that $A = A'$ will ensure that $A = A^2$. This completes the proof.

Let us look into some generalizations of the Theorem 3.4.1. Let the $p \times 1$ vector have a real Gaussian distribution $X \sim N_p(O, \Sigma)$, $\Sigma > O$, that is, X is a Gaussian vector with the null vector as its mean value and a real positive definite matrix as its covariance matrix. When Σ is positive definite, we can define $\Sigma^{\frac{1}{2}}$. Letting $Z = \Sigma^{-\frac{1}{2}}X$, Z will be distributed as a standard Gaussian vector, that is, $Z \sim N_p(O, I)$, since Z is a linear function of X with $E[Z] = O$ and $\text{Cov}(Z) = I$. Now, Theorem 3.4.1 is applicable to Z . Then $u = X'AX$, $A = A'$, becomes

$$u = Z' \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} Z, \quad \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}} = (\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})',$$

and it follows from Theorem 3.4.1 that the next result holds:

Theorem 3.4.2. Let the $p \times 1$ vector X have a real p -variate Gaussian density $X \sim N_p(O, \Sigma)$, $\Sigma > O$. Then $q = X'AX$, $A = A'$, is a real chisquare with r degrees of freedom if and only if $\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$ is idempotent and of rank r or, equivalently, if and only if $A = A\Sigma A$ and the rank of A is r .

Now, let us consider the general case. Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let $q = X'AX$, $A = A'$. Then, referring to representation (2.2.1), we can express q as

$$\lambda_1(u_1 + b_1)^2 + \cdots + \lambda_p(u_p + b_p)^2 \equiv \lambda_1 w_1^2 + \cdots + \lambda_p w_p^2 \quad (3.4.4)$$

where $U = (u_1, \dots, u_p)' \sim N_p(O, I)$, the λ_j 's, $j = 1, \dots, p$, are the eigenvalues of $\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}$ and b_i is the i -th component of $P'\Sigma^{-\frac{1}{2}}\mu$, P being a $p \times p$ orthonormal matrix whose j -th column consists of the normalized eigenvectors corresponding to λ_j , $j = 1, \dots, p$. When $\mu = O$, w_j^2 is a real central chisquare random variable having one degree of freedom; otherwise, it is a real noncentral chisquare random variable with one degree of freedom and noncentrality parameter $\frac{1}{2}b_j^2$. Thus, in general, (3.4.4) is a linear function of independently distributed real noncentral chisquare random variables having one degree of freedom each.

Example 3.4.1. Let $X \sim N_3(O, \Sigma)$, $q = X'AX$ where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \Sigma = \frac{1}{3} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

(1) Show that $q \sim \chi_1^2$ by applying Theorem 3.4.2 as well as independently; (2) If the mean value vector $\mu' = [-1, 1, -2]$, what is then the distribution of q ?

Solution 3.4.1. In (1) $\mu = O$ and

$$X'AX = x_1^2 + x_2^2 + x_3^2 + 2(x_1x_2 + x_1x_3 + x_2x_3) = (x_1 + x_2 + x_3)^2.$$

Let $y_1 = x_1 + x_2 + x_3$. Then $E[y_1] = 0$ and

$$\begin{aligned} \text{Var}(y_1) &= \text{Var}(x_1) + \text{Var}(x_2) + \text{Var}(x_3) + 2[\text{Cov}(x_1, x_2) + \text{Cov}(x_1, x_3) + \text{Cov}(x_2, x_3)] \\ &= \frac{1}{3}[2 + 2 + 3 + 0 - 2 - 2] = \frac{3}{3} = 1. \end{aligned}$$

Hence, $y_1 = x_1 + x_2 + x_3$ has $E[y_1] = 0$ and $\text{Var}(y_1) = 1$, and since it is a linear function of the real normal vector X , y_1 is a standard normal. Accordingly, $q = y_1^2 \sim \chi_1^2$. In order to apply Theorem 3.4.2, consider $A\Sigma A$:

$$A\Sigma A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \left(\frac{1}{3} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix} \right) \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = A.$$

Then, by Theorem 3.4.2, $q = X'AX \sim \chi_r^2$ where r is the rank of A . In this case, the rank of A is 1 and hence $y \sim \chi_1^2$. This completes the calculations in connection with respect to (1). When $\mu \neq O$, $u \sim \chi_1^2(\lambda)$, a noncentral chisquare with noncentrality parameter $\lambda = \frac{1}{2}\mu' \Sigma^{-1}\mu$. Let us compute Σ^{-1} by making use of the formula $\Sigma^{-1} = \frac{1}{|\Sigma|}[\text{Cof}(\Sigma)]'$ where $\text{Cof}(\Sigma)$ is the matrix of cofactors of Σ wherein each of its elements is replaced by its cofactor. Now,

$$\Sigma^{-1} = \left\{ \frac{1}{3} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix} \right\}^{-1} = \frac{3}{8} \begin{bmatrix} 5 & 1 & 2 \\ 1 & 5 & 2 \\ 2 & 2 & 4 \end{bmatrix} = \Sigma^{-1}.$$

Then,

$$\lambda = \frac{1}{2}\mu' \Sigma^{-1}\mu = \frac{3}{16}[-1, 1, -2] \begin{bmatrix} 5 & 1 & 2 \\ 1 & 5 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix} = \frac{3}{16} \times 24 = \frac{9}{2}.$$

This completes the computations for the second part.

3.4.1. Independence of quadratic forms

Another relevant result in the real case pertains to the independence of quadratic forms. The concept of chisquaredness and the independence of quadratic forms are prominently encountered in the theoretical underpinnings of statistical techniques such as the Analysis of Variance, Regression and Model Building when it is assumed that the errors are normally distributed. First, we state a result on the independence of quadratic forms in Gaussian vectors whose components are independently distributed.

Theorem 3.4.3. *Let $u_1 = X'AX$, $A = A'$, and $u_2 = X'BX$, $B = B'$, be two quadratic forms in $X \sim N_p(\mu, I)$. Then u_1 and u_2 are independently distributed if and only if $AB = O$.*

Note that independence property holds whether $\mu = O$ or $\mu \neq O$. The result will still be valid if the covariance matrix is $\sigma^2 I$ where σ^2 is a positive real scalar quantity. If the covariance matrix is $\Sigma > O$, the statement of Theorem 3.4.3 needs modification.

Proof: Since $AB = O$, we have $AB = O = O' = (AB)' = B'A' = BA$, which means that A and B commute. Then there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, that diagonalizes both A and B , and

$$\begin{aligned} AB = O &\Rightarrow PABP = O \Rightarrow P'APP'BP = D_1D_2 = O, \\ D_1 &= \text{diag}(\lambda_1, \dots, \lambda_p), \quad D_2 = \text{diag}(v_1, \dots, v_p), \end{aligned} \quad (3.4.5)$$

where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A and v_1, \dots, v_p are the eigenvalues of B . Let $Y = P'X$, then the canonical representations of u_1 and u_2 are the following:

$$u_1 = \lambda_1 y_1^2 + \dots + \lambda_p y_p^2 \quad (3.4.6)$$

$$u_2 = v_1 y_1^2 + \dots + v_p y_p^2 \quad (3.4.7)$$

where y_j 's are real and independently distributed. But $D_1D_2 = O$ means that whenever a $\lambda_j \neq 0$ then the corresponding $v_j = 0$ and vice versa. In other words, whenever a y_j is present in (3.4.6), it is absent in (3.4.7) and vice versa, or the independent variables y_j 's are separated in (3.4.6) and (3.4.7), which implies that u_1 and u_2 are independently distributed.

The necessity part of the proof which consists in showing that $AB = O$ given that $A = A'$, $B = B'$ and u_1 and u_2 are independently distributed, cannot be established by retracing the steps utilized for proving the sufficiency as it requires more matrix manipulations. We note that there are several incorrect or incomplete proofs of Theorem 3.4.3 in the statistical literature. A correct proof for the central case is given in Mathai and Provost (1992).

If $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$, consider the transformation $Y = \Sigma^{-\frac{1}{2}}X \sim N_p(\Sigma^{-\frac{1}{2}}\mu, I)$. Then, $u_1 = X'AX = Y'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Y$, $u_2 = X'BX = Y'\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}}Y$, and we can apply Theorem 3.4.3. In that case, the matrices being orthogonal means

$$\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}B\Sigma^{\frac{1}{2}} = O \Rightarrow A\Sigma B = O.$$

Thus we have the following result:

Theorem 3.4.4. *Let $u_1 = X'AX$, $A = A'$ and $u_2 = X'BX$, $B = B'$ where $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Then u_1 and u_2 are independently distributed if and only if $A\Sigma B = O$.*

What about the distribution of the quadratic form $y = (X - \mu)' \Sigma^{-1} (X - \mu)$ that is present in the exponent of the p -variate real Gaussian density? Let us first determine the mgf of y , that is,

$$\begin{aligned} M_y(t) &= E[e^{ty}] = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \int_X e^{t(X-\mu)' \Sigma^{-1} (X-\mu) - \frac{1}{2} (X-\mu)' \Sigma^{-1} (X-\mu)} dX \\ &= \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \int_X e^{-\frac{1}{2} (1-2t) (X-\mu)' \Sigma^{-1} (X-\mu)} dX \\ &= (1 - 2t)^{-\frac{p}{2}} \text{ for } (1 - 2t) > 0. \end{aligned} \tag{3.4.8}$$

This is the mgf of a real chisquare random variable having p degrees of freedom. Hence we have the following result:

Theorem 3.4.5. When $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$,

$$y = (X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi_p^2, \tag{3.4.9}$$

and if $y_1 = X' \Sigma^{-1} X$, then $y_1 \sim \chi_p^2(\lambda)$, that is, a real non-central chisquare with p degrees of freedom and noncentrality parameter $\lambda = \frac{1}{2} \mu' \Sigma^{-1} \mu$.

Example 3.4.2. Let $X \sim N_3(\mu, \Sigma)$ and consider the quadratic forms $u_1 = X' A X$ and $u_2 = X' B X$ where

$$\begin{aligned} X &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \Sigma = \frac{1}{3} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \\ B &= \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}. \end{aligned}$$

Show that u_1 and u_2 are independently distributed.

Solution 3.4.2. Let J be a 3×1 column vector of unities or 1's as its elements. Then observe that $A = J J'$ and $B = I - \frac{1}{3} J J'$. Further, $J' J = 3$, $J' \Sigma = J'$ and hence $A \Sigma = J J' \Sigma = J J'$. Then $A \Sigma B = J J' [I - \frac{1}{3} J J'] = J J' - J J' = O$. It then follows from Theorem 3.4.4 that u_1 and u_2 are independently distributed. Now, let us prove the result independently without resorting to Theorem 3.4.4. Note that $u_3 = x_1 + x_2 + x_3 = J' X$ has a standard normal distribution as shown in Example 3.4.1. Consider the B in $B X$, namely $I - \frac{1}{3} J J'$. The first component of $B X$ is of the

form $\frac{1}{3}[2, -1, -1]X = \frac{1}{3}[2x_1 - x_2 - x_3]$, which shall be denoted by u_4 . Then u_3 and u_4 are linear functions of the same real normal vector X , and hence u_3 and u_4 are real normal variables. Let us compute the covariance between u_3 and u_4 , observing that $J'\Sigma = J' = J'\text{Cov}(X)$:

$$\text{Cov}(u_3, u_4) = \frac{1}{3}[1, 1, 1]\text{Cov}(X) \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} = \frac{1}{3}[1, 1, 1] \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} = 0.$$

Thus, u_3 and u_4 are independently distributed. As a similar result can be established with respect to the second and third component of BX , u_3 and BX are indeed independently distributed. This implies that $u_3^2 = (J'X)^2 = X'JJ'X = X'AX$ and $(BX)'(BX) = X'B'BX = X'BX$ are independently distributed. Observe that since B is symmetric and idempotent, $B'B = B$. This solution makes use of the following property: if Y_1 and Y_2 are real vectors or matrices that are independently distributed, then $Y_1'Y_1$ and $Y_2'Y_2$ are also independently distributed. It should be noted that the converse does not necessarily hold.

3.4a. Chisquaredness and Independence in the Complex Gaussian Case

Let the $p \times 1$ vector \tilde{X} in the complex domain have a p -variate complex Gaussian density $\tilde{X} \sim \tilde{N}_p(O, I)$. Let $\tilde{u} = \tilde{X}^*A\tilde{X}$ be a Hermitian form, $A = A^*$ where A^* denotes the conjugate transpose of A . Then there exists a unitary matrix Q , $QQ^* = I$, $Q^*Q = I$, such that $Q^*AQ = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A . It can be shown that when A is Hermitian, which means in the real case that $A = A'$ (symmetric), all the eigenvalues of A are real. Let $\tilde{Y} = Q^*\tilde{X}$ then

$$\tilde{u} = \tilde{X}^*A\tilde{X} = \tilde{Y}^*Q^*AQ\tilde{Y} = \lambda_1|\tilde{y}_1|^2 + \dots + \lambda_p|\tilde{y}_p|^2 \quad (3.4a.1)$$

where $|\tilde{y}_j|$ denotes the absolute value or modulus of \tilde{y}_j . If $\tilde{y}_j = y_{j1} + iy_{j2}$ where y_{j1} and y_{j2} are real, $i = \sqrt{-1}$, then $|\tilde{y}_j|^2 = y_{j1}^2 + y_{j2}^2$. We can obtain the following result which is the counterpart of Theorem 3.4.1:

Theorem 3.4a.1. *Let $\tilde{X} \sim \tilde{N}_p(O, I)$ and $\tilde{u} = \tilde{X}^*A\tilde{X}$, $A = A^*$. Then $\tilde{u} \sim \tilde{\chi}_r^2$, a chisquare random variable having r degrees of freedom in the complex domain, if and only if $A = A^2$ (idempotent) and A is of rank r .*

Proof: The definition of an idempotent matrix A as $A = A^2$ holds whether the elements of A are real or complex. Let A be idempotent and of rank r . Then r of the eigenvalues of A are unities and the remaining ones are zeros. Then the representation given in (3.4a.1) becomes

$$\tilde{u} = |\tilde{y}_1|^2 + \dots + |\tilde{y}_r|^2 \sim \tilde{\chi}_r^2,$$

a chisquare with r degrees of freedom in the complex domain, that is, a real gamma with the parameters $(\alpha = r, \beta = 1)$ whose mgf is $(1 - t)^{-r}, 1 - t > 0$. For proving the necessity, let us assume that $\tilde{u} \sim \tilde{\chi}_r^2$, its mgf being $M_{\tilde{u}}(t) = (1 - t)^{-r}$ for $1 - t > 0$. But from (3.4a.1), $|\tilde{y}_j|^2 \sim \tilde{\chi}_1^2$ and its mgf is $(1 - t)^{-1}$ for $1 - t > 0$. Hence the mgf of $\lambda_j |\tilde{y}_j|^2$ is $M_{\lambda_j |\tilde{y}_j|^2}(t) = (1 - \lambda_j t)^{-1}$ for $1 - \lambda_j t > 0$, and we have the following identity:

$$(1 - t)^{-r} = \prod_{j=1}^p (1 - \lambda_j t)^{-1}. \tag{3.4a.2}$$

Take the natural logarithm on both sides of (3.4a.2), expand and compare the coefficients of $t, \frac{t^2}{2}, \dots$ to obtain

$$r = \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \lambda_j^2 = \dots \tag{3.4a.3}$$

The only possibility for the λ_j 's in (3.4a.3) is that r of them are unities and the remaining ones, zeros. This property, combined with $A = A^*$ guarantees that $A = A^2$ and A is of rank r . This completes the proof.

An extension of Theorem 3.4a.1 which is the counterpart of Theorem 3.4.2 can also be obtained. We will simply state it as the proof is parallel to that provided in the real case.

Theorem 3.4a.2. *Let $\tilde{X} \sim \tilde{N}_p(O, \Sigma), \Sigma > O$ and $\tilde{u} = \tilde{X}^* A \tilde{X}, A = A^*$, be a Hermitian form. Then $\tilde{u} \sim \tilde{\chi}_r^2$, a chisquare random variable having r degrees of freedom in the complex domain, if and only if $A = A \Sigma A$ and A is of rank r .*

Example 3.4a.1. Let $\tilde{X} \sim \tilde{N}_3(\tilde{\mu}, \Sigma), \tilde{u} = \tilde{X}^* A \tilde{X}$ where

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \Sigma = \frac{1}{3} \begin{bmatrix} 3 & -(1+i) & -(1-i) \\ -(1-i) & 3 & -(1+i) \\ -(1+i) & -(1-i) & 3 \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \tilde{\mu} = \begin{bmatrix} 2+i \\ -i \\ 2i \end{bmatrix}.$$

First determine whether Σ can be a covariance matrix. Then determine the distribution of \tilde{u} by making use of Theorem 3.4a.2 as well as independently, that is, without using Theorem 3.4a.2, for the cases (1) $\tilde{\mu} = O$; (2) $\tilde{\mu}$ as given above.

Solution 3.4a.1. Note that $\Sigma = \Sigma^*$, that is, Σ is Hermitian. Let us verify that Σ is a Hermitian positive definite matrix. Note that Σ must be either positive definite or positive semi-definite to be a covariance matrix. In the semi-definite case, the density of \tilde{X} does not exist. Let us check the leading minors: $\det((3)) = 3 > 0$, $\det\left(\begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}\right) = 9 - 2 = 7 > 0$, $\det(\Sigma) = \frac{13}{3^3} > 0$ [evaluated by using the cofactor expansion which is the same in the complex case]. Hence Σ is Hermitian positive definite. In order to apply Theorem 3.4a.2, we must now verify that $A\Sigma A = A$ when $\tilde{\mu} = O$. Observe the following: $A = JJ'$, $J'A = 3J'$, $J'J = 3$ where $J' = [1, 1, 1]$. Hence $A\Sigma A = (JJ')\Sigma(JJ') = J(J'\Sigma)JJ' = \frac{1}{3}(JJ')(JJ') = \frac{1}{3}J(J'J)J' = \frac{1}{3}J(3)J' = JJ' = A$. Thus the condition holds and by Theorem 3.4a.2, $\tilde{u} \sim \tilde{\chi}_1^2$ in the complex domain, that is, \tilde{u} a real gamma random variable with parameters $(\alpha = 1, \beta = 1)$ when $\tilde{\mu} = O$. Now, let us derive this result without using Theorem 3.4a.2. Let $\tilde{u}_1 = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3$ and $A_1 = (1, 1, 1)'$. Note that $A_1'\tilde{X} = \tilde{u}_1$, the sum of the components of \tilde{X} . Hence $\tilde{u}_1^*\tilde{u}_1 = \tilde{X}^*A_1A_1'\tilde{X} = \tilde{X}^*A\tilde{X}$. For $\tilde{\mu} = O$, we have $E[\tilde{u}_1] = 0$ and

$$\begin{aligned} \text{Var}(\tilde{u}_1) &= \text{Var}(\tilde{x}_1) + \text{Var}(\tilde{x}_2) + \text{Var}(\tilde{x}_3) + [\text{Cov}(\tilde{x}_1, \tilde{x}_2) + \text{Cov}(\tilde{x}_2, \tilde{x}_1)] \\ &\quad + [\text{Cov}(\tilde{x}_1, \tilde{x}_3) + \text{Cov}(\tilde{x}_3, \tilde{x}_1)] + [\text{Cov}(\tilde{x}_2, \tilde{x}_3) + \text{Cov}(\tilde{x}_3, \tilde{x}_2)] \\ &= \frac{1}{3}\{3 + 3 + 3 + [-(1+i) - (1-i)] + [-(1-i) - (1+i)] \\ &\quad + [-(1+i) - (1-i)]\} = \frac{1}{3}[9 - 6] = 1. \end{aligned}$$

Thus, \tilde{u}_1 is a standard normal random variable in the complex domain and $\tilde{u}_1^*\tilde{u}_1 \sim \tilde{\chi}_1^2$, a chisquare random variable with one degree of freedom in the complex domain, that is, a real gamma random variable with parameters $(\alpha = 1, \beta = 1)$.

For $\tilde{\mu} = (2 + i, -i, 2i)'$, this chisquare random variable is noncentral with noncentrality parameter $\lambda = \tilde{\mu}^*\Sigma^{-1}\tilde{\mu}$. Hence, the inverse of Σ has to be evaluated. To do so, we will employ the formula $\Sigma^{-1} = \frac{1}{|\Sigma|}[\text{Cof}(\Sigma)]'$, which also holds for the complex case. Earlier, the determinant was found to be equal to $\frac{13}{3^3}$ and

$$\begin{aligned} \frac{1}{|\Sigma|}[\text{Cof}(\Sigma)] &= \frac{3^3}{13} \begin{bmatrix} 7 & 3-i & 3+i \\ 3+i & 7 & 3-i \\ 3-i & 3+i & 7 \end{bmatrix}; \text{ then} \\ \Sigma^{-1} &= \frac{1}{|\Sigma|}[\text{Cof}(\Sigma)]' = \frac{3^3}{13} \begin{bmatrix} 7 & 3+i & 3-i \\ 3-i & 7 & 3+i \\ 3+i & 3-i & 7 \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned}\lambda &= \tilde{\mu}^* \Sigma^{-1} \tilde{\mu} = \frac{3^3}{13} [2 - i, i, -2i] \begin{bmatrix} 7 & 3 + i & 3 - i \\ 3 - i & 7 & 3 + i \\ 3 + i & 3 - i & 7 \end{bmatrix} \begin{bmatrix} 2 + i \\ -i \\ 2i \end{bmatrix} \\ &= \frac{(76)(3^3)}{13} = \frac{2052}{13} \approx 157.85.\end{aligned}$$

This completes the computations.

3.4a.1. Independence of Hermitian forms

We shall mainly state certain results in connection with Hermitian forms in this section since they parallel those pertaining to the real case.

Theorem 3.4a.3. *Let $\tilde{u}_1 = \tilde{X}^* A \tilde{X}$, $A = A^*$, and $\tilde{u}_2 = \tilde{X}^* B \tilde{X}$, $B = B^*$, where $\tilde{X} \sim \tilde{N}(\mu, I)$. Then, \tilde{u}_1 and \tilde{u}_2 are independently distributed if and only if $AB = O$.*

Proof: Let us assume that $AB = O$. Then

$$AB = O = O^* = (AB)^* = B^* A^* = BA. \quad (3.4a.4)$$

This means that there exists a unitary matrix Q , $QQ^* = I$, $Q^*Q = I$, that will diagonalize both A and B . That is, $Q^*AQ = \text{diag}(\lambda_1, \dots, \lambda_p) = D_1$, $Q^*BQ = \text{diag}(v_1, \dots, v_p) = D_2$ where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A and v_1, \dots, v_p are the eigenvalues of B . But $AB = O$ implies that $D_1 D_2 = O$. As well,

$$\tilde{u}_1 = \tilde{X}^* A \tilde{X} = \tilde{Y}^* Q^* A Q \tilde{Y} = \lambda_1 |\tilde{y}_1|^2 + \dots + \lambda_p |\tilde{y}_p|^2, \quad (3.4a.5)$$

$$\tilde{u}_2 = \tilde{X}^* B \tilde{X} = \tilde{Y}^* Q^* B Q \tilde{Y} = v_1 |\tilde{y}_1|^2 + \dots + v_p |\tilde{y}_p|^2. \quad (3.4a.6)$$

Since $D_1 D_2 = O$, whenever a $\lambda_j \neq 0$, the corresponding $v_j = 0$ and vice versa. Thus the independent variables \tilde{y}_j 's are separated in (3.4a.5) and (3.4a.6) and accordingly, \tilde{u}_1 and \tilde{u}_2 are independently distributed. The proof of the necessity which requires more matrix algebra, will not be provided herein. The general result can be stated as follows:

Theorem 3.4a.4. *Letting $\tilde{X} \sim \tilde{N}_p(\mu, \Sigma)$, $\Sigma > O$, the Hermitian forms $\tilde{u}_1 = \tilde{X}^* A \tilde{X}$, $A = A^*$, and $\tilde{u}_2 = \tilde{X}^* B \tilde{X}$, $B = B^*$, are independently distributed if and only if $A \Sigma B = O$.*

Now, consider the density of the exponent in the p -variate complex Gaussian density. What will then be the density of $\tilde{y} = (\tilde{X} - \tilde{\mu})^* \Sigma^{-1} (\tilde{X} - \tilde{\mu})$? Let us evaluate the mgf of \tilde{y} . Observing that \tilde{y} is real so that we may take $E[e^{t\tilde{y}}]$ where t is a real parameter, we have

$$\begin{aligned}
M_{\tilde{y}}(t) &= E[e^{t\tilde{y}}] = \frac{1}{\pi^p |\det(\Sigma)|} \int_{\tilde{X}} e^{t(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu}) - (\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})} d\tilde{X} \\
&= \frac{1}{\pi^p |\det(\Sigma)|} \int_{\tilde{X}} e^{-(1-t)(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})} d\tilde{X} \\
&= (1-t)^{-p} \text{ for } 1-t > 0.
\end{aligned} \tag{3.4a.7}$$

This is the mgf of a real gamma random variable with the parameters $(\alpha = p, \beta = 1)$ or a chisquare random variable in the complex domain with p degrees of freedom. Hence we have the following result:

Theorem 3.4a.5. *When $\tilde{X} \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$ then $\tilde{y} = (\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu})$ is distributed as a real gamma random variable with the parameters $(\alpha = p, \beta = 1)$ or a chisquare random variable in the complex domain with p degrees of freedom, that is,*

$$\tilde{y} \sim \text{gamma}(\alpha = p, \beta = 1) \text{ or } \tilde{y} \sim \tilde{\chi}_p^2. \tag{3.4a.8}$$

Example 3.4a.2. Let $\tilde{X} \sim \tilde{N}_3(\tilde{\mu}, \Sigma)$, $\tilde{u}_1 = \tilde{X}^* A \tilde{X}$, $\tilde{u}_2 = \tilde{X}^* B \tilde{X}$ where

$$\begin{aligned}
\tilde{X} &= \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} 2-i \\ 3+2i \\ 1-i \end{bmatrix}, \quad \Sigma = \frac{1}{3} \begin{bmatrix} 3 & -(1+i) & -(1-i) \\ -(1-i) & 3 & -(1+i) \\ -(1+i) & -(1-i) & 3 \end{bmatrix}, \\
A &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.
\end{aligned}$$

- (1) By making use of Theorem 3.4a.4, show that \tilde{u}_1 and \tilde{u}_2 are independently distributed.
- (2) Show the independence of \tilde{u}_1 and \tilde{u}_2 without using Theorem 3.4a.4.

Solution 3.4a.2. In order to use Theorem 3.4a.4, we have to show that $A\Sigma B = O$ irrespective of $\tilde{\mu}$. Note that $A = JJ'$, $J' = [1, 1, 1]$, $J'J = 3$, $J'\Sigma = \frac{1}{3}J'$, $J'B = O$. Hence $A\Sigma = JJ'\Sigma = J(J'\Sigma) = \frac{1}{3}JJ' \Rightarrow A\Sigma B = \frac{1}{3}JJ'B = \frac{1}{3}J(J'B) = O$. This proves the result that \tilde{u}_1 and \tilde{u}_2 are independently distributed through Theorem 3.4a.4. This will now be established without resorting to Theorem 3.4a.4. Let $\tilde{u}_3 = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = J'X$ and $\tilde{u}_4 = \frac{1}{3}[2\tilde{x}_1 - \tilde{x}_2 - \tilde{x}_3]$ or the first row of $B\tilde{X}$. Since independence is not affected by the relocation of the variables, we may assume, without any loss of generality, that $\tilde{\mu} = O$ when considering the independence of \tilde{u}_3 and \tilde{u}_4 . Let us compute the covariance between \tilde{u}_3 and \tilde{u}_4 :

$$\text{Cov}(\tilde{u}_3, \tilde{u}_4) = \frac{1}{3}[1, 1, 1]\Sigma \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} = \frac{1}{3^2}[1, 1, 1] \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} = 0.$$

Thus, \tilde{u}_3 and \tilde{u}_4 are uncorrelated and hence independently distributed since both are linear functions of the normal vector \tilde{X} . This property holds for each row of $B\tilde{X}$ and therefore \tilde{u}_3 and $B\tilde{X}$ are independently distributed. However, $\tilde{u}_1 = \tilde{X}^*A\tilde{X} = \tilde{u}_3^*\tilde{u}_3$ and hence \tilde{u}_1 and $(B\tilde{X})^*(B\tilde{X}) = \tilde{X}^*B^*B\tilde{X} = \tilde{X}^*B\tilde{X} = \tilde{u}_2$ are independently distributed. This completes the computations. The following property was utilized: Let \tilde{U} and \tilde{V} be vectors or matrices that are independently distributed. Then, all the pairs (\tilde{U}, \tilde{V}^*) , $(\tilde{U}, \tilde{V}\tilde{V}^*)$, $(\tilde{U}, \tilde{V}^*\tilde{V})$, \dots , $(\tilde{U}\tilde{U}^*, \tilde{V}\tilde{V}^*)$, are independently distributed whenever the quantities are defined. The converses need not hold when quadratic terms are involved; for instance, $(\tilde{U}\tilde{U}^*, \tilde{V}\tilde{V}^*)$ being independently distributed need not imply that (\tilde{U}, \tilde{V}) are independently distributed.

Exercises 3.4

3.4.1. In the real case on the right side of (3.4.4), compute the densities of the following items: (i) z_1^2 , (ii) $\lambda_1 z_1^2$, (iii) $\lambda_1 z_1^2 + \lambda_2 z_2^2$, (iv) $\lambda_1 z_1^2 + \dots + \lambda_4 z_4^2$ if $\lambda_1 = \lambda_2$, $\lambda_3 = \lambda_4$ for $\mu = O$.

3.4.2. Compute the density of $u = X'AX$, $A = A'$ in the real case when (i) $X \sim N_p(O, \Sigma)$, $\Sigma > O$, (ii) $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$.

3.4.3. Modify the statement in Theorem 3.4.1 if (i) $X \sim N_p(O, \sigma^2 I)$, $\sigma^2 > 0$, (ii) $X \sim N_p(\mu, \sigma^2 I)$, $\mu \neq O$.

3.4.4. Prove the only if part in Theorem 3.4.3

3.4.5. Establish the cases (i), (ii), (iii) of Exercise 3.4.1 in the corresponding complex domain.

3.4.6. Supply the proof for the only if part in Theorem 3.4a.3.

3.4.7. Can a matrix A having at least one complex element be Hermitian and idempotent at the same time? Prove your statement.

3.4.8. Let the $p \times 1$ vector X have a real Gaussian density $N_p(O, \Sigma)$, $\Sigma > O$. Let $u = X'AX$, $A = A'$. Evaluate the density of u for $p = 2$ and show that this density can be written in terms of a hypergeometric series of the ${}_1F_1$ type.

3.4.9. Repeat Exercise 3.4.8 if \tilde{X} is in the complex domain, $\tilde{X} \sim \tilde{N}_p(O, \Sigma)$, $\Sigma > O$.

3.4.10. Supply the proofs for the only if part in Theorems 3.4.4 and 3.4a.4.

3.5. Samples from a p -variate Real Gaussian Population

Let the $p \times 1$ real vectors X_1, \dots, X_n be iid as $N_p(\mu, \Sigma)$, $\Sigma > O$. Then, the collection X_1, \dots, X_n is called a simple random sample of size n from this $N_p(\mu, \Sigma)$, $\Sigma > O$. Then the joint density of X_1, \dots, X_n is the following:

$$\begin{aligned} L &= \prod_{j=1}^n f(X_j) = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \\ &= [(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}]^{-1} e^{-\frac{1}{2} \sum_{j=1}^n (X_j - \mu)' \Sigma^{-1} (X_j - \mu)}. \end{aligned} \quad (3.5.1)$$

This L at an observed set of X_1, \dots, X_n is called the *likelihood function*. Let the sample matrix, which is $p \times n$, be denoted by a bold-faced \mathbf{X} . In order to avoid too many symbols, we will use \mathbf{X} to denote the $p \times n$ matrix in this section. In earlier sections, we had used X to denote a $p \times 1$ vector. Then

$$\mathbf{X} = [X_1, \dots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix}, \quad X_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{pk} \end{bmatrix}, \quad k = 1, \dots, n. \quad (i)$$

Let the sample average be denoted by $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. Then \bar{X} will be of the following form:

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}, \quad \bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik} = \text{average on the } i\text{-th component of any } X_j. \quad (ii)$$

Let the bold-faced $\bar{\mathbf{X}}$ be defined as follows:

$$\bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}] = \begin{bmatrix} \bar{x}_1 & \bar{x}_1 & \dots & \bar{x}_1 \\ \bar{x}_2 & \bar{x}_2 & \dots & \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_p & \bar{x}_p & \dots & \bar{x}_p \end{bmatrix}.$$

Then,

$$\mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \dots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \dots & x_{2n} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \dots & x_{pn} - \bar{x}_p \end{bmatrix},$$

and

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = (s_{ij}),$$

so that

$$s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

S is called the *sample sum of products matrix* or the *corrected sample sum of products matrix*, corrected in the sense that the averages are deducted from the observations. As well, $\frac{1}{n}s_{ii}$ is called the sample variance on the component x_i of any vector X_k , referring to (i) above, and $\frac{1}{n}s_{ij}$, $i \neq j$, is called the sample covariance on the components x_i and x_j of any X_k , $\frac{1}{n}S$ being referred to as the sample covariance matrix. The exponent in L can be simplified by making use of the following properties: (1) When u is a 1×1 matrix or a scalar quantity, then $\text{tr}(u) = \text{tr}(u') = u = u'$. (2) For two matrices A and B , whenever AB and BA are defined, $\text{tr}(AB) = \text{tr}(BA)$ where AB need not be equal to BA . Observe that the following quantity is real scalar and hence, it is equal to its trace:

$$\begin{aligned} \sum_{j=1}^n (X_j - \mu)' \Sigma^{-1} (X_j - \mu) &= \text{tr} \left[\sum_{j=1}^n (X_j - \mu)' \Sigma^{-1} (X_j - \mu) \right] \\ &= \text{tr} \left[\Sigma^{-1} \sum_{j=1}^n (X_j - \mu)(X_j - \mu)' \right] \\ &= \text{tr} \left[\Sigma^{-1} \sum_{j=1}^n (X_j - \bar{X} + \bar{X} - \mu)(X_j - \bar{X} + \bar{X} - \mu)' \right] \\ &= \text{tr} \left[\Sigma^{-1} (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \right] + n \text{tr} \left[\Sigma^{-1} (\bar{X} - \mu)(\bar{X} - \mu)' \right] \\ &= \text{tr}(\Sigma^{-1} S) + n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) \end{aligned}$$

because

$$\text{tr}(\Sigma^{-1} (\bar{X} - \mu)(\bar{X} - \mu)') = \text{tr}((\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)) = (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu).$$

The right-hand side expression being 1×1 , it is equal to its trace, and L can be written as

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}. \quad (3.5.2)$$

If we wish to estimate the parameters μ and Σ from a set of observation vectors corresponding to X_1, \dots, X_n , one method consists in maximizing L with respect to μ and Σ given those observations and estimating the parameters. By resorting to calculus, L is

differentiated partially with respect to μ and Σ , the resulting expressions are equated to null vectors and matrices, respectively, and these equations are then solved to obtain the solutions for μ and Σ . Those estimates will be called *the maximum likelihood estimates or MLE's*. We will explore this aspect later.

Example 3.5.1. Let the 3×1 vector X_1 be real Gaussian, $X_1 \sim N_3(\mu, \Sigma)$, $\Sigma > O$. Let X_j , $j = 1, 2, 3, 4$ be iid as X_1 . Compute the 3×4 sample matrix \mathbf{X} , the sample average \bar{X} , the matrix of sample means $\bar{\mathbf{X}}$, the sample sum of products matrix S , the maximum likelihood estimates of μ and Σ , based on the following set of observations on X_j , $j = 1, 2, 3, 4$:

$$X_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}, X_4 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

Solution 3.5.1. The 3×4 sample matrix and the sample average are

$$\mathbf{X} = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 2 & 4 & 3 \end{bmatrix}, \bar{X} = \frac{1}{4} \begin{bmatrix} 2+1+1+0 \\ 0-1+0+1 \\ -1+2+4+3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

Then $\bar{\mathbf{X}}$ and $\mathbf{X} - \bar{\mathbf{X}}$ are the following:

$$\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \bar{X}, \bar{X}] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 \end{bmatrix}, \mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ -3 & 0 & 2 & 1 \end{bmatrix},$$

and the sample sum of products matrix S is the following:

$$S = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]' = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ -3 & 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \\ -1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -4 \\ -1 & 2 & 1 \\ -4 & 1 & 14 \end{bmatrix}.$$

Then, the maximum likelihood estimates of μ and Σ , denoted with a hat, are

$$\hat{\mu} = \bar{X} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \hat{\Sigma} = \frac{1}{n}S = \frac{1}{4} \begin{bmatrix} 2 & -1 & -4 \\ -1 & 2 & 1 \\ -4 & 1 & 14 \end{bmatrix}.$$

This completes the computations.

3.5a. Simple Random Sample from a p -variate Complex Gaussian Population

Our population density is given by the following:

$$\tilde{f}(\tilde{X}_j) = \frac{e^{-(\tilde{X}_j - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X}_j - \tilde{\mu})}}{\pi^p |\det(\tilde{\Sigma})|}, \quad \tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma}), \quad \tilde{\Sigma} = \tilde{\Sigma}^* > O.$$

Let $\tilde{X}_1, \dots, \tilde{X}_n$ be a collection of complex vector random variables iid as $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} > O$. This collection is called a simple random sample of size n from this complex Gaussian population $\tilde{f}(\tilde{X}_j)$. We will use notations parallel to those utilized in the real case. Let $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_n]$, $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$, $\bar{\tilde{\mathbf{X}}} = (\bar{\tilde{X}}, \dots, \bar{\tilde{X}})$, and $\tilde{S} = (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})(\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})^* = \tilde{S} = (\tilde{s}_{ij})$. Then

$$\tilde{s}_{ij} = \sum_{k=1}^n (\tilde{x}_{ik} - \bar{\tilde{x}}_i)(\tilde{x}_{jk} - \bar{\tilde{x}}_j)^*$$

with $\frac{1}{n}\tilde{s}_{ij}$ being the sample covariance between the components \tilde{x}_i and \tilde{x}_j , $i \neq j$, of any \tilde{X}_k , $k = 1, \dots, n$, $\frac{1}{n}\tilde{s}_{ii}$ being the sample variance on the component \tilde{x}_i . The joint density of $\tilde{X}_1, \dots, \tilde{X}_n$, denoted by \tilde{L} , is given by

$$\tilde{L} = \prod_{j=1}^n \frac{e^{-(\tilde{X}_j - \mu)^* \tilde{\Sigma}^{-1} (\tilde{X}_j - \mu)}}{\pi^p |\det(\tilde{\Sigma})|} = \frac{e^{-\sum_{j=1}^n (\tilde{X}_j - \mu)^* \tilde{\Sigma}^{-1} (\tilde{X}_j - \mu)}}{\pi^{np} |\det(\tilde{\Sigma})|^n}, \quad (3.5a.1)$$

which can be simplified to the following expression by making use of steps parallel to those utilized in the real case:

$$L = \frac{e^{-\text{tr}(\tilde{\Sigma}^{-1} \tilde{S}) - n(\bar{\tilde{X}} - \mu)^* \tilde{\Sigma}^{-1} (\bar{\tilde{X}} - \mu)}}{\pi^{np} |\det(\tilde{\Sigma})|^n}. \quad (3.5a.2)$$

Example 3.5a.1. Let the 3×1 vector \tilde{X}_1 in the complex domain have a complex trivariate Gaussian distribution $\tilde{X}_1 \sim \tilde{N}_3(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} > O$. Let \tilde{X}_j , $j = 1, 2, 3, 4$ be iid as \tilde{X}_1 . With our usual notations, compute the 3×4 sample matrix $\tilde{\mathbf{X}}$, the sample average $\bar{\tilde{X}}$, the 3×4 matrix of sample averages $\bar{\tilde{\mathbf{X}}}$, the sample sum of products matrix \tilde{S} and the maximum likelihood estimates of $\tilde{\mu}$ and $\tilde{\Sigma}$ based on the following set of observations on \tilde{X}_j , $j = 1, 2, 3, 4$:

$$\tilde{X}_1 = \begin{bmatrix} 1 + i \\ 2 - i \\ 1 - i \end{bmatrix}, \quad \tilde{X}_2 = \begin{bmatrix} -1 + 2i \\ 3i \\ -1 + i \end{bmatrix}, \quad \tilde{X}_3 = \begin{bmatrix} -2 + 2i \\ 3 + i \\ 4 + 2i \end{bmatrix}, \quad \tilde{X}_4 = \begin{bmatrix} -2 + 3i \\ 3 + i \\ -4 + 2i \end{bmatrix}.$$

Solution 3.5a.1. The sample matrix $\tilde{\mathbf{X}}$ and the sample average $\bar{\tilde{\mathbf{X}}}$ are

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1+i & -1+2i & -2+2i & -2+3i \\ 2-i & 3i & 3+i & 3+i \\ 1-i & -1+i & 4+2i & -4+2i \end{bmatrix}, \quad \bar{\tilde{\mathbf{X}}} = \begin{bmatrix} -1+2i \\ 2+i \\ i \end{bmatrix}.$$

Then, with our usual notations, $\bar{\tilde{\mathbf{X}}}$ and $\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}$ are the following:

$$\bar{\tilde{\mathbf{X}}} = \begin{bmatrix} -1+2i & -1+2i & -1+2i & -1+2i \\ 2+i & 2+i & 2+i & 2+i \\ i & i & i & i \end{bmatrix},$$

$$\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}} = \begin{bmatrix} 2-i & 0 & -1 & -1+i \\ -2i & -2+2i & 1 & 1 \\ 1-2i & -1 & 4+i & -4+i \end{bmatrix}.$$

Thus, the sample sum of products matrix $\tilde{\mathbf{S}}$ is

$$\begin{aligned} \tilde{\mathbf{S}} &= [\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}] [\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}]^* \\ &= \begin{bmatrix} 2-i & 0 & -1 & -1+i \\ -2i & -2+2i & 1 & 1 \\ 1-2i & -1 & 4+i & -4+i \end{bmatrix} \begin{bmatrix} 2+i & 2i & 1+2i \\ 0 & -2-2i & -1 \\ -1 & 1 & 4-i \\ -1-i & 1 & -4-i \end{bmatrix} \\ &= \begin{bmatrix} 8 & 5i & 5+i \\ -5i & 14 & 6-6i \\ 5-i & 6+6i & 40 \end{bmatrix}. \end{aligned}$$

The maximum likelihood estimates are as follows:

$$\hat{\mu} = \bar{\tilde{\mathbf{X}}} = \begin{bmatrix} -1+2i \\ 2+i \\ i \end{bmatrix}, \quad \hat{\Sigma} = \frac{1}{4} \tilde{\mathbf{S}}$$

where $\tilde{\mathbf{S}}$ is given above. This completes the computations.

3.5.1. Some simplifications of the sample matrix in the real Gaussian case

The $p \times n$ sample matrix is

$$\mathbf{X} = [X_1, \dots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix}, \quad X_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{pk} \end{bmatrix}$$

where the rows are iid variables on the components of the p -vector X_1 . For example, $(x_{11}, x_{12}, \dots, x_{1n})$ are iid variables distributed as the first component of X_1 . Let

$$\begin{aligned} \bar{X} &= \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{k=1}^n x_{1k} \\ \vdots \\ \frac{1}{n} \sum_{k=1}^n x_{pk} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}(x_{11}, \dots, x_{1n})J \\ \vdots \\ \frac{1}{n}(x_{p1}, \dots, x_{pn})J \end{bmatrix} = \frac{1}{n} \mathbf{X}J, \quad J = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad n \times 1. \end{aligned}$$

Consider the matrix

$$\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X}) = \frac{1}{n} \mathbf{X}J J' = \mathbf{X}B, \quad B = \frac{1}{n} J J'.$$

Then,

$$\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}A, \quad A = I - B = I - \frac{1}{n} J J'.$$

Observe that $A = A^2$, $B = B^2$, $AB = O$, $A = A'$ and $B = B'$ where both A and B are $n \times n$ matrices. Then $\mathbf{X}A$ and $\mathbf{X}B$ are $p \times n$ and, in order to determine the mgf, we will take the $p \times n$ parameter matrices T_1 and T_2 . Accordingly, the mgf of $\mathbf{X}A$ is $M_{\mathbf{X}A}(T_1) = E[e^{\text{tr}(T_1' \mathbf{X}A)}]$, that of $\mathbf{X}B$ is $M_{\mathbf{X}B}(T_2) = E[e^{\text{tr}(T_2' \mathbf{X}B)}]$ and the joint mgf is $E[e^{\text{tr}(T_1' \mathbf{X}A) + \text{tr}(T_2' \mathbf{X}B)}]$. Let us evaluate the joint mgf for $X_j \sim N_p(O, I)$,

$$E[e^{\text{tr}(T_1' \mathbf{X}A) + \text{tr}(T_2' \mathbf{X}B)}] = \int_{\mathbf{X}} \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{\text{tr}(T_1' \mathbf{X}A) + \text{tr}(T_2' \mathbf{X}B) - \frac{1}{2} \text{tr}(\mathbf{X}\mathbf{X}')} d\mathbf{X}.$$

Let us simplify the exponent,

$$-\frac{1}{2} \{ \text{tr}(\mathbf{X}\mathbf{X}') - 2\text{tr}[\mathbf{X}(AT_1' + BT_2')] \}. \tag{i}$$

If we expand $\text{tr}[(\mathbf{X} - C)(\mathbf{X} - C)']$ for some C , we have

$$\begin{aligned} \text{tr}(\mathbf{X}\mathbf{X}') - \text{tr}(C\mathbf{X}') - \text{tr}(\mathbf{X}C') + \text{tr}(CC') \\ = \text{tr}(\mathbf{X}\mathbf{X}') - 2\text{tr}(\mathbf{X}C') + \text{tr}(CC') \end{aligned} \tag{ii}$$

as $\text{tr}(\mathbf{X}C') = \text{tr}(C\mathbf{X}')$ even though $C\mathbf{X}' \neq \mathbf{X}C'$. On comparing (i) and (ii), we have $C' = AT_1' + BT_2'$, and then

$$\begin{aligned} \text{tr}(CC') &= \text{tr}[(T_1A' + T_2B')(AT_1' + BT_2')] \\ &= \text{tr}(T_1A'AT_1') + \text{tr}(T_2B'BT_2') + \text{tr}(T_1A'BT_2') + \text{tr}(T_2B'AT_1'). \end{aligned} \tag{iii}$$

Since the integral over $\mathbf{X} - C$ will absorb the normalizing constant and give 1, the joint mgf is $e^{\text{tr}(CC')}$. Proceeding exactly the same way, it is seen that the mgf of $\mathbf{X}A$ and $\mathbf{X}B$ are respectively

$$M_{\mathbf{X}A}(T_1) = e^{\frac{1}{2}\text{tr}(T_1 A' A T_1')} \quad \text{and} \quad M_{\mathbf{X}B}(T_2) = e^{\frac{1}{2}\text{tr}(T_2 B' B T_2')}.$$

The independence of $\mathbf{X}A$ and $\mathbf{X}B$ implies that the joint mgf should be equal to the product of the individual mgf's. In this instance, this is the case as $A'B = O$, $B'A = O$. Hence, the following result:

Theorem 3.5.1. *Assuming that X_1, \dots, X_n are iid as $X_j \sim N_p(O, I)$, let the $p \times n$ matrix $\mathbf{X} = (X_1, \dots, X_n)$ and $\bar{\mathbf{X}} = \frac{1}{n}\mathbf{X}J$, $J' = (1, 1, \dots, 1)$. Let $\bar{\mathbf{X}} = \mathbf{X}B$ and $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}A$ so that $A = A'$, $B = B'$, $A^2 = A$, $B^2 = B$, $AB = O$. Letting $U_1 = \mathbf{X}B$ and $U_2 = \mathbf{X}A$, it follows that U_1 and U_2 are independently distributed.*

Now, appealing to a general result to the effect that if U and V are independently distributed then U and VV' as well as U and $V'V$ are independently distributed whenever VV' and $V'V$ are defined, the next result follows.

Theorem 3.5.2. *For the $p \times n$ matrix \mathbf{X} , let $\mathbf{X}A$ and $\mathbf{X}B$ be as defined in Theorem 3.5.1. Then $\mathbf{X}B$ and $\mathbf{X}AA'\mathbf{X}' = \mathbf{X}A\mathbf{X}' = S$ are independently distributed and, consequently, the sample mean $\bar{\mathbf{X}}$ and the sample sum of products matrix S are independently distributed.*

As μ is absent from the previous derivations, the results hold for a $N_p(\mu, I)$ population. If the population is $N_p(\mu, \Sigma)$, $\Sigma > O$, it suffices to make the transformation $Y_j = \Sigma^{-\frac{1}{2}}X_j$ or $\mathbf{Y} = \Sigma^{-\frac{1}{2}}\mathbf{X}$, in which case $\mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{Y}$. Then, $\text{tr}(T_1'\mathbf{X}A) = \text{tr}(T_1'\Sigma^{\frac{1}{2}}\mathbf{Y}A) = \text{tr}[(T_1'\Sigma^{\frac{1}{2}})\mathbf{Y}A]$ so that $\Sigma^{\frac{1}{2}}$ is combined with T_1' , which does not affect $\mathbf{Y}A$. Thus, we have the general result that is stated next.

Theorem 3.5.3. *Letting the population be $N_p(\mu, \Sigma)$, $\Sigma > O$, and \mathbf{X} , A , B , S , and $\bar{\mathbf{X}}$ be as defined in Theorem 3.5.1, it then follows that $U_1 = \mathbf{X}A$ and $U_2 = \mathbf{X}B$ are independently distributed and thereby, that the sample mean $\bar{\mathbf{X}}$ and the sample sum of products matrix S are independently distributed.*

3.5.2. Linear functions of the sample vectors

Let the X_j 's, $j = 1, \dots, n$, be iid as $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let us consider a linear function $a_1X_1 + a_2X_2 + \dots + a_nX_n$ where a_1, \dots, a_n are real scalar constants. Then the mgf's of X_j , a_jX_j , $U = \sum_{j=1}^n a_jX_j$ are obtained as follows:

$$M_{X_j}(T) = E[e^{T'X_j}] = e^{T'\mu + \frac{1}{2}T'\Sigma T}, \quad M_{a_jX_j}(T) = e^{T'(a_j\mu) + \frac{1}{2}a_j^2T'\Sigma T}$$

$$M_{\sum_{j=1}^n a_jX_j}(T) = \prod_{j=1}^n M_{a_jX_j}(T) = e^{T'\mu(\sum_{j=1}^n a_j) + \frac{1}{2}(\sum_{j=1}^n a_j^2)T'\Sigma T},$$

which implies that $U = \sum_{j=1}^n a_jX_j$ is distributed as a real normal vector random variable with parameters $(\sum_{j=1}^n a_j)\mu$ and $(\sum_{j=1}^n a_j^2)\Sigma$, that is, $U \sim N_p(\mu(\sum_{j=1}^n a_j), (\sum_{j=1}^n a_j^2)\Sigma)$. Thus, the following result:

Theorem 3.5.4. *Let the X_j 's be iid $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, and $U = a_1X_1 + \dots + a_nX_n$ be a linear function of the X_j 's, $j = 1, \dots, n$, where a_1, \dots, a_n are real scalar constants. Then U is distributed as a p -variate real Gaussian vector random variable with parameters $[(\sum_{j=1}^n a_j)\mu, (\sum_{j=1}^n a_j^2)\Sigma]$, that is, $U \sim N_p((\sum_{j=1}^n a_j)\mu, (\sum_{j=1}^n a_j^2)\Sigma)$, $\Sigma > O$.*

If, in Theorem 3.5.4, $a_j = \frac{1}{n}$, $j = 1, \dots, n$, then $\sum_j a_j = \sum_{j=1}^n \frac{1}{n} = 1$ and $\sum_{j=1}^n a_j^2 = \sum_{j=1}^n (\frac{1}{n})^2 = \frac{1}{n}$. However, when $a_j = \frac{1}{n}$, $j = 1, \dots, n$, $U = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. Hence we have the following corollary.

Corollary 3.5.1. *Let the X_j 's be $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$. Then, the sample mean $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ is distributed as a p -variate real Gaussian with the parameters μ and $\frac{1}{n}\Sigma$, that is, $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$, $\Sigma > O$.*

From the representation given in Sect. 3.5.1, let \mathbf{X} be the sample matrix, $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, the sample average, and the $p \times n$ matrix $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$, $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}(I - \frac{1}{n}J J') = \mathbf{X}A$, $J' = (1, \dots, 1)$. Since A is idempotent of rank $n - 1$, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that $P'AP = \text{diag}(1, \dots, 1, 0) \equiv D$, $A = PDP'$ and $\mathbf{X}A = \mathbf{X}PDP' = \mathbf{Z}DP'$. Note that $A = A'$, $A^2 = A$ and $D^2 = D$. Thus, the sample sum of products matrix has the following representations:

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \mathbf{X}A A' \mathbf{X} = \mathbf{X}A \mathbf{X}' = \mathbf{Z}D D' \mathbf{Z}' = \mathbf{Z}_{n-1} \mathbf{Z}'_{n-1} \quad (3.5.3)$$

where \mathbf{Z}_{n-1} is a $p \times (n - 1)$ matrix consisting of the first $n - 1$ columns of $\mathbf{Z} = \mathbf{X}P$. When

$$D = \begin{bmatrix} I_{n-1} & O \\ O & 0 \end{bmatrix}, \quad \mathbf{Z} = (\mathbf{Z}_{n-1}, Z_{(n)}), \quad \mathbf{Z}D\mathbf{Z}' = \mathbf{Z}_{n-1}\mathbf{Z}'_{n-1},$$

where $Z_{(n)}$ denotes the last column of \mathbf{Z} . For a p -variate real normal population wherein the X_j 's are iid $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, $X_j - \bar{X} = (X_j - \mu) - (\bar{X} - \mu)$ and

hence the population can be taken to be distributed as $N_p(O, \Sigma)$, $\Sigma > O$ without any loss of generality. Then the $n - 1$ columns of \mathbf{Z}_{n-1} will be iid standard normal $N_p(O, I)$. After discussing the real matrix-variate gamma distribution in the next chapter, we will show that whenever $(n - 1) \geq p$, $\mathbf{Z}_{n-1}\mathbf{Z}'_{n-1}$ has a real matrix-variate gamma distribution, or equivalently, that it is Wishart distributed with $n - 1$ degrees of freedom.

3.5a.1. Some simplifications of the sample matrix in the complex Gaussian case

Let the $p \times 1$ vector \tilde{X}_1 in the complex domain have a complex Gaussian density $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. Let $\tilde{X}_1, \dots, \tilde{X}_n$ be iid as $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$ or $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_n]$ is the sample matrix of a simple random sample of size n from a $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. Let the sample mean vector or the sample average be $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$ and the matrix of sample means be the bold-faced $p \times n$ matrix $\tilde{\tilde{\mathbf{X}}}$. Let $\tilde{S} = (\tilde{\mathbf{X}} - \tilde{\tilde{\mathbf{X}}})(\tilde{\mathbf{X}} - \tilde{\tilde{\mathbf{X}}})^*$. Then $\tilde{\tilde{\mathbf{X}}} = \frac{1}{n}\tilde{\mathbf{X}}\mathbf{J}\mathbf{J}' = \tilde{\tilde{\mathbf{X}}}\mathbf{B}$, $\tilde{\mathbf{X}} - \tilde{\tilde{\mathbf{X}}} = \tilde{\mathbf{X}}(I - \frac{1}{n}\mathbf{J}\mathbf{J}') = \tilde{\mathbf{X}}\mathbf{A}$. Then, $\mathbf{A} = \mathbf{A}^2$, $\mathbf{A} = \mathbf{A}' = \mathbf{A}^*$, $\mathbf{B} = \mathbf{B}' = \mathbf{B}^*$, $\mathbf{B} = \mathbf{B}^2$, $\mathbf{A}\mathbf{B} = \mathbf{O}$, $\mathbf{B}\mathbf{A} = \mathbf{O}$. Thus, results parallel to Theorems 3.5.1 and 3.5.2 hold in the complex domain, and we now state the general result.

Theorem 3.5a.1. *Let the population be complex p -variate Gaussian $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. Let the $p \times n$ sample matrix be $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$ where $\tilde{X}_1, \dots, \tilde{X}_n$ are iid as $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. Let $\tilde{\mathbf{X}}$, $\bar{\tilde{X}}$, \tilde{S} , $\tilde{\tilde{\mathbf{X}}}\mathbf{A}$, $\tilde{\tilde{\mathbf{X}}}\mathbf{B}$ be as defined above. Then, $\tilde{\tilde{\mathbf{X}}}\mathbf{A}$ and $\tilde{\tilde{\mathbf{X}}}\mathbf{B}$ are independently distributed, and thereby the sample mean $\bar{\tilde{X}}$ and the sample sum of products matrix \tilde{S} are independently distributed.*

3.5a.2. Linear functions of the sample vectors in the complex domain

Let $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$ be a p -variate complex Gaussian vector random variable. Consider a simple random sample of size n from this population, in which case the \tilde{X}_j 's, $j = 1, \dots, n$, are iid as $\tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} > O$. Let the linear function $\tilde{U} = a_1\tilde{X}_1 + \dots + a_n\tilde{X}_n$ where a_1, \dots, a_n are real or complex scalar constants. Then, following through steps parallel to those provided in Sect. 3.5.2, we obtain the following mgf:

$$\tilde{M}_{\tilde{U}}(\tilde{T}) = e^{\Re(\tilde{T}^*\tilde{\mu}(\sum_{j=1}^n a_j)) + \frac{1}{4}(\sum_{j=1}^n a_j a_j^*)\tilde{T}^*\tilde{\Sigma}\tilde{T}}$$

where $\sum_{j=1}^n a_j a_j^* = |\tilde{a}_1|^2 + \dots + |\tilde{a}_n|^2$. For example, if $a_j = \frac{1}{n}$, $j = 1, \dots, n$, then $\sum_{j=1}^n a_j = 1$ and $\sum_{j=1}^n a_j a_j^* = \frac{1}{n}$. Hence, we have the following result and the resulting corollary.

Theorem 3.5a.2. Let the $p \times 1$ complex vector have a p -variate complex Gaussian distribution $\tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$. Consider a simple random sample of size n from this population, with the \tilde{X}_j 's, $j = 1, \dots, n$, being iid as this p -variate complex Gaussian. Let a_1, \dots, a_n be scalar constants, real or complex. Consider the linear function $\tilde{U} = a_1\tilde{X}_1 + \dots + a_n\tilde{X}_n$. Then $\tilde{U} \sim \tilde{N}_p(\tilde{\mu}(\sum_{j=1}^n a_j), (\sum_{j=1}^n a_j a_j^*)\tilde{\Sigma})$, that is, \tilde{U} has a p -variate complex Gaussian distribution with the parameters $(\sum_{j=1}^n a_j)\tilde{\mu}$ and $(\sum_{j=1}^n a_j a_j^*)\tilde{\Sigma}$.

Corollary 3.5a.1. Let the population and sample be as defined in Theorem 3.5a.2. Then the sample mean $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$ is distributed as a p -variate complex Gaussian with the parameters $\tilde{\mu}$ and $\frac{1}{n}\tilde{\Sigma}$.

Proceeding as in the real case, we can show that the sample sum of products matrix \tilde{S} can have a representation of the form

$$\tilde{S} = \tilde{\mathbf{Z}}_{n-1}\tilde{\mathbf{Z}}_{n-1}^* \quad (3.5a.3)$$

where the columns of $\tilde{\mathbf{Z}}_{n-1}$ are iid standard normal vectors in the complex domain if the population is a p -variate Gaussian in the complex domain. In this case, it will be shown later, that \tilde{S} is distributed as a complex Wishart matrix with $(n-1) \geq p$ degrees of freedom.

3.5.3. Maximum likelihood estimators of the p -variate real Gaussian distribution

Letting L denote the joint density of the sample values X_1, \dots, X_n , which are $p \times 1$ iid Gaussian vectors constituting a simple random sample of size n , we have

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{1}{2} n (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \quad (3.5.4)$$

where, as previously denoted, \mathbf{X} is the $p \times n$ matrix

$$\mathbf{X} = (X_1, \dots, X_n), \quad \bar{X} = \frac{1}{n}(X_1 + \dots + X_n), \quad \bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X}),$$

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = (s_{ij}), \quad s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

In this case, the parameters are the $p \times 1$ vector μ and the $p \times p$ real positive definite matrix Σ . If we resort to Calculus to maximize L , then we would like to differentiate L , or

the one-to-one function $\ln L$, with respect to μ and Σ directly, rather than differentiating with respect to each element comprising μ and Σ . For achieving this, we need to further develop the differential operators introduced in Chap. 1.

Definition 3.5.1. *Derivative with respect to a matrix.* Let $Y = (y_{ij})$ be a $p \times q$ matrix where the elements y_{ij} 's are distinct real scalar variables. The operator $\frac{\partial}{\partial Y}$ will be defined as $\frac{\partial}{\partial Y} = \left(\frac{\partial}{\partial y_{ij}} \right)$ and this operator applied to a real scalar quantity f will be defined as

$$\frac{\partial}{\partial Y} f = \left(\frac{\partial f}{\partial y_{ij}} \right).$$

For example, if $f = y_{11}^2 + y_{12}^2 + y_{13}^2 - y_{11}y_{12} + y_{21} + y_{22}^2 + y_{23}$ and the 2×3 matrix Y is

$$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \end{bmatrix} \Rightarrow \frac{\partial f}{\partial Y} = \begin{bmatrix} \frac{\partial f}{\partial y_{11}} & \frac{\partial f}{\partial y_{12}} & \frac{\partial f}{\partial y_{13}} \\ \frac{\partial f}{\partial y_{21}} & \frac{\partial f}{\partial y_{22}} & \frac{\partial f}{\partial y_{23}} \end{bmatrix},$$

$$\frac{\partial f}{\partial Y} = \begin{bmatrix} 2y_{11} - y_{12} & 2y_{12} - y_{11} & 2y_{13} \\ 1 & 2y_{22} & 1 \end{bmatrix}.$$

There are numerous examples of real-valued scalar functions of matrix argument. The determinant and the trace are two scalar functions of a square matrix A . The derivative with respect to a vector has already been defined in Chap. 1. The loglikelihood function $\ln L$ which is available from (3.5.4) has to be differentiated with respect to μ and with respect to Σ and the resulting expressions have to be respectively equated to a null vector and a null matrix. These equations are then solved to obtain the critical points where the L as well as $\ln L$ may have a local maximum, a local minimum or a saddle point. However, $\ln L$ contains a determinant and a trace. Hence we need to develop some results on differentiating a determinant and a trace with respect to a matrix, and the following results will be helpful in this regard.

Theorem 3.5.5. *Let the $p \times p$ matrix $Y = (y_{ij})$ be nonsingular, the y_{ij} 's being distinct real scalar variables. Let $f = |Y|$, the determinant of Y . Then,*

$$\frac{\partial}{\partial Y} |Y| = \begin{cases} |Y|(Y^{-1})' & \text{for a general } Y \\ |Y|[2Y^{-1} - \text{diag}(Y^{-1})] & \text{for } Y = Y' \end{cases}$$

where $\text{diag}(Y^{-1})$ is a diagonal matrix whose diagonal elements coincide with those of Y^{-1} .

Proof: A determinant can be obtained by expansions along any row (or column), the resulting sums involving the corresponding elements and their associated cofactors. More specifically, $|Y| = y_{i1}C_{i1} + \dots + y_{ip}C_{ip}$ for each $i = 1, \dots, p$, where C_{ij} is the cofactor of y_{ij} . This expansion holds whether the elements in the matrix are real or complex. Then,

$$\frac{\partial}{\partial y_{ij}}|Y| = \begin{cases} C_{ij} & \text{for a general } Y \\ 2C_{ij} & \text{for } Y = Y', i \neq j \\ C_{jj} & \text{for } Y = Y', i = j. \end{cases}$$

Thus, $\frac{\partial}{\partial Y}|Y| =$ the matrix of cofactors $= |Y|(Y^{-1})'$ for a general Y . When $Y = Y'$, then

$$\begin{aligned} \frac{\partial}{\partial Y}|Y| &= \begin{bmatrix} C_{11} & 2C_{12} & \dots & 2C_{1p} \\ 2C_{21} & C_{22} & \dots & 2C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 2C_{p1} & 2C_{p2} & \dots & C_{pp} \end{bmatrix} \\ &= |Y|[2Y^{-1} - \text{diag}(Y^{-1})]. \end{aligned}$$

Hence the result.

Theorem 3.5.6. Let A and $Y = (y_{ij})$ be $p \times p$ matrices where A is a constant matrix and the y_{ij} 's are distinct real scalar variables. Then,

$$\frac{\partial}{\partial Y}[\text{tr}(AY)] = \begin{cases} A' & \text{for a general } Y \\ A + A' - \text{diag}(A) & \text{for } Y = Y'. \end{cases}$$

Proof: $\text{tr}(AY) = \sum_{ij} a_{ji}y_{ij}$ for a general Y , so that $\frac{\partial}{\partial Y}[\text{tr}(AY)] = A'$ for a general Y . When $Y = Y'$, $\frac{\partial}{\partial y_{jj}}[\text{tr}(AY)] = a_{jj}$ and $\frac{\partial}{\partial y_{ij}}[\text{tr}(AY)] = a_{ij} + a_{ji}$ for $i \neq j$. Hence, $\frac{\partial}{\partial Y}[\text{tr}(AY)] = A + A' - \text{diag}(A)$ for $Y = Y'$. Thus, the result is established.

With the help of Theorems 3.5.5 and 3.5.6, we can optimize L or $\ln L$ with L as specified in Eq. (3.5.4). For convenience, we take $\ln L$ which is given by

$$\ln L = -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1}S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu). \quad (3.5.5)$$

Then,

$$\begin{aligned} \frac{\partial}{\partial \mu} \ln L = 0 &\Rightarrow 0 - \frac{n}{2} \frac{\partial}{\partial \mu} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) = 0 \\ &\Rightarrow n \Sigma^{-1} (\bar{X} - \mu) = 0 \Rightarrow \bar{X} - \mu = 0 \\ &\Rightarrow \mu = \bar{X}, \end{aligned}$$

referring to the vector derivatives defined in Chap. 1. The extremal value, denoted with a hat, is $\hat{\mu} = \bar{X}$. When differentiating with respect to Σ , we may take $B = \Sigma^{-1}$ for convenience and differentiate with respect to B . We may also substitute \bar{X} to μ because the critical point for Σ must correspond to $\hat{\mu} = \bar{X}$. Accordingly, $\ln L$ at $\mu = \bar{X}$ is

$$\ln L(\hat{\mu}, B) = -\frac{np}{2} \ln(2\pi) + \frac{n}{2} \ln |B| - \frac{1}{2} \text{tr}(BS).$$

Noting that $B = B'$,

$$\begin{aligned} \frac{\partial}{\partial B} \ln L(\hat{\mu}, B) = O &\Rightarrow \frac{n}{2} [2B^{-1} - \text{diag}(B^{-1})] - \frac{1}{2} [2S - \text{diag}(S)] = O \\ &\Rightarrow n[2\Sigma - \text{diag}(\Sigma)] = 2S - \text{diag}(S) \\ &\Rightarrow \hat{\sigma}_{jj} = \frac{1}{n} s_{jj}, \hat{\sigma}_{ij} = \frac{1}{n} s_{ij}, i \neq j \\ &\Rightarrow (\hat{\mu} = \bar{X}, \hat{\Sigma} = \frac{1}{n} S). \end{aligned}$$

Hence, the only critical point is $(\hat{\mu}, \hat{\Sigma}) = (\bar{X}, \frac{1}{n} S)$. Does this critical point correspond to a local maximum or a local minimum or something else? For $\hat{\mu} = \bar{X}$, consider the behavior of $\ln L$. For convenience, we may convert the problem in terms of the eigenvalues of B . Letting $\lambda_1, \dots, \lambda_p$ be the eigenvalues of B , observe that $\lambda_j > 0$, $j = 1, \dots, p$, that the determinant is the product of the eigenvalues and the trace is the sum of the eigenvalues. Examining the behavior of $\ln L$ for all possible values of λ_1 when $\lambda_2, \dots, \lambda_p$ are fixed, we see that $\ln L$ at $\hat{\mu}$ goes from $-\infty$ to $-\infty$ through finite values. For each λ_j , the behavior of $\ln L$ is the same. Hence the only critical point must correspond to a local maximum. Therefore $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \frac{1}{n} S$ are the maximum likelihood estimators (MLE's) of μ and Σ respectively. The observed values of $\hat{\mu}$ and $\hat{\Sigma}$ are the maximum likelihood estimates of μ and Σ , for which the same abbreviation MLE is utilized. While maximum likelihood estimators are random variables, maximum likelihood estimates are numerical values. Observe that, in order to have an estimate for Σ , we must have that the sample size $n \geq p$.

In the derivation of the MLE of Σ , we have differentiated with respect to $B = \Sigma^{-1}$ instead of differentiating with respect to the parameter Σ . Could this affect final result? Given any θ and any non-trivial differentiable function of θ , $\phi(\theta)$, whose derivative is not identically zero, that is, $\frac{d}{d\theta} \phi(\theta) \neq 0$ for any θ , it follows from basic calculus that for any differentiable function $g(\theta)$, the equations $\frac{d}{d\theta} g(\theta) = 0$ and $\frac{d}{d\phi} g(\theta) = 0$ will lead to the same solution for θ . Hence, whether we differentiate with respect to $B = \Sigma^{-1}$ or Σ , the procedures will lead to the same estimator of Σ . As well, if $\hat{\theta}$ is the MLE of θ ,

then $g(\hat{\theta})$ will also be the MLE of $g(\theta)$ whenever $g(\theta)$ is a one-to-one function of θ . The numerical evaluation of maximum likelihood estimates for μ and Σ has been illustrated in Example 3.5.1.

3.5a.3. MLE's in the complex p -variate Gaussian case

Let the $p \times 1$ vectors in the complex domain $\tilde{X}_1, \dots, \tilde{X}_n$ be iid as $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. and let the joint density of the \tilde{X}_j 's, $j = 1, \dots, n$, be denoted by \tilde{L} . Then

$$\begin{aligned} \tilde{L} &= \prod_{j=1}^n \frac{e^{-(\tilde{X}_j - \tilde{\mu})^* \Sigma^{-1} (\tilde{X}_j - \tilde{\mu})}}{\pi^p |\det(\Sigma)|} = \frac{e^{-\sum_{j=1}^n (\tilde{X}_j - \tilde{\mu})^* \Sigma^{-1} (\tilde{X}_j - \tilde{\mu})}}{\pi^{np} |\det(\Sigma)|^n} \\ &= \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{S}) - n(\tilde{\bar{X}} - \tilde{\mu})^* \Sigma^{-1} (\tilde{\bar{X}} - \tilde{\mu})}}{\pi^{np} |\det(\Sigma)|^n} \end{aligned}$$

where $|\det(\Sigma)|$ denotes the absolute value of the determinant of Σ ,

$$\begin{aligned} \tilde{S} &= (\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})(\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})^* = (\tilde{s}_{ij}), \quad \tilde{s}_{ij} = \sum_{k=1}^n (\tilde{x}_{ik} - \tilde{\bar{x}}_i)(\tilde{x}_{jk} - \tilde{\bar{x}}_j)^*, \\ \tilde{\mathbf{X}} &= [\tilde{X}_1, \dots, \tilde{X}_n], \quad \tilde{\bar{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n), \quad \tilde{\bar{\mathbf{X}}} = [\tilde{\bar{X}}, \dots, \tilde{\bar{X}}], \end{aligned}$$

where $\tilde{\mathbf{X}}$ and $\tilde{\bar{\mathbf{X}}}$ are $p \times n$. Hence,

$$\ln \tilde{L} = -np \ln \pi - n \ln |\det(\Sigma)| - \text{tr}(\Sigma^{-1} \tilde{S}) - n(\tilde{\bar{X}} - \tilde{\mu})^* \Sigma^{-1} (\tilde{\bar{X}} - \tilde{\mu}). \quad (3.5a.4)$$

3.5a.4. Matrix derivatives in the complex domain

Consider $\text{tr}(\tilde{B} \tilde{S}^*)$, $\tilde{B} = \tilde{B}^* > O$, $\tilde{S} = \tilde{S}^* > O$. Let $\tilde{B} = B_1 + iB_2$, $\tilde{S} = S_1 + iS_2$, $i = \sqrt{-1}$. Then B_1 and S_1 are real symmetric and B_2 and S_2 are real skew symmetric since \tilde{B} and \tilde{S} are Hermitian. What is then $\frac{\partial}{\partial \tilde{B}} [\text{tr}(\tilde{B} \tilde{S}^*)]$? Consider

$$\begin{aligned} \tilde{B} \tilde{S}^* &= (B_1 + iB_2)(S_1' - iS_2') = B_1 S_1' + B_2 S_2' + i(B_2 S_1' - B_1 S_2'), \\ \text{tr}(\tilde{B} \tilde{S}^*) &= \text{tr}(B_1 S_1' + B_2 S_2') + i[\text{tr}(B_2 S_1') - \text{tr}(B_1 S_2')]. \end{aligned}$$

It can be shown that when B_2 and S_2 are real skew symmetric and B_1 and S_1 are real symmetric, then $\text{tr}(B_2 S_1') = 0$, $\text{tr}(B_1 S_2') = 0$. This will be stated as a lemma.

Lemma 3.5a.1. Consider two $p \times p$ real matrices A and B where $A = A'$ (symmetric) and $B = -B'$ (skew symmetric). Then, $\text{tr}(AB) = 0$.

Proof: $\text{tr}(AB) = \text{tr}(AB)' = \text{tr}(B'A') = -\text{tr}(BA) = -\text{tr}(AB)$, which implies that $\text{tr}(AB) = 0$.

Thus, $\text{tr}(\tilde{B}\tilde{S}^*) = \text{tr}(B_1S_1' + B_2S_2')$. The diagonal elements of S_1 in $\text{tr}(B_1S_1')$ are multiplied once by the diagonal elements of B_1 and the non-diagonal elements in S_1 are multiplied twice each by the corresponding elements in B_1 . Hence,

$$\frac{\partial}{\partial B_1} \text{tr}(B_1S_1') = 2S_1 - \text{diag}(S_1).$$

In B_2 and S_2 , the diagonal elements are zeros and hence

$$\frac{\partial}{\partial B_2} \text{tr}(B_2S_2') = 2S_2.$$

Therefore

$$\left(\frac{\partial}{\partial B_1} + i \frac{\partial}{\partial B_2} \right) \text{tr}(B_1S_1' + B_2S_2') = 2(S_1 + iS_2) - \text{diag}(S_1) = 2\tilde{S} - \text{diag}(\tilde{S}).$$

Thus, the following result:

Theorem 3.5a.3. Let $\tilde{S} = \tilde{S}^* > O$ and $\tilde{B} = \tilde{B}^* > O$ be $p \times p$ Hermitian matrices. Let $\tilde{B} = B_1 + iB_2$ and $\tilde{S} = S_1 + iS_2$ where the $p \times p$ matrices B_1 and S_1 are symmetric and B_2 and S_2 are skew symmetric real matrices. Letting $\frac{\partial}{\partial \tilde{B}} = \frac{\partial}{\partial B_1} + i \frac{\partial}{\partial B_2}$, we have

$$\frac{\partial}{\partial \tilde{B}} \text{tr}(\tilde{B}\tilde{S}^*) = 2\tilde{S} - \text{diag}(\tilde{S}).$$

Theorem 3.5a.4. Let $\tilde{\Sigma} = (\tilde{\sigma}_{ij}) = \tilde{\Sigma}^* > O$ be a Hermitian positive definite $p \times p$ matrix. Let $\det(\Sigma)$ be the determinant and $|\det(\Sigma)|$ be the absolute value of the determinant respectively. Let $\frac{\partial}{\partial \tilde{\Sigma}} = \frac{\partial}{\partial \Sigma_1} + i \frac{\partial}{\partial \Sigma_2}$ be the differential operator, where $\tilde{\Sigma} = \Sigma_1 + i\Sigma_2$, $i = \sqrt{-1}$, Σ_1 being real symmetric and Σ_2 , real skew symmetric. Then,

$$\frac{\partial}{\partial \tilde{\Sigma}} \ln |\det(\tilde{\Sigma})| = 2\tilde{\Sigma}^{-1} - \text{diag}(\tilde{\Sigma}^{-1}).$$

Proof: Note that for two scalar complex quantities, $\tilde{x} = x_1 + ix_2$ and $\tilde{y} = y_1 + iy_2$ where $i = \sqrt{-1}$ and x_1, x_2, y_1, y_2 are real, and for the operator $\frac{\partial}{\partial \tilde{x}} = \frac{\partial}{\partial x_1} + i \frac{\partial}{\partial x_2}$, the following results hold, which will be stated as a lemma.

Lemma 3.5a.2. Given \tilde{x}, \tilde{y} and the operator $\frac{\partial}{\partial x_1} + i \frac{\partial}{\partial x_2}$ defined above,

$$\begin{aligned} \frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{y}) &= 0, \quad \frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{y}^*) = 0, \quad \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{y}) = 2\tilde{y}, \quad \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{y}^*) = 2\tilde{y}^*, \\ \frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{x}^*) &= \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{x}) = 2\tilde{x}, \quad \frac{\partial}{\partial \tilde{x}^*}(\tilde{x}^*\tilde{x}) = \frac{\partial}{\partial \tilde{x}^*}(\tilde{x}\tilde{x}^*) = 2\tilde{x}^* \end{aligned}$$

where, for example, \tilde{x}^* which, in general, is the conjugate transpose of \tilde{x} , is only the conjugate in this case since \tilde{x} is a scalar quantity.

Observe that for a $p \times p$ Hermitian positive definite matrix \tilde{X} , the absolute value of the determinant, namely, $|\det(\tilde{X})| = \sqrt{\det(\tilde{X})\det(\tilde{X}^*)} = \det(\tilde{X}) = \det(\tilde{X}^*)$ since $\tilde{X} = \tilde{X}^*$. Consider the following cofactor expansion of $\det(\tilde{X})$ (in general, a cofactor expansion is valid whether the elements of the matrix are real or complex). Letting C_{ij} denote the cofactor of x_{ij} in $\tilde{X} = (x_{ij})$ when x_{ij} is real or complex,

$$\det(X) = x_{11}C_{11} + x_{12}C_{12} + \cdots + x_{1p}C_{1p} \tag{1}$$

$$= x_{21}C_{21} + x_{22}C_{22} + \cdots + x_{2p}C_{2p} \tag{2}$$

⋮

$$= x_{p1}C_{p1} + x_{p2}C_{p2} + \cdots + x_{pp}C_{pp}. \tag{p}$$

When $\tilde{X} = \tilde{X}^*$, the diagonal elements x_{jj} 's are all real. From Lemma 3.5a.2 and equation (1), we have

$$\frac{\partial}{\partial x_{11}}(x_{11}C_{11}) = C_{11}, \quad \frac{\partial}{\partial x_{1j}}(x_{1j}C_{1j}) = 0, \quad j = 2, \dots, p.$$

From Eq. (2), note that $x_{21} = x_{12}^*$, $C_{21} = C_{12}^*$ since $\tilde{X} = \tilde{X}^*$. Then from Lemma 3.5a.2 and (2), we have

$$\frac{\partial}{\partial x_{12}}(x_{12}^*C_{12}^*) = C_{12}^*, \quad \frac{\partial}{\partial x_{22}}(x_{22}^*C_{22}^*) = C_{22}^*, \quad \frac{\partial}{\partial x_{2j}}(x_{2j}C_{2j}) = 0, \quad j = 3, \dots, p,$$

observing that $x_{22}^* = x_{22}$ and $C_{22}^* = C_{22}$. Now, continuing the process with Eqs. (3), (4), \dots , (p), we have the following result:

$$\frac{\partial}{\partial x_{ij}}[\det(\tilde{X})] = \begin{cases} C_{jj}^*, & j = 1, \dots, p \\ 2C_{ij}^* & \text{for all } i \neq j. \end{cases}$$

Observe that for $\tilde{\Sigma}^{-1} = \tilde{B} = \tilde{B}^*$,

$$\frac{\partial}{\partial \tilde{B}} [\ln(\det(\tilde{B}))] = \frac{1}{\det(\tilde{B})} \begin{bmatrix} B_{11}^* & 2B_{12}^* & \cdots & B_{1p}^* \\ 2B_{21}^* & B_{22}^* & \cdots & 2B_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ 2B_{p1}^* & 2B_{p2}^* & \cdots & B_{pp}^* \end{bmatrix}$$

$$= 2\tilde{B}^{-1} - \text{diag}(\tilde{B}^{-1}) = 2\tilde{\Sigma} - \text{diag}(\tilde{\Sigma})$$

where B_{rs} is the cofactor of \tilde{b}_{rs} , $\tilde{B} = (\tilde{b}_{rs})$. Therefore, at $\hat{\mu} = \bar{X}$, for $\tilde{\Sigma}^{-1} = \tilde{B}$, and from Theorems 3.5a.5 and 3.5a.6, we have

$$\begin{aligned} \frac{\partial}{\partial \tilde{B}} [\ln \tilde{L}] = O &\Rightarrow n[\tilde{\Sigma} - \text{diag}(\tilde{\Sigma})] - [\tilde{S} - \text{diag}(\tilde{S})] = O \\ &\Rightarrow \tilde{\Sigma} = \frac{1}{n}\tilde{S} \Rightarrow \hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S} \text{ for } n \geq p, \end{aligned}$$

where a hat denotes the estimate/estimator.

Again, from Lemma 3.5a.2 we have the following:

$$\begin{aligned} \frac{\partial}{\partial \tilde{\mu}} [(\bar{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\bar{X} - \tilde{\mu})] &= \frac{\partial}{\partial \tilde{\mu}} \{ \bar{X}^* \tilde{\Sigma}^{-1} \bar{X} + \tilde{\mu}^* \tilde{\Sigma}^{-1} \tilde{\mu} - \bar{X}^* \tilde{\Sigma}^{-1} \tilde{\mu} + \tilde{\mu}^* \tilde{\Sigma}^{-1} \bar{X} \} = O \\ &\Rightarrow O + 2\tilde{\Sigma}^{-1} \tilde{\mu} - O - 2\tilde{\Sigma}^{-1} \bar{X} = O \\ &\Rightarrow \hat{\tilde{\mu}} = \bar{X}. \end{aligned}$$

Thus, the MLE of $\tilde{\mu}$ and $\tilde{\Sigma}$ are respectively $\hat{\tilde{\mu}} = \bar{X}$ and $\hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S}$ for $n \geq p$. It is not difficult to show that the only critical point $(\hat{\tilde{\mu}}, \hat{\tilde{\Sigma}}) = (\bar{X}, \frac{1}{n}\tilde{S})$ corresponds to a local maximum for \tilde{L} . Consider $\ln \tilde{L}$ at $\hat{\tilde{\mu}} = \bar{X}$. Let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of $\tilde{B} = \tilde{\Sigma}^{-1}$ where the λ_j 's are real as \tilde{B} is Hermitian. Examine the behavior of $\ln \tilde{L}$ when a λ_j is increasing from 0 to ∞ . Then $\ln \tilde{L}$ goes from $-\infty$ back to $-\infty$ through finite values. Hence, the only critical point corresponds to a local maximum. Thus, \bar{X} and $\frac{1}{n}\tilde{S}$ are the MLE's of $\tilde{\mu}$ and $\tilde{\Sigma}$, respectively.

Theorems 3.5.7, 3.5a.5. For the p -variate real Gaussian with the parameters μ and $\Sigma > O$ and the p -variate complex Gaussian with the parameters $\tilde{\mu}$ and $\tilde{\Sigma} > O$, the maximum likelihood estimators (MLE's) are $\hat{\mu} = \bar{X}$, $\hat{\Sigma} = \frac{1}{n}S$, $\hat{\tilde{\mu}} = \bar{X}$, $\hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S}$ where n is the sample size, \bar{X} and S are the sample mean and sample sum of products matrix in

the real case, and $\bar{\tilde{X}}$ and \tilde{S} are the sample mean and the sample sum of products matrix in the complex case, respectively.

A numerical illustration of the maximum likelihood estimates of $\tilde{\mu}$ and $\tilde{\Sigma}$ in the complex domain has already been given in Example 3.5a.1.

It can be shown that the MLE of μ and Σ in the real and complex p -variate Gaussian cases are such that $E[\bar{X}] = \mu$, $E[\tilde{X}] = \tilde{\mu}$, $E[S] = \frac{n-1}{n}\Sigma$, $E[\tilde{S}] = \frac{n-1}{n}\tilde{\Sigma}$. For these results to hold, the population need not be Gaussian. Any population for which the covariance matrix exists will have these properties. This will be stated as a result.

Theorems 3.5.8, 3.5a.6. *Let X_1, \dots, X_n be a simple random sample from any p -variate population with mean value vector μ and covariance matrix $\Sigma = \Sigma' > O$ in the real case and mean value vector $\tilde{\mu}$ and covariance matrix $\tilde{\Sigma} = \tilde{\Sigma}^* > O$ in the complex case, respectively, and let Σ and $\tilde{\Sigma}$ exist in the sense all the elements therein exist. Let $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$ and let S and \tilde{S} be the sample sum of products matrices in the real and complex cases, respectively. Then $E[\bar{X}] = \mu$, $E[\bar{\tilde{X}}] = \tilde{\mu}$, $E[\hat{\Sigma}] = E[\frac{1}{n}S] = \frac{n-1}{n}\Sigma \rightarrow \Sigma$ as $n \rightarrow \infty$ and $E[\hat{\tilde{\Sigma}}] = E[\frac{1}{n}\tilde{S}] = \frac{n-1}{n}\tilde{\Sigma} \rightarrow \tilde{\Sigma}$ as $n \rightarrow \infty$.*

Proof: $E[\bar{X}] = \frac{1}{n}\{E[X_1] + \dots + E[X_n]\} = \frac{1}{n}\{\mu + \dots + \mu\} = \mu$. Similarly, $E[\bar{\tilde{X}}] = \tilde{\mu}$. Let $\mathbf{M} = (\mu, \mu, \dots, \mu)$, that is, \mathbf{M} is a $p \times n$ matrix wherein every column is the $p \times 1$ vector μ . Let $\tilde{\mathbf{X}} = (\tilde{X}, \dots, \tilde{X})$, that is, $\tilde{\mathbf{X}}$ is a $p \times n$ matrix wherein every column is \tilde{X} . Now, consider

$$E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'] = E\left[\sum_{j=1}^n (X_j - \mu)(X_j - \mu)'\right] = \sum_{j=1}^n \{\Sigma + \dots + \Sigma\} = n\Sigma.$$

As well,

$$\begin{aligned} (\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})' &= (\mathbf{X} - \bar{\tilde{X}} + \bar{\tilde{X}} - \mathbf{M})(\mathbf{X} - \bar{\tilde{X}} + \bar{\tilde{X}} - \mathbf{M})' \\ &= (\mathbf{X} - \bar{\tilde{X}})(\mathbf{X} - \bar{\tilde{X}})' + (\mathbf{X} - \bar{\tilde{X}})(\bar{\tilde{X}} - \mathbf{M})' \\ &\quad + (\bar{\tilde{X}} - \mathbf{M})(\mathbf{X} - \bar{\tilde{X}})' + (\bar{\tilde{X}} - \mathbf{M})(\bar{\tilde{X}} - \mathbf{M})' \Rightarrow \\ (\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})' &= S + \sum_{j=1}^n (X_j - \bar{\tilde{X}})(\bar{\tilde{X}} - \mu)' + \sum_{j=1}^n (\bar{\tilde{X}} - \mu)(X_j - \bar{\tilde{X}})' \\ &\quad + \sum_{j=1}^n (\bar{\tilde{X}} - \mu)(\bar{\tilde{X}} - \mu)' \\ &= S + O + O + n(\bar{\tilde{X}} - \mu)(\bar{\tilde{X}} - \mu)' \Rightarrow \end{aligned}$$

$$n\Sigma = E[S] + O + O + n\text{Cov}(\bar{X}) = E[S] + n\left[\frac{1}{n}\Sigma\right] = E[S] + \Sigma \Rightarrow$$

$$E[S] = (n-1)\Sigma \Rightarrow E[\hat{S}] = E\left[\frac{1}{n}S\right] = \frac{n-1}{n}\Sigma \rightarrow \Sigma \text{ as } n \rightarrow \infty.$$

Observe that $\sum_{j=1}^n (X_j - \bar{X}) = O$, this result having been utilized twice in the above derivations. The complex case can be established in a similar manner. This completes the proof.

3.5.4. Properties of maximum likelihood estimators

Definition 3.5.2 Unbiasedness. Let $g(\theta)$ be a function of the parameter θ which stands for all the parameters associated with a population's distribution. Let the independently distributed random variables x_1, \dots, x_n constitute a simple random sample of size n from a univariate population. Let $T(x_1, \dots, x_n)$ be an observable function of the sample values x_1, \dots, x_n . This definition for a statistic holds when the iid variables are scalar, vector or matrix variables, whether in the real or complex domains. Then T is called *a statistic* (the plural form, statistics, is not to be confused with the subject of Statistics). If $E[T] = g(\theta)$ for all θ in the parameter space, then T is said to be unbiased for $g(\theta)$ or an unbiased estimator of $g(\theta)$.

We will look at some properties of the MLE of the parameter or parameters represented by θ in a given population specified by its density/probability function $f(x, \theta)$. Consider a simple random sample of size n from this population. The sample will be of the form x_1, \dots, x_n if the population is univariate or of the form X_1, \dots, X_n if the population is multivariate or matrix-variate. Some properties of estimators in the scalar variable case will be illustrated first. Then the properties will be extended to the vector/matrix-variate cases. The joint density of the sample values will be denoted by L . Thus, in the univariate case,

$$L = L(x_1, \dots, x_n, \theta) = \prod_{j=1}^n f(x_j, \theta) \Rightarrow \ln L = \sum_{j=1}^n \ln f(x_j, \theta).$$

Since the total probability is 1, we have the following, taking for example the variable to be continuous and a scalar parameter θ :

$$\int_X L \, dX = 1 \Rightarrow \frac{\partial}{\partial \theta} \int_X L \, dX = 0, \quad X' = (x_1, \dots, x_n).$$

We are going to assume that the support of x is free of θ and the differentiation can be done inside the integral sign. Then,

$$0 = \int_X \frac{\partial}{\partial \theta} L \, dX = \int_X \frac{1}{L} \left(\frac{\partial}{\partial \theta} L \right) L \, dX = \int_X \left[\frac{\partial}{\partial \theta} \ln L \right] L \, dX.$$

Noting that $\int_X (\cdot) L \, dX = E[(\cdot)]$, we have

$$E \left[\frac{\partial}{\partial \theta} \ln L \right] = 0 \Rightarrow E \left[\sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta) \right] = 0. \quad (3.5.6)$$

Let $\hat{\theta}$ be the MLE of θ . Then

$$\begin{aligned} \frac{\partial}{\partial \theta} L|_{\theta=\hat{\theta}} = 0 &\Rightarrow \frac{\partial}{\partial \theta} \ln L|_{\theta=\hat{\theta}} = 0 \\ &\Rightarrow E \left[\sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\hat{\theta}} \right] = 0. \end{aligned} \quad (3.5.7)$$

If θ is scalar, then the above are single equations, otherwise they represent a system of equations as the derivatives are then vector or matrix derivatives. Here (3.5.6) is the likelihood equation giving rise to the maximum likelihood estimators (MLE) of θ . However, by the weak law of large numbers (see Sect. 2.6),

$$\frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\hat{\theta}} \rightarrow E \left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta_o) \right] \text{ as } n \rightarrow \infty \quad (3.5.8)$$

where θ_o is the true value of θ . Noting that $E \left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta_o) \right] = 0$ owing to the fact that $\int_{-\infty}^{\infty} f(x) dx = 1$, we have the following results:

$$\sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\hat{\theta}} = 0, \quad E \left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\theta_o} \right] = 0.$$

This means that $E[\hat{\theta}] = \theta_o$ or $E[\hat{\theta}] \rightarrow \theta_o$ as $n \rightarrow \infty$, that is, $\hat{\theta}$ is asymptotically unbiased for the true value θ_o of θ . As well, $\hat{\theta} \rightarrow \theta_o$ as $n \rightarrow \infty$ almost surely or with probability 1, except on a set having probability measure zero. Thus, the MLE of θ is asymptotically unbiased and consistent for the true value θ_o , which is stated next as a theorem:

Theorem 3.5.9. *In a given population's distribution whose parameter or set of parameters is denoted by θ , the MLE of θ , denoted by $\hat{\theta}$, is asymptotically unbiased and consistent for the true value θ_o .*

Definition 3.5.3. Consistency of an estimator If $Pr\{\hat{\theta} \rightarrow \theta_o\} \rightarrow 1$ as $n \rightarrow \infty$, then we say that $\hat{\theta}$ is consistent for θ_o , where $\hat{\theta}$ is an estimator for θ .

Example 3.5.2. Consider a real p -variate Gaussian population $N_p(\mu, \Sigma)$, $\Sigma > O$. Show that the MLE of μ is unbiased and consistent for μ and that the MLE of Σ is asymptotically unbiased for Σ .

Solution 3.5.2. We have $\hat{\mu} = \bar{X}$ = the sample mean or sample average and $\hat{\Sigma} = \frac{1}{n}S$ where S is the sample sum of products matrix. From Theorem 3.5.4, $E[\bar{X}] = \mu$ and $\text{Cov}(\bar{X}) = \frac{1}{n}\Sigma \rightarrow O$ as $n \rightarrow \infty$. Therefore, $\hat{\mu} = \bar{X}$ is unbiased and consistent for μ . From Theorem 3.5.8, $E[\hat{\Sigma}] = \frac{n-1}{n}\Sigma \rightarrow \Sigma$ as $n \rightarrow \infty$ and hence $\hat{\Sigma}$ is asymptotically unbiased for Σ .

Another desirable property for point estimators is referred to as sufficiency. If T is a statistic used to estimate a real scalar, vector or matrix parameter θ and if the conditional distribution of the sample values, given this statistic T , is free of θ , then no more information about θ can be secured from that sample once the statistic T is known. Accordingly, all the information that can be obtained from the sample is contained in T or, in this sense, T is *sufficient* or a *sufficient estimator* for θ .

Definition 3.5.4. Sufficiency of estimators Let θ be a scalar, vector or matrix parameter associated with a given population's distribution. Let $T = T(X_1, \dots, X_n)$ be an estimator of θ , where X_1, \dots, X_n are iid as the given population. If the conditional distribution of the sample values X_1, \dots, X_n , given T , is free of θ , then we say that this T is a *sufficient estimator* for θ . If there are several scalar, vector or matrix parameters $\theta_1, \dots, \theta_k$ associated with a given population and if $T_1(X_1, \dots, X_n), \dots, T_r(X_1, \dots, X_n)$ are r statistics, where r may be greater, smaller or equal to k , then if the conditional distribution of X_1, \dots, X_n , given T_1, \dots, T_r , is free of $\theta_1, \dots, \theta_k$, then we say that T_1, \dots, T_r are *jointly sufficient* for $\theta_1, \dots, \theta_k$. If there are several sets of statistics, where each set is sufficient for $\theta_1, \dots, \theta_k$, then that set of statistics which allows for the maximal reduction of the data is called the *minimal sufficient set of statistics* for $\theta_1, \dots, \theta_k$.

Example 3.5.3. Show that the MLE of μ in a $N_p(\mu, \Sigma)$, $\Sigma > O$, is sufficient for μ .

Solution 3.5.3. Let X_1, \dots, X_n be a simple random sample from a $N_p(\mu, \Sigma)$. Then the joint density of X_1, \dots, X_n can be written as

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu)}, \quad (i)$$

referring to (3.5.2). Since \bar{X} is a function of X_1, \dots, X_n , the joint density of X_1, \dots, X_n and \bar{X} is L itself. Hence, the conditional density of X_1, \dots, X_n , given \bar{X} , is $L/f_1(\bar{X})$

where $f_1(\bar{X})$ is the marginal density of \bar{X} . However, appealing to Corollary 3.5.1, $f_1(\bar{X})$ is $N_p(\mu, \frac{1}{n}\Sigma)$. Hence

$$\frac{L}{f_1(\bar{X})} = \frac{1}{(2\pi)^{\frac{n(p-1)}{2}} n^{\frac{p}{2}} |\Sigma|^{\frac{n-1}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}, \quad (ii)$$

which is free of μ so that $\hat{\mu}$ is sufficient for μ .

Note 3.5.1. We can also show that $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \frac{1}{n}S$ are jointly sufficient for μ and Σ in a $N_p(\mu, \Sigma)$, $\Sigma > O$, population. This results requires the density of S , which will be discussed in Chap. 5.

An additional property of interest for a point estimator is that of *relative efficiency*. If $g(\theta)$ is a function of θ and if $T = T(x_1, \dots, x_n)$ is an estimator of $g(\theta)$, then $E|T - g(\theta)|^2$ is a squared mathematical distance between T and $g(\theta)$. We can consider the following criterion: *the smaller the distance, the more efficient the estimator is*, as we would like this distance to be as small as possible when we are estimating $g(\theta)$ by making use of T . If $E[T] = g(\theta)$, then T is unbiased for $g(\theta)$ and, in this case, $E|T - g(\theta)|^2 = \text{Var}(T)$, the variance of T . In the class of unbiased estimators, we seek that particular estimator which has the smallest variance.

Definition 3.5.5. Relative efficiency of estimators If T_1 and T_2 are two estimators of the same function $g(\theta)$ of θ and if $E[|T_1 - g(\theta)|^2] < E[|T_2 - g(\theta)|^2]$, then T_1 is said to be relatively more efficient for estimating $g(\theta)$. If T_1 and T_2 are unbiased for $g(\theta)$, the criterion becomes $\text{Var}(T_1) < \text{Var}(T_2)$.

Let u be an unbiased estimator of $g(\theta)$, a function of the parameter θ associated with any population, and let T be a sufficient statistic for θ . Let the conditional expectation of u , given T , be denoted by $h(T)$, that is, $E[u|T] \equiv h(T)$. We have the two following general properties on conditional expectations, refer to Mathai and Haubold (2017), for example. For any two real scalar random variables x and y having a joint density/probability function,

$$E[y] = E[E(y|x)] \quad (3.5.9)$$

and

$$\text{Var}(y) = \text{Var}(E[y|x]) + E[\text{Var}(y|x)] \quad (3.5.10)$$

whenever the expected values exist. From (3.5.9),

$$g(\theta) = E[u] = E[E(u|T)] = E[h(T)] \Rightarrow E[h(T)] = g(\theta). \quad (3.5.11)$$

Then,

$$\begin{aligned}\text{Var}(u) &= E[u - g(\theta)]^2 = \text{Var}(E[u|T]) + E[\text{Var}(E[u|T])] = \text{Var}(h(T)) + \delta, \quad \delta \geq 0 \\ &\Rightarrow \text{Var}(u) \geq \text{Var}(h(T)),\end{aligned}\quad (3.5.12)$$

which means that if we have a sufficient statistic T for θ , then the variance of $h(T)$, with $h(T) = E[u|T]$ where u is any unbiased estimator of $g(\theta)$, is smaller than or equal to the variance of any unbiased estimator of $g(\theta)$. Accordingly, we should restrict ourselves to the class of $h(T)$ when seeking minimum variance estimators. Observe that since δ in (3.5.12) is the expected value of the variance of a real variable, it is nonnegative. The inequality in (3.5.12) is known in the literature as the *Rao-Blackwell Theorem*.

It follows from (3.5.6) that $E[\frac{\partial}{\partial\theta} \ln L] = \int_X (\frac{\partial}{\partial\theta} \ln L)L \, dX = 0$. Differentiating once again with respect to θ , we have

$$\begin{aligned}0 &= \int_X \frac{\partial}{\partial\theta} \left[\left(\frac{\partial}{\partial\theta} \ln L \right) L \right] dX = 0 \\ &\Rightarrow \int_X \left\{ \left(\frac{\partial^2}{\partial\theta^2} \ln L \right) L + \left(\frac{\partial}{\partial\theta} \ln L \right)^2 \right\} dX = 0 \\ &\Rightarrow \int_X \left(\frac{\partial}{\partial\theta} \ln L \right)^2 L \, dX = - \int_X \left(\frac{\partial^2}{\partial\theta^2} \ln L \right) L \, dX,\end{aligned}$$

so that

$$\begin{aligned}\text{Var}\left(\frac{\partial}{\partial\theta} \ln L\right) &= E\left[\frac{\partial}{\partial\theta} \ln L\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2} \ln L\right] \\ &= nE\left[\frac{\partial}{\partial\theta} \ln f(x_j, \theta)\right]^2 = -nE\left[\frac{\partial^2}{\partial\theta^2} \ln f(x_j, \theta)\right].\end{aligned}\quad (3.5.13)$$

Let T be any estimator for θ , where θ is a real scalar parameter. If T is unbiased for θ , then $E[T] = \theta$; otherwise, let $E[T] = \theta + b(\theta)$ where $b(\theta)$ is some function of θ , which is called the bias. Then, differentiating both sides with respect to θ ,

$$\begin{aligned}\int_X T L \, dX &= \theta + b(\theta) \Rightarrow \\ 1 + b'(\theta) &= \int_X T \frac{\partial}{\partial\theta} L \, dX, \quad b'(\theta) = \frac{d}{d\theta} b(\theta) \\ &\Rightarrow E\left[T \left(\frac{\partial}{\partial\theta} \ln L\right)\right] = 1 + b'(\theta) \\ &= \text{Cov}\left(T, \frac{\partial}{\partial\theta} \ln L\right)\end{aligned}$$

because $E[\frac{\partial}{\partial\theta} \ln L] = 0$. Hence,

$$\begin{aligned} [\text{Cov}(T, \frac{\partial}{\partial\theta} \ln L)]^2 &= [1 + b'(\theta)]^2 \leq \text{Var}(T) \text{Var}\left(\frac{\partial}{\partial\theta} \ln L\right) \Rightarrow \\ \text{Var}(T) &\geq \frac{[1 + b'(\theta)]^2}{\text{Var}(\frac{\partial}{\partial\theta} \ln L)} = \frac{[1 + b'(\theta)]^2}{n \text{Var}(\frac{\partial}{\partial\theta} \ln f(x_j, \theta))} \\ &= \frac{[1 + b'(\theta)]^2}{E[\frac{\partial}{\partial\theta} \ln L]^2} = \frac{[1 + b'(\theta)]^2}{n E[\frac{\partial}{\partial\theta} \ln f(x_j, \theta)]^2}, \end{aligned} \quad (3.5.14)$$

which is a lower bound for the variance of any estimator for θ . This inequality is known as the *Cramér-Rao inequality* in the literature. When T is unbiased for θ , then $b'(\theta) = 0$ and then

$$\text{Var}(T) \geq \frac{1}{I_n(\theta)} = \frac{1}{n I_1(\theta)} \quad (3.5.15)$$

where

$$\begin{aligned} I_n(\theta) &= \text{Var}\left(\frac{\partial}{\partial\theta} \ln L\right) = E\left[\frac{\partial}{\partial\theta} \ln L\right]^2 = n E\left[\frac{\partial}{\partial\theta} \ln f(x_j, \theta)\right]^2 \\ &= -E\left[\frac{\partial^2}{\partial\theta^2} \ln L\right] = -n E\left[\frac{\partial^2}{\partial\theta^2} \ln f(x_j, \theta)\right] = n I_1(\theta) \end{aligned} \quad (3.5.16)$$

is known as *Fisher's information* about θ which can be obtained from a sample of size n , $I_1(\theta)$ being *Fisher's information* in one observation or a sample of size 1. Observe that Fisher's information is different from the *information* in *Information Theory*. For instance, some aspects of Information Theory are discussed in Mathai and Rathie (1975).

Asymptotic efficiency and normality of MLE's

We have already established that

$$0 = \frac{\partial}{\partial\theta} \ln L(X, \theta)|_{\theta=\hat{\theta}}, \quad (i)$$

which is the likelihood equation giving rise to the MLE. Let us expand (i) in a neighborhood of the true parameter value θ_o :

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta} \ln L(X, \theta)|_{\theta=\theta_o} + (\hat{\theta} - \theta_o) \frac{\partial^2}{\partial\theta^2} \ln L(X, \theta)|_{\theta=\theta_o} \\ &\quad + \frac{(\hat{\theta} - \theta_o)^2}{2} \frac{\partial^3}{\partial\theta^3} \ln L(X, \theta)|_{\theta=\theta_o} \end{aligned} \quad (ii)$$

where $|\hat{\theta} - \theta_1| < |\hat{\theta} - \theta_o|$. Multiplying both sides by \sqrt{n} and rearranging terms, we have the following:

$$\sqrt{n}(\hat{\theta} - \theta_o) = \frac{-\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \ln L(X, \theta)|_{\theta=\theta_o}}{\frac{1}{n} \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta)|_{\theta=\theta_o} + \frac{1}{n} \frac{(\hat{\theta} - \theta_o)}{2} \frac{\partial^3}{\partial \theta^3} \ln L(X, \theta)|_{\theta=\theta_o}}. \quad (iii)$$

The second term in the denominator of (iii) goes to zero because $\hat{\theta} \rightarrow \theta_o$ as $n \rightarrow \infty$, and the third derivative is assumed to be bounded. Then the first term in the denominator is such that

$$\begin{aligned} \frac{1}{n} \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta)|_{\theta=\theta_o} &= \frac{1}{n} \sum_{j=1}^n \frac{\partial^2}{\partial \theta^2} \ln f(x_j, \theta)|_{\theta=\theta_o} \\ &\rightarrow E \left[\frac{\partial^2}{\partial \theta^2} \ln f(x_j, \theta) \right] = -I_1(\theta)|_{\theta_o} \\ &= -\text{Var} \left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta) \right] \Big|_{\theta=\theta_o}, \\ I_1(\theta)|_{\theta=\theta_o} &= \text{Var} \left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta) \right] \Big|_{\theta=\theta_o}, \end{aligned}$$

which is the information bound $I_1(\theta_o)$. Thus,

$$\frac{1}{n} \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta)|_{\theta=\theta_o} \rightarrow -I_1(\theta_o), \quad (iv)$$

and we may write (iii) as follows:

$$\sqrt{I_1(\theta_o)} \sqrt{n}(\hat{\theta} - \theta_o) \approx \frac{\sqrt{n}}{\sqrt{I_1(\theta_o)}} \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\theta_o}, \quad (v)$$

where $\frac{\partial}{\partial \theta} \ln f(x_j, \theta)$ has zero as its expected value and $I_1(\theta_o)$ as its variance. Further, $f(x_j, \theta)$, $j = 1, \dots, n$ are iid variables. Hence, by the central limit theorem which is stated in Sect. 2.6,

$$\frac{\sqrt{n}}{\sqrt{I_1(\theta_o)}} \frac{1}{n} \sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta) \rightarrow N_1(0, 1) \text{ as } n \rightarrow \infty. \quad (3.5.17)$$

where $N_1(0, 1)$ is a univariate standard normal random variable. This may also be re-expressed as follows since $I_1(\theta_o)$ is free of n :

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta=\theta_o} \rightarrow N_1(0, I_1(\theta_o))$$

or

$$\sqrt{I_1(\theta_o)}\sqrt{n}(\hat{\theta} - \theta_o) \rightarrow N_1(0, 1) \text{ as } n \rightarrow \infty. \quad (3.5.18)$$

Since $I_1(\theta_o)$ is free of n , this result can also be written as

$$\sqrt{n}(\hat{\theta} - \theta_o) \rightarrow N_1\left(0, \frac{1}{I_1(\theta_o)}\right). \quad (3.5.19)$$

Thus, the MLE $\hat{\theta}$ is asymptotically unbiased, consistent and asymptotically normal, referring to (3.5.18) or (3.5.19).

Example 3.5.4. Show that the MLE of the parameter θ in a real scalar exponential population is unbiased, consistent, efficient and that asymptotic normality holds as in (3.5.18).

Solution 3.5.4. As per the notations introduced in this section,

$$f(x_j, \theta) = \frac{1}{\theta}e^{-\frac{x_j}{\theta}}, \quad 0 \leq x_j < \infty, \quad \theta > 0,$$

$$L = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{j=1}^n x_j}.$$

In the exponential population, $E[x_j] = \theta$, $\text{Var}(x_j) = \theta^2$, $j = 1, \dots, n$, the MLE of θ is $\hat{\theta} = \bar{x}$, $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ and $\text{Var}(\hat{\theta}) = \frac{\theta^2}{n} \rightarrow 0$ as $n \rightarrow \infty$. Thus, $E[\hat{\theta}] = \theta$ and $\text{Var}(\hat{\theta}) \rightarrow 0$ as $n \rightarrow \infty$. Hence, $\hat{\theta}$ is unbiased and consistent for θ . Note that

$$\ln f(x_j, \theta) = -\ln \theta - \frac{1}{\theta}x_j \Rightarrow -E\left[\frac{\partial^2}{\partial \theta^2} f(x_j, \theta)\right] = -\frac{1}{\theta^2} + 2\frac{E[x_j]}{\theta^3} = \frac{1}{\theta^2} = \frac{1}{\text{Var}(x_j)}.$$

Accordingly, the information bound is attained, that is, $\hat{\theta}$ is minimum variance unbiased or most efficient. Letting the true value of θ be θ_o , by the central limit theorem, we have

$$\frac{\bar{x} - \theta_o}{\sqrt{\text{Var}(\bar{x})}} = \frac{\sqrt{n}(\hat{\theta} - \theta_o)}{\theta_o} \rightarrow N_1(0, 1) \text{ as } n \rightarrow \infty,$$

and hence the asymptotic normality is also verified. Is $\hat{\theta}$ sufficient for θ ? Let us consider the statistic $u = x_1 + \dots + x_n$, the sample sum. If u is sufficient, then $\bar{x} = \hat{\theta}$ is also sufficient. The mgf of u is given by

$$M_u(t) = \prod_{j=1}^n (1 - \theta t)^{-1} = (1 - \theta t)^{-n}, \quad 1 - \theta t > 0 \Rightarrow u \sim \text{gamma}(\alpha = n, \beta = \theta)$$

whose density is $f_1(u) = \frac{u^{n-1}}{\theta^n \Gamma(n)} e^{-\frac{u}{\theta}}$, $u = x_1 + \dots + x_n$. However, the joint density of x_1, \dots, x_n is $L = \frac{1}{\theta^n} e^{-\frac{1}{\theta}(x_1 + \dots + x_n)}$. Accordingly, the conditional density of x_1, \dots, x_n given $\hat{\theta} = \bar{x}$ is

$$\frac{L}{f_1(u)} = \frac{\Gamma(n)}{u^{n-1}},$$

which is free of θ , and hence $\hat{\theta}$ is also sufficient.

3.5.5. Some limiting properties in the p -variate case

The p -variate extension of the central limit theorem is now being considered. Let the $p \times 1$ real vectors X_1, \dots, X_n be iid with common mean value vector μ and the common covariance matrix $\Sigma > O$, that is, $E(X_j) = \mu$ and $\text{Cov}(X_j) = \Sigma > O$, $j = 1, \dots, n$. Assume that $\|\Sigma\| < \infty$ where $\|\cdot\|$ denotes a norm of (\cdot) . Letting $Y_j = \Sigma^{-\frac{1}{2}} X_j$, $E[Y_j] = \Sigma^{-\frac{1}{2}} \mu$ and $\text{Cov}(Y_j) = I$, $j = 1, \dots, n$, and letting $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, $\bar{Y} = \Sigma^{-\frac{1}{2}} \bar{X}$, $E(\bar{X}) = \mu$ and $E(\bar{Y}) = \Sigma^{-\frac{1}{2}} \mu$. If we let

$$U = \sqrt{n} \Sigma^{-\frac{1}{2}} (\bar{X} - \mu), \quad (3.5.20)$$

the following result holds:

Theorem 3.5.10. *Let the $p \times 1$ vector U be as defined in (3.5.20). Then, as $n \rightarrow \infty$, $U \rightarrow N_p(O, I)$.*

Proof: Let $L' = (a_1, \dots, a_p)$ be an arbitrary constant vector such that $L'L = 1$. Then, $L'X_j$, $j = 1, \dots, n$, are iid with common mean $L'\mu$ and common variance $\text{Var}(L'X_j) = L'\Sigma L$. Let $Y_j = \Sigma^{-\frac{1}{2}} X_j$ and $u_j = L'Y_j = L'\Sigma^{-\frac{1}{2}} X_j$. Then, the common mean of the u_j 's is $L'\Sigma^{-\frac{1}{2}} \mu$ and their common variance is $\text{Var}(u_j) = L'\Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} L = L'L = 1$, $j = 1, \dots, n$. Note that $\bar{u} = \frac{1}{n}(u_1 + \dots + u_n) = L'\bar{Y} = L'\Sigma^{-\frac{1}{2}} \bar{X}$ and that $\text{Var}(\bar{u}) = \frac{1}{n} L'L = \frac{1}{n}$. Then, in light of the univariate central limit theorem as stated in Sect. 2.6, we have $\sqrt{n} L' \Sigma^{-\frac{1}{2}} (\bar{X} - \mu) \rightarrow N_1(0, 1)$ as $n \rightarrow \infty$. If, for some p -variate vector W , $L'W$ is univariate normal for arbitrary L , it follows from a characterization of the multivariate normal distribution that W is p -variate normal vector. Thus,

$$U = \sqrt{n} \Sigma^{-\frac{1}{2}} (\bar{X} - \mu) \rightarrow N_p(O, I) \text{ as } n \rightarrow \infty, \quad (3.5.21)$$

which completes the proof.

A parallel result also holds in the complex domain. Let \tilde{X}_j , $j = 1, \dots, n$, be iid from some complex population with mean $\tilde{\mu}$ and Hermitian positive definite covariance matrix $\tilde{\Sigma} = \tilde{\Sigma}^* > O$ where $\|\tilde{\Sigma}\| < \infty$. Letting $\tilde{\bar{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$, we have

$$\sqrt{n} \tilde{\Sigma}^{-\frac{1}{2}}(\tilde{\bar{X}} - \tilde{\mu}) \rightarrow N_p(O, I) \text{ as } n \rightarrow \infty. \quad (3.5a.5)$$

Exercises 3.5

3.5.1. By making use of the mgf or otherwise, show that the sample mean $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ in the real p -variate Gaussian case, $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, is again Gaussian distributed with the parameters μ and $\frac{1}{n}\Sigma$.

3.5.2. Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$ and iid. Let $\mathbf{X} = (X_1, \dots, X_n)$ be the $p \times n$ sample matrix. Derive the density of (1) $\text{tr}(\Sigma^{-1}(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})')$ where $\mathbf{M} = (\mu, \dots, \mu)$ or a $p \times n$ matrix where all the columns are μ ; (2) $\text{tr}(\mathbf{X}\mathbf{X}')$. Derive the densities in both the cases, including the noncentrality parameter.

3.5.3. Let the $p \times 1$ real vector $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$ for $j = 1, \dots, n$ and iid. Let $\mathbf{X} = (X_1, \dots, X_n)$ the $p \times n$ sample matrix. Derive the density of $\text{tr}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ where $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ is the $p \times n$ matrix where every column is \bar{X} .

3.5.4. Repeat Exercise 3.5.1 for the p -variate complex Gaussian case.

3.5.5. Repeat Exercise 3.5.2 for the complex Gaussian case and write down the density explicitly.

3.5.6. Consider a real bivariate normal density with the parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. Write down the density explicitly. Consider a simple random sample of size n , X_1, \dots, X_n , from this population where X_j is 2×1 , $j = 1, \dots, n$. Then evaluate the MLE of these five parameters by (1) by direct evaluation, (2) by using the general formula.

3.5.7. In Exercise 3.5.6 evaluate the maximum likelihood estimates of the five parameters if the following is an observed sample from this bivariate normal population:

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

3.5.8. Repeat Exercise 3.5.6 if the population is a bivariate normal in the complex domain.

3.5.9. Repeat Exercise 3.5.7 if the following is an observed sample from the complex bivariate normal population referred to in Exercise 3.5.8:

$$\begin{bmatrix} 1 + 2i \\ i \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3i \\ 1 - i \end{bmatrix}, \begin{bmatrix} 2 - i \\ i \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

3.5.10. Let the $p \times 1$ real vector X_1 be Gaussian distributed, $X_1 \sim N_p(O, I)$. Consider the quadratic forms $u_1 = X_1' A_1 X_1$, $u_2 = X_1' A_2 X_1$. Let $A_j = A_j^2$, $j = 1, 2$ and $A_1 + A_2 = I$. What can you say about the chisquaredness and independence of u_1 and u_2 ? Prove your assertions.

3.5.11. Let $X_1 \sim N_p(O, I)$. Let $u_j = X_1' A_j X_1$, $A_j = A_j^2$, $j = 1, \dots, k$, $A_1 + \dots + A_k = I$. What can you say about the chisquaredness and independence of the u_j 's? Prove your assertions.

5.3.12. Repeat Exercise 3.5.11 for the complex case.

5.3.13. Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$ and iid. Let $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$. Show that the exponent in the density of \bar{X} , excluding $-\frac{1}{2}$, namely, $\sqrt{n}(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu) \sim \chi_p^2$. Derive the density of $\text{tr}(\bar{X}' \Sigma^{-1} \bar{X})$.

3.5.14. Let $Q = \sqrt{n}(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu)$ as in Exercise 5.3.13. For a given α consider the probability statement $Pr\{Q \geq b\} = \alpha$. Show that $b = \chi_{p,\alpha}^2$ where $Pr\{\chi_p^2 \geq \chi_{p,\alpha}^2\} = \alpha$.

3.5.15. Let $Q_1 = \sqrt{n}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)$ where \bar{X} , Σ and μ are all as defined in Exercise 5.3.14. If $\mu_o \neq \mu$, show that $Q_1 \sim \chi_p^2(\lambda)$ where the noncentrality parameter $\lambda = \frac{1}{2}(\mu - \mu_o)' \Sigma^{-1}(\mu - \mu_o)$.

3.6. Elliptically Contoured Distribution, Real Case

Let X be a real $p \times 1$ vector of distinct real scalar variables with x_1, \dots, x_p as its components. For some $p \times 1$ parameter vector B and $p \times p$ positive definite constant matrix $A > O$, consider the positive definite quadratic form $(X - B)' A(X - B)$. We have encountered such a quadratic form in the exponent of a real p -variate Gaussian density, in which case $B = \mu$ is the mean value vector and $A = \Sigma^{-1}$, Σ being the positive definite covariance matrix. Let $g(\cdot) \geq 0$ be a non-negative function such that $|A|^{\frac{1}{2}} g((X - B)' A(X - B)) \geq 0$ and

$$\int_X |A|^{\frac{1}{2}} g((X - B)' A(X - B)) dX = 1, \quad (3.6.1)$$

so that $|A|^{\frac{1}{2}} g((X - B)' A(X - B))$ is a statistical density. Such a density is referred to as an *elliptically contoured density*.

3.6.1. Some properties of elliptically contoured distributions

Let $Y = A^{\frac{1}{2}}(X - B)$. Then, from Theorem 1.6.1, $dX = |A|^{-\frac{1}{2}} dY$ and from (3.6.1),

$$\int_Y g(Y' Y) dY = 1 \quad (3.6.2)$$

where

$$Y'Y = y_1^2 + y_2^2 + \cdots + y_p^2, \quad Y' = (y_1, \dots, y_p).$$

We can further simplify (3.6.2) via a general polar coordinate transformation:

$$\begin{aligned} y_1 &= r \sin \theta_1 \\ y_2 &= r \cos \theta_1 \sin \theta_2 \\ y_3 &= r \cos \theta_1 \cos \theta_2 \sin \theta_3 \\ &\vdots \\ y_{p-2} &= r \cos \theta_1 \cdots \cos \theta_{p-3} \sin \theta_{p-2} \\ y_{p-1} &= r \cos \theta_1 \cdots \cos \theta_{p-2} \sin \theta_{p-1} \\ y_p &= r \cos \theta_1 \cdots \cos \theta_{p-1} \end{aligned} \quad (3.6.3)$$

for $-\frac{\pi}{2} < \theta_j \leq \frac{\pi}{2}$, $j = 1, \dots, p-2$, $-\pi < \theta_{p-1} \leq \pi$, $0 \leq r < \infty$. It then follows that

$$dy_1 \wedge \cdots \wedge dy_p = r^{p-1} (\cos \theta_1)^{p-1} \cdots (\cos \theta_{p-1}) dr \wedge d\theta_1 \wedge \cdots \wedge d\theta_{p-1}. \quad (3.6.4)$$

Thus,

$$y_1^2 + \cdots + y_p^2 = r^2.$$

Given (3.6.3) and (3.6.4), observe that $r, \theta_1, \dots, \theta_{p-1}$ are mutually independently distributed. Separating the factors containing θ_i from (3.6.4) and then, normalizing it, we have

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_i)^{p-i-1} d\theta_i = 1 \Rightarrow 2 \int_0^{\frac{\pi}{2}} (\cos \theta_i)^{p-i-1} d\theta_i = 1. \quad (i)$$

Let $u = \sin \theta_i \Rightarrow du = \cos \theta_i d\theta_i$. Then (i) becomes $2 \int_0^1 (1-u^2)^{\frac{p-i}{2}-1} du = 1$, and letting $v = u^2$ gives (i) as

$$\int_0^1 v^{\frac{1}{2}-1} (1-v)^{\frac{p-i}{2}-1} dv = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{p-i}{2})}{\Gamma(\frac{p-i+1}{2})}. \quad (ii)$$

Thus, the density of θ_j , denoted by $f_j(\theta_j)$, is

$$f_j(\theta_j) = \frac{\Gamma(\frac{p-j+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{p-j}{2})} (\cos \theta_j)^{p-j-1}, \quad -\frac{\pi}{2} < \theta_j \leq \frac{\pi}{2}, \quad (3.6.5)$$

and zero, elsewhere, for $j = 1, \dots, p-2$, and

$$f_{p-1}(\theta_{p-1}) = \frac{1}{2\pi}, \quad -\pi < \theta_{p-1} \leq \pi, \quad (iii)$$

and zero elsewhere. Taking the product of the $p - 2$ terms in (ii) and (iii), the total integral over the θ_j 's is available as

$$\left\{ \prod_{j=1}^{p-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_j)^{p-j-1} d\theta_j \right\} \int_{-\pi}^{\pi} d\theta_{p-1} = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}. \quad (3.6.6)$$

The expression in (3.6.6), excluding 2, can also be obtained by making the transformation $s = Y'Y$ and then writing ds in terms of dY by appealing to Theorem 4.2.3.

3.6.2. The density of $u = r^2$

From (3.6.2) and (3.6.3),

$$\frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^{\infty} r^{p-1} g(r^2) dr = 1, \quad (3.6.7)$$

that is,

$$2 \int_{r=0}^{\infty} r^{p-1} g(r^2) dr = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}. \quad (iv)$$

Letting $u = r^2$, we have

$$\int_0^{\infty} u^{\frac{p}{2}-1} g(u) du = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}, \quad (v)$$

and the density of r , denoted by $f_r(r)$, is available from (3.6.7) as

$$f_r(r) = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} r^{p-1} g(r^2), \quad 0 \leq r < \infty, \quad (3.6.8)$$

and zero, elsewhere. The density of $u = r^2$ is then

$$f_u(u) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} u^{\frac{p}{2}-1} g(u), \quad 0 \leq u < \infty, \quad (3.6.9)$$

and zero, elsewhere. Considering the density of Y given in (3.6.2), we may observe that y_1, \dots, y_p are identically distributed.

Theorem 3.6.1. *If $y_j = r u_j$, $j = 1, \dots, p$, in the transformation in (3.6.3), then $E[u_j^2] = \frac{1}{p}$, $j = 1, \dots, p$, and u_1, \dots, u_p are uniformly distributed.*

Proof: From (3.6.3), $y_j = ru_j$, $j = 1, \dots, p$. We may observe that $u_1^2 + \dots + u_p^2 = 1$ and that u_1, \dots, u_p are identically distributed. Hence $E[u_1^2] + E[u_2^2] + \dots + E[u_p^2] = 1 \Rightarrow E[u_j^2] = \frac{1}{p}$.

Theorem 3.6.2. Consider the y_j 's in Eq. (3.6.2). If $g(u)$ is free of p and if $E[u] < \infty$, then $E[y_j^2] = \frac{1}{2\pi}$; otherwise, $E[y_j^2] = \frac{1}{p}E[u]$ provided $E[u]$ exists.

Proof: Since r and u_j are independently distributed and since $E[u_j^2] = \frac{1}{p}$ in light of Theorem 3.6.1, $E[y_j^2] = E[r^2]E[u_j^2] = \frac{1}{p}E[r^2] = \frac{1}{p}E[u]$. From (3.6.9),

$$\int_0^\infty u^{\frac{p}{2}-1} g(u) du = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}. \tag{vi}$$

However,

$$E[u] = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_{u=0}^\infty u^{\frac{p}{2}+1-1} g(u) du. \tag{vii}$$

Thus, assuming that $g(u)$ is free of p , that $\frac{p}{2}$ can be taken as a parameter and that (vii) is convergent,

$$E[y_j^2] = \frac{1}{p}E[r^2] = \frac{1}{p}E[u] = \frac{1}{p} \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \frac{\Gamma(\frac{p}{2} + 1)}{\pi^{\frac{p}{2}+1}} = \frac{1}{2\pi}; \tag{3.6.10}$$

otherwise, $E[y_j^2] = \frac{1}{p}E[u]$ as long as $E[u] < \infty$.

3.6.3. Mean value vector and covariance matrix

From (3.6.1),

$$E[X] = |A|^{\frac{1}{2}} \int_X X g((X - B)'A(X - B)) dX.$$

Noting that

$$\begin{aligned} E[X] &= E[X - B + B] = B + E[X - B] \\ &= B + |A|^{\frac{1}{2}} \int_X (X - B) g((X - B)'A(X - B)) dX \end{aligned}$$

and letting $Y = A^{\frac{1}{2}}(X - B)$, we have

$$E[X] = B + \int_Y Y g(Y'Y) dY, \quad -\infty < y_j < \infty, \quad j = 1, \dots, p.$$

But $Y'Y$ is even whereas each element in Y is linear and odd. Hence, if the integral exists, $\int_Y Y g(Y'Y) dY = O$ and so, $E[X] = B \equiv \mu$. Let $V = \text{Cov}(X)$, the covariance matrix associated with X . Then

$$V = E[(X - \mu)(X - \mu)'] = A^{-\frac{1}{2}} \left[\int_Y (YY') g(Y'Y) dY \right] A^{-\frac{1}{2}},$$

where

$$Y = A^{\frac{1}{2}}(X - \mu), \quad (3.6.11)$$

$$YY' = \begin{bmatrix} y_1^2 & y_1 y_2 & \cdots & y_1 y_p \\ y_2 y_1 & y_2^2 & \cdots & y_2 y_p \\ \vdots & \vdots & \ddots & \vdots \\ y_p y_1 & y_p y_2 & \cdots & y_p^2 \end{bmatrix}. \quad (\text{viii})$$

Since the non-diagonal elements, $y_i y_j$, $i \neq j$, are odd and $g(Y'Y)$ is even, the integrals over the non-diagonal elements are equal to zero whenever the second moments exist. Since $E[Y] = O$, $V = E(YY')$. It has already been determined in (3.6.10) that $E[y_j^2] = \frac{1}{2\pi}$ for $j = 1, \dots, p$, whenever $g(u)$ is free of p and $E[u]$ exists, the density of u being as specified in (3.6.9). If $g(u)$ is not free of p , the diagonal elements will each integrate out to $\frac{1}{p}E[r^2]$. Accordingly,

$$\text{Cov}(X) = V = \frac{1}{2\pi} A^{-1} \quad \text{or} \quad V = \frac{1}{p} E[r^2] A^{-1}. \quad (3.6.12)$$

Theorem 3.6.3. *When X has the p -variate elliptically contoured distribution defined in (3.6.1), the mean value vector of X , $E[X] = B$ and the covariance of X , denoted by Σ , is such that $\Sigma = \frac{1}{p} E[r^2] A^{-1}$ where A is the parameter matrix in (3.6.1), $u = r^2$ and r is defined in the transformation (3.6.3).*

3.6.4. Marginal and conditional distributions

Consider the density

$$f(X) = |A|^{\frac{1}{2}} g((X - \mu)' A (X - \mu)), \quad A > O, \quad -\infty < x_j < \infty, \quad -\infty < \mu_j < \infty \quad (3.6.13)$$

where $X' = (x_1, \dots, x_p)$, $\mu' = (\mu_1, \dots, \mu_p)$, $A = (a_{ij}) > O$. Consider the following partitioning of X , μ and A :

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where X_1 , μ_1 are $p_1 \times 1$, X_2 , μ_2 are $p_2 \times 1$, A_{11} is $p_1 \times p_1$ and A_{22} is $p_2 \times p_2$, $p_1 + p_2 = p$. Then, as was established in Sect. 3.3,

$$\begin{aligned}
(X - \mu)'A(X - \mu) &= (X_1 - \mu_{(1)})'A_{11}(X_1 - \mu_{(1)}) + 2(X_2 - \mu_{(2)})'A_{21}(X_1 - \mu_{(1)}) \\
&\quad + (X_2 - \mu_{(2)})'A_{22}(X_2 - \mu_{(2)}) \\
&= (X_1 - \mu_{(1)})'[A_{11} - A_{12}A_{22}^{-1}A_{21}](X_1 - \mu_{(1)}) \\
&\quad + (X_2 - \mu_{(2)} + C)'A_{22}(X_2 - \mu_{(2)} + C), \quad C = A_{22}^{-1}A_{21}(X_1 - \mu_{(1)}).
\end{aligned}$$

In order to obtain the marginal density of X_1 , we integrate out X_2 from $f(X)$. Let the marginal densities of X_1 and X_2 be respectively denoted by $g_1(X_1)$ and $g_2(X_2)$. Then

$$\begin{aligned}
g_1(X_1) &= |A|^{\frac{1}{2}} \int_{X_2} g((X_1 - \mu_{(1)})'[A_{11} - A_{12}A_{22}^{-1}A_{21}](X_1 - \mu_{(1)}) \\
&\quad + (X_2 - \mu_{(2)} + C)'A_{22}(X_2 - \mu_{(2)} + C))dX_2.
\end{aligned}$$

Letting $A_{22}^{\frac{1}{2}}(X_2 - \mu_{(2)} + C) = Y_2$, $dY_2 = |A_{22}|^{\frac{1}{2}} dX_2$ and

$$g_1(X_1) = |A|^{\frac{1}{2}} |A_{22}|^{-\frac{1}{2}} \int_{Y_2} g((X_1 - \mu_{(1)})'[A_{11} - A_{12}A_{22}^{-1}A_{21}](X_1 - \mu_{(1)}) + Y_2'Y_2) dY_2.$$

Note that $|A| = |A_{22}| |A_{11} - A_{12}A_{22}^{-1}A_{21}|$ from the results on partitioned matrices presented in Sect. 1.3 and thus, $|A|^{\frac{1}{2}} |A_{22}|^{-\frac{1}{2}} = |A_{11} - A_{12}A_{22}^{-1}A_{21}|^{\frac{1}{2}}$. We have seen that Σ^{-1} is a constant multiple of A where Σ is the covariance matrix of the $p \times 1$ vector X . Then

$$(\Sigma^{11})^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

which is a constant multiple of $A_{11} - A_{12}A_{22}^{-1}A_{21}$. If $Y_2'Y_2 = s_2$, then from Theorem 4.2.3,

$$dY_2 = \frac{\pi^{\frac{p_2}{2}}}{\Gamma(\frac{p_2}{2})} \int_{s_2 > 0} s_2^{\frac{p_2}{2}-1} g(s_2 + u_1) ds_2 \quad (3.6.14)$$

where $u_1 = (X_1 - \mu_{(1)})'[A_{11} - A_{12}A_{22}^{-1}A_{21}](X_1 - \mu_{(1)})$. Note that (3.6.14) is elliptically contoured or X_1 has an elliptically contoured distribution. Similarly, X_2 has an elliptically contoured distribution. Letting $Y_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{\frac{1}{2}}(X_1 - \mu_{(1)})$, then Y_{11} has a spherically symmetric distribution. Denoting the density of Y_{11} by $g_{11}(Y_{11})$, we have

$$g_1(Y_{11}) = \frac{\pi^{\frac{p_2}{2}}}{\Gamma(\frac{p_2}{2})} \int_{s_2 > 0} s_2^{\frac{p_2}{2}-1} g(s_2 + Y_{11}'Y_{11}) ds_2. \quad (3.6.15)$$

By a similar argument, the marginal density of X_2 , namely $g_2(X_2)$, and the density of Y_{22} , namely $g_{22}(Y_{22})$, are as follows:

$$\begin{aligned} g_2(X_2) &= |A_{22} - A_{21}A_{11}^{-1}A_{12}|^{\frac{1}{2}} \frac{\pi^{\frac{p_1}{2}}}{\Gamma(\frac{p_1}{2})} \\ &\quad \times \int_{s_1>0} s_1^{\frac{p_1}{2}-1} g(s_1 + (X_2 - \mu_{(2)})'[A_{22} - A_{21}A_{11}^{-1}A_{12}](X_2 - \mu_{(2)})) ds_1, \\ g_{22}(Y_{22}) &= \frac{\pi^{\frac{p_1}{2}}}{\Gamma(\frac{p_1}{2})} \int_{s_1>0} s_1^{\frac{p_1}{2}-1} g(s_1 + Y_{22}'Y_{22}) ds_1. \end{aligned} \quad (3.6.16)$$

3.6.5. The characteristic function of an elliptically contoured distribution

Let T be a $p \times 1$ parameter vector, $T' = (t_1, \dots, t_p)$, so that $T'X = t_1x_1 + \dots + t_px_p$. Then, the characteristic function of X , denoted by $\phi_X(T)$, is $E[e^{iT'X}]$ where E denotes the expected value and $i = \sqrt{-1}$, that is,

$$\phi_X(T) = E[e^{iT'X}] = \int_X e^{iT'X} |A|^{\frac{1}{2}} g((X - \mu)'A(X - \mu)) dX. \quad (3.6.17)$$

Writing X as $X - \mu + \mu$ and then making the transformation $Y = A^{\frac{1}{2}}(X - \mu)$, we have

$$\phi_X(T) = e^{iT'\mu} \int_Y e^{iT'A^{-\frac{1}{2}}Y} g(Y'Y) dY. \quad (3.6.18)$$

However, $g(Y'Y)$ is invariant under orthonormal transformation of the type $Z = PY$, $PP' = I$, $P'P = I$, as $Z'Z = Y'Y$ so that $g(Y'Y) = g(Z'Z)$ for all orthonormal matrices. Thus,

$$\phi_X(T) = e^{iT'\mu} \int_Z e^{iT'A^{-\frac{1}{2}}P'Z} g(Z'Z) dZ \quad (3.6.19)$$

for all P . This means that the integral in (3.6.19) is a function of $(T'A^{-\frac{1}{2}})(T'A^{-\frac{1}{2}})' = T'A^{-1}T$, say $\psi(T'A^{-1}T)$. Then,

$$\phi_X(T) = e^{iT'\mu} \psi(T'A^{-1}T) \quad (3.6.20)$$

where A^{-1} is proportional to Σ , the covariance matrix of X , and

$$\frac{\partial}{\partial T} \phi_X(T)|_{T=0} = i\mu \Rightarrow E(X) = \mu;$$

the reader may refer to Chap. 1 for vector/matrix derivatives. Now, considering $\phi_{X-\mu}(T)$, we have

$$\begin{aligned}\frac{\partial}{\partial T}\phi_{X-\mu}(T) &= \frac{\partial}{\partial T}\psi(T'A^{-1}T) = \psi'(TA^{-1}T)2A^{-1}T \\ &\Rightarrow \frac{\partial}{\partial T'}\psi(T'A^{-1}T) = \psi'(T'A^{-1}T)2T'A^{-1} \\ &\Rightarrow \frac{\partial}{\partial T}\frac{\partial}{\partial T'}\psi(T'A^{-1}T) = \psi''(T'A^{-1}T)(2A^{-1}T)(2T'A^{-1}) + \psi'(TA^{-1}T)2A^{-1} \\ &\Rightarrow \frac{\partial}{\partial T}\frac{\partial}{\partial T'}\psi(T'A^{-1}T)|_{T=O} = 2A^{-1},\end{aligned}$$

assuming that $\psi'(T'A^{-1}T)|_{T=O} = 1$ and $\psi''(T'A^{-1}T)|_{T=O} = 1$, where $\psi'(u) = \frac{d}{du}\psi(u)$ for a real scalar variable u and $\psi''(u)$ denotes the second derivative of ψ with respect to u . The same procedure can be utilized to obtain higher order moments of the type $E[\cdots XX'XX']$ by repeatedly applying vector derivatives to $\phi_X(T)$ as $\cdots \frac{\partial}{\partial T}\frac{\partial}{\partial T'}$ operating on $\phi_X(T)$ and then evaluating the result at $T = O$. Similarly, higher order central moments of the type $E[\cdots (X - \mu)(X - \mu)'(X - \mu)(X - \mu)']$ are available by applying the vector differential operator $\cdots \frac{\partial}{\partial T}\frac{\partial}{\partial T'}$ on $\psi(T'A^{-1}T)$ and then evaluating the result at $T = O$. However, higher moments with respect to individual variables, such as $E[x_j^k]$, are available by differentiating $\phi_X(T)$ partially k times with respect to t_j , and then evaluating the resulting expression at $T = O$. If central moments are needed then the differentiation is done on $\psi(T'A^{-1}T)$.

Thus, we can obtain results parallel to those derived for the p -variate Gaussian distribution by applying the same procedures on elliptically contoured distributions. Accordingly, further discussion of elliptically contoured distributions will not be taken up in the coming chapters.

Exercises 3.6

3.6.1. Let x_1, \dots, x_k be independently distributed real scalar random variables with density functions $f_j(x_j)$, $j = 1, \dots, k$. If the joint density of x_1, \dots, x_k is of the form $f_1(x_1)\cdots f_k(x_k) = g(x_1^2 + \cdots + x_k^2)$ for some differentiable function g , show that x_1, \dots, x_k are identically distributed as Gaussian random variables.

3.6.2. Letting the real scalar random variables x_1, \dots, x_k have a joint density such that $f(x_1, \dots, x_k) = c$ for $x_1^2 + \cdots + x_k^2 \leq r^2$, $r > 0$, show that (1) (x_1, \dots, x_k) is uniformly distributed over the volume of the k -dimensional sphere; (2) $E[x_j] = 0$, $\text{Cov}(x_i, x_j) = 0$, $i \neq j = 1, \dots, k$; (3) x_1, \dots, x_k are not independently distributed.

3.6.3. Let $u = (X - B)'A(X - B)$ in Eq. (3.6.1), where $A > O$ and A is a $p \times p$ matrix. Let $g(u) = c_1(1 - au)^\rho$, $a > 0$, $1 - au > 0$ and c_1 is an appropriate constant. If $|A|^{1/2}g(u)$ is a density, show that (1) this density is elliptically contoured; (2) evaluate its normalizing constant and specify the conditions on the parameters.

3.6.4. Solve Exercise 3.6.3 for $g(u) = c_2(1 + au)^{-\rho}$, where c_2 is an appropriate constant.

3.6.5. Solve Exercise 3.6.3 for $g(u) = c_3 u^{\gamma-1}(1 - au)^{\beta-1}$, $a > 0$, $1 - au > 0$ and c_3 is an appropriate constant.

3.6.6. Solve Exercise 3.6.3 for $g(u) = c_4 u^{\gamma-1}e^{-au}$, $a > 0$ where c_4 is an appropriate constant.

3.6.7. Solve Exercise 3.6.3 for $g(u) = c_5 u^{\gamma-1}(1 + au)^{-(\rho+\gamma)}$, $a > 0$ where c_5 is an appropriate constant.

3.6.8. Solve Exercises 3.6.3 to 3.6.7 by making use of the general polar coordinate transformation.

3.6.9. Let $s = y_1^2 + \dots + y_p^2$ where y_j , $j = 1, \dots, p$, are real scalar random variables. Let $dY = dy_1 \wedge \dots \wedge dy_p$ and let ds be the differential in s . Then, it can be shown that $dY = \frac{\pi^{p/2}}{\Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} ds$. By using this fact, solve Exercises 3.6.3–3.6.7.

3.6.10. If $A = \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix}$, write down the elliptically contoured density in (3.6.1) explicitly by taking an arbitrary $b = E[X] = \mu$, if (1) $g(u) = (a - cu)^\alpha$, $a > 0$, $c > 0$, $a - cu > 0$; (2) $g(u) = (a + cu)^{-\beta}$, $a > 0$, $c > 0$, and specify the conditions on α and β .

References

- A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.
- A.M. Mathai and Hans J. Haubold (2017): *Probability and Statistics, A Course for Physicists and Engineers*, De Gruyter, Germany.
- A.M. Mathai and P.N. Rathie (1975): *Basic Concepts and Information Theory and Statistics: Axiomatic Foundations and Applications*, Wiley Halsted, New York.
- A.M. Mathai and Serge B. Provost (1992): *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, New York.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

The Matrix-Variate Gaussian Distribution



4.1. Introduction

This chapter relies on various results presented in Chap. 1. We will introduce a class of integrals called the real matrix-variate Gaussian integrals and complex matrix-variate Gaussian integrals wherefrom a statistical density referred to as the matrix-variate Gaussian density and, as a special case, the multivariate Gaussian or normal density will be obtained, both in the real and complex domains.

The notations introduced in Chap. 1 will also be utilized in this chapter. Scalar variables, mathematical and random, will be denoted by lower case letters, vector/matrix variables will be denoted by capital letters, and complex variables will be indicated by a tilde. Additionally, the following notations will be used. All the matrices appearing in this chapter are $p \times p$ real positive definite or Hermitian positive definite unless stated otherwise. $X > O$ will mean that the $p \times p$ real symmetric matrix X is positive definite and $\tilde{X} > O$, that the $p \times p$ matrix \tilde{X} in the complex domain is Hermitian, that is, $\tilde{X} = \tilde{X}^*$ where \tilde{X}^* denotes the conjugate transpose of \tilde{X} and \tilde{X} is positive definite. $O < A < X < B$ will indicate that the $p \times p$ real positive definite matrices are such that $A > O$, $B > O$, $X > O$, $X - A > O$, $B - X > O$. $\int_X f(X) dX$ represents a real-valued scalar function $f(X)$ being integrated out over all X in the domain of X where dX stands for the wedge product of differentials of all distinct elements in X . If $X = (x_{ij})$ is a real $p \times q$ matrix, the x_{ij} 's being distinct real scalar variables, then $dX = dx_{11} \wedge dx_{12} \wedge \dots \wedge dx_{pq}$ or $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. If $X = X'$, that is, X is a real symmetric matrix of dimension $p \times p$, then $dX = \wedge_{i \geq j=1}^p dx_{ij} = \wedge_{i \leq j=1}^p dx_{ij}$, which involves only $p(p+1)/2$ differential elements dx_{ij} . When taking the wedge product, the elements x_{ij} 's may be taken in any convenient order to start with. However, that order has to be maintained until the computations are completed. If $\tilde{X} = X_1 + iX_2$, where X_1 and X_2 are real $p \times q$ matrices, $i = \sqrt{-1}$, then $d\tilde{X}$ will be defined as $d\tilde{X} = dX_1 \wedge dX_2$. $\int_{A < \tilde{X} < B} f(\tilde{X}) d\tilde{X}$ represents the real-valued scalar function f of complex matrix argument \tilde{X} being integrated out over

all $p \times p$ matrix \tilde{X} such that $A > O$, $\tilde{X} > O$, $B > O$, $\tilde{X} - A > O$, $B - \tilde{X} > O$ (all Hermitian positive definite), where A and B are constant matrices in the sense that they are free of the elements of \tilde{X} . The corresponding integral in the real case will be denoted by $\int_{A < X < B} f(X) dX = \int_A^B f(X) dX$, $A > O$, $X > O$, $X - A > O$, $B > O$, $B - X > O$, where A and B are constant matrices, all the matrices being of dimension $p \times p$.

4.2. Real Matrix-variate and Multivariate Gaussian Distributions

Let $X = (x_{ij})$ be a $p \times q$ matrix whose elements x_{ij} are distinct real variables. For any real matrix X , be it square or rectangular, $\text{tr}(XX') = \text{tr}(X'X) = \text{sum of the squares of all the elements of } X$. Note that XX' need not be equal to $X'X$. Thus, $\text{tr}(XX') = \sum_{i=1}^p \sum_{j=1}^q x_{ij}^2$ and, in the complex case, $\text{tr}(\tilde{X}\tilde{X}^*) = \sum_{i=1}^p \sum_{j=1}^q |\tilde{x}_{ij}|^2$ where if $\tilde{x}_{rs} = x_{rs1} + ix_{rs2}$ where x_{rs1} and x_{rs2} are real, $i = \sqrt{-1}$, with $|\tilde{x}_{rs}| = +[x_{rs1}^2 + x_{rs2}^2]^{\frac{1}{2}}$. Consider the following integrals over the real rectangular $p \times q$ matrix X :

$$\begin{aligned} I_1 &= \int_X e^{-\text{tr}(XX')} dX = \int_X e^{-\sum_{i=1}^p \sum_{j=1}^q x_{ij}^2} dX = \prod_{i,j} \int_{-\infty}^{\infty} e^{-x_{ij}^2} dx_{ij} \\ &= \prod_{i,j} \sqrt{\pi} = \pi^{\frac{pq}{2}}, \end{aligned} \quad (i)$$

$$I_2 = \int_X e^{-\frac{1}{2}\text{tr}(XX')} dX = (2\pi)^{\frac{pq}{2}}. \quad (ii)$$

Let $A > O$ be $p \times p$ and $B > O$ be $q \times q$ constant positive definite matrices. Then we can define the unique positive definite square roots $A^{\frac{1}{2}}$ and $B^{\frac{1}{2}}$. For the discussions to follow, we need only the representations $A = A_1 A_1'$, $B = B_1 B_1'$ with A_1 and B_1 nonsingular, a prime denoting the transpose. For an $m \times n$ real matrix X , consider

$$\begin{aligned} \text{tr}(AXBX') &= \text{tr}(A^{\frac{1}{2}} A^{\frac{1}{2}} X B^{\frac{1}{2}} B^{\frac{1}{2}} X') = \text{tr}(A^{\frac{1}{2}} X B^{\frac{1}{2}} B^{\frac{1}{2}} X' A^{\frac{1}{2}}) \\ &= \text{tr}(YY'), \quad Y = A^{\frac{1}{2}} X B^{\frac{1}{2}}. \end{aligned} \quad (iii)$$

In order to obtain the above results, we made use of the property that for any two matrices P and Q such that PQ and QP are defined, $\text{tr}(PQ) = \text{tr}(QP)$ where PQ need not be equal to QP . As well, letting $Y = (y_{ij})$, $\text{tr}(YY') = \sum_{i=1}^p \sum_{j=1}^q y_{ij}^2$. YY' is real positive definite when Y is $p \times q$, $p \leq q$, is of full rank p . Observe that any real square matrix U that can be written as $U = VV'$ for some matrix V where V may be square or rectangular, is either positive definite or at least positive semi-definite. When V is a $p \times q$ matrix, $q \geq p$, whose rank is p , VV' is positive definite; if the rank of V is less than p , then VV' is positive semi-definite. From Result 1.6.4,

$$\begin{aligned} Y = A^{\frac{1}{2}} X B^{\frac{1}{2}} &\Rightarrow dY = |A|^{\frac{q}{2}} |B|^{\frac{p}{2}} dX \\ &\Rightarrow dX = |A|^{-\frac{q}{2}} |B|^{-\frac{p}{2}} dY \end{aligned} \quad (iv)$$

where we use the standard notation $|\cdot| = \det(\cdot)$ to denote the determinant of (\cdot) in general and $|\det(\cdot)|$ to denote the absolute value or modulus of the determinant of (\cdot) in the complex domain. Let

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2} \text{tr}(AXBX')}, \quad A > O, \quad B > O \quad (4.2.1)$$

for $X = (x_{ij})$, $-\infty < x_{ij} < \infty$ for all i and j . From the steps (i) to (iv), we see that $f_{p,q}(X)$ in (4.2.1) is a statistical density over the real rectangular $p \times q$ matrix X . This function $f_{p,q}(X)$ is known as the *real matrix-variate Gaussian density*. We introduced a $\frac{1}{2}$ in the exponent so that particular cases usually found in the literature agree with the real p -variate Gaussian distribution. Actually, this $\frac{1}{2}$ factor is quite unnecessary from a mathematical point of view as it complicates computations rather than simplifying them. In the complex case, the factor $\frac{1}{2}$ does not appear in the exponent of the density, which is consistent with the current particular cases encountered in the literature.

Note 4.2.1. If the factor $\frac{1}{2}$ is omitted in the exponent, then 2π is to be replaced by π in the denominator of (4.2.1), namely,

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(\pi)^{\frac{pq}{2}}} e^{-\text{tr}(AXBX')}, \quad A > O, \quad B > O. \quad (4.2.2)$$

When $p = 1$, the matrix X is $1 \times q$ and we let $X = (x_1, \dots, x_q)$ where X is a row vector whose components are x_1, \dots, x_q . When $p = 1$, A is 1×1 or a scalar quantity. Letting $A = 1$ and $B = V^{-1}$, $V > O$, be of dimension $q \times q$, then in the real case,

$$\begin{aligned} f_{1,q}(X) &= \frac{|\frac{1}{2}V^{-1}|^{\frac{1}{2}}}{\pi^{\frac{q}{2}}} e^{-\frac{1}{2}XV^{-1}X'}, \quad X = (x_1, \dots, x_q), \\ &= \frac{1}{(2\pi)^{\frac{q}{2}} |V|^{\frac{1}{2}}} e^{-\frac{1}{2}XV^{-1}X'}, \end{aligned} \quad (4.2.3)$$

which is the usual real nonsingular Gaussian density with parameter matrix V , that is, $X' \sim N_q(O, V)$. If a location parameter vector $\mu = (\mu_1, \dots, \mu_q)$ is introduced or, equivalently, if X is replaced by $X - \mu$, then we have

$$f_{1,q}(X) = [(2\pi)^{\frac{q}{2}} |V|^{\frac{1}{2}}]^{-1} e^{-\frac{1}{2}(X-\mu)V^{-1}(X-\mu)'}, \quad V > O. \quad (4.2.4)$$

On the other hand, when $q = 1$, a real p -variate Gaussian or normal density is available from (4.2.1) wherein $B = 1$; in this case, $X \sim N_p(\mu, A^{-1})$ where X and the location

parameter vector μ are now $p \times 1$ column vectors. This density is given by

$$f_{p,1}(X) = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(X-\mu)'A(X-\mu)}, \quad A > O. \quad (4.2.5)$$

Example 4.2.1. Write down the exponent and the normalizing constant explicitly in a real matrix-variate Gaussian density where

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}, \quad E[X] = M = \begin{bmatrix} 1 & 0 & -1 \\ -1 & -2 & 0 \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix},$$

where the x_{ij} 's are real scalar random variables.

Solution 4.2.1. Note that $A = A'$ and $B = B'$, the leading minors in A being $|(1)| = 1 > 0$ and $|A| = 1 > 0$ so that $A > O$. The leading minors in B are $|(1)| = 1 > 0$, $\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1 > 0$ and

$$|B| = (1) \begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} - (1) \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} + (1) \begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} = 2 > 0,$$

and hence $B > O$. The density is of the form

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2} \text{tr}(A(X-M)B(X-M)')}$$

where the normalizing constant is $\frac{(1)^{\frac{3}{2}}(2)^{\frac{2}{2}}}{(2\pi)^{\frac{(2)(3)}{2}}} = \frac{2}{(2\pi)^3} = \frac{1}{4\pi^3}$. Let X_1 and X_2 be the two rows

of X and let $Y = X - M = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$. Then $Y_1 = (y_{11}, y_{12}, y_{13}) = (x_{11} - 1, x_{12}, x_{13} + 1)$, $Y_2 = (y_{21}, y_{22}, y_{23}) = (x_{21} + 1, x_{22} + 2, x_{23})$. Now

$$(X - M)B(X - M)' = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} B \begin{bmatrix} Y_1' & Y_2' \end{bmatrix} = \begin{bmatrix} Y_1 B Y_1' & Y_1 B Y_2' \\ Y_2 B Y_1' & Y_2 B Y_2' \end{bmatrix},$$

$$A(X - M)B(X - M)' = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} Y_1 B Y_1' & Y_1 B Y_2' \\ Y_2 B Y_1' & Y_2 B Y_2' \end{bmatrix}$$

$$= \begin{bmatrix} Y_1 B Y_1' + Y_2 B Y_1' & Y_1 B Y_2' + Y_2 B Y_2' \\ Y_1 B Y_1' + 2Y_2 B Y_1' & Y_1 B Y_2' + 2Y_2 B Y_2' \end{bmatrix}.$$

Thus,

$$\begin{aligned}\text{tr}[A(X - M)B(X - M)'] &= Y_1BY_1' + Y_2BY_1' + Y_1BY_2' + 2Y_2BY_2' \\ &= Y_1BY_1' + 2Y_1BY_2' + 2Y_2BY_2', \equiv Q,\end{aligned}\quad (i)$$

noting that Y_1BY_2' and Y_2BY_1' are equal since both are real scalar quantities and one is the transpose of the other. Here are now the detailed computations of the various items:

$$Y_1BY_1' = y_{11}^2 + 2y_{11}y_{12} + 2y_{11}y_{13} + 2y_{12}^2 + 2y_{12}y_{13} + 3y_{13}^2 \quad (ii)$$

$$Y_2BY_2' = y_{21}^2 + 2y_{21}y_{22} + 2y_{21}y_{23} + 2y_{22}^2 + 2y_{22}y_{23} + 3y_{23}^2 \quad (iii)$$

$$\begin{aligned}Y_1BY_2' &= y_{11}y_{21} + y_{11}y_{22} + y_{11}y_{23} + y_{12}y_{21} + 2y_{12}y_{22} + y_{12}y_{23} \\ &\quad + y_{13}y_{21} + y_{13}y_{22} + 3y_{13}y_{23}\end{aligned}\quad (iv)$$

where the y_{1j} 's and y_{2j} 's and the various quadratic and bilinear forms are as specified above. The density is then

$$f_{2,3}(X) = \frac{1}{4\pi^3} e^{-\frac{1}{2}(Y_1BY_1' + 2Y_1BY_2' + Y_2BY_2')}$$

where the terms in the exponent are given in (ii)-(iv). This completes the computations.

4.2a. The Matrix-variate Gaussian Density, Complex Case

In the following discussion, the absolute value of a determinant will be denoted by $|\det(A)|$ where A is a square matrix. For example, if $\det(A) = a + ib$ with a and b real scalar and $i = \sqrt{-1}$, the determinant of the conjugate transpose of A is $\det(A^*) = a - ib$. Then the absolute value of the determinant is

$$|\det(A)| = +\sqrt{(a^2 + b^2)} = +[(a+ib)(a-ib)]^{\frac{1}{2}} = +[\det(A)\det(A^*)]^{\frac{1}{2}} = +[\det(AA^*)]^{\frac{1}{2}}. \quad (4.2a.1)$$

The matrix-variate Gaussian density in the complex case, which is the counterpart to that given in (4.2.1) for the real case, is

$$\tilde{f}_{p,q}(\tilde{X}) = \frac{|\det(A)|^q |\det(B)|^p}{\pi^{pq}} e^{-\text{tr}(A\tilde{X}B\tilde{X}^*)} \quad (4.2a.2)$$

for $A > O$, $B > O$, $\tilde{X} = (\tilde{x}_{ij})$, $|\cdot|$ denoting the absolute value of (\cdot) . When $p = 1$ and $A = 1$, the usual multivariate Gaussian density in the complex domain is obtained:

$$\tilde{f}_{1,q}(\tilde{X}) = \frac{|\det(B)|}{\pi^q} e^{-(\tilde{X}-\mu)B(\tilde{X}-\mu)^*}, \quad \tilde{X}' \sim \tilde{N}_q(\tilde{\mu}', B^{-1}) \quad (4.2a.3)$$

where $B > O$ and \tilde{X} and μ are $1 \times q$ row vectors, μ being a location parameter vector. When $q = 1$ in (4.2a.1), we have the p -variate Gaussian or normal density in the complex case which is given by

$$\tilde{f}_{p,1}(\tilde{X}) = \frac{|\det(A)|}{\pi^p} e^{-(\tilde{X}-\mu)^* A (\tilde{X}-\mu)}, \quad \tilde{X} \sim \tilde{N}_p(\mu, A^{-1}) \quad (4.2a.4)$$

where \tilde{X} and the location parameter also denoted by μ are now $p \times 1$ vectors.

Example 4.2a.1. Consider a 2×3 complex matrix-variate Gaussian density. Write down the normalizing constant and the exponent explicitly if

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix}, \quad E[\tilde{X}] = \tilde{M} = \begin{bmatrix} i & -i & 1+i \\ 0 & 1-i & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 1+i & i \\ 1-i & 2 & 1-i \\ -i & 1+i & 3 \end{bmatrix},$$

where the \tilde{x}_{ij} 's are scalar complex random variables.

Solution 4.2a.1. Let us verify the definiteness of A and B . It is obvious that $A = A^*$, $B = B^*$ and hence they are Hermitian. The leading minors of A are $|(3)| = 3 > 0$, $|A| = 4 > 0$ and hence $A > O$. The leading minors of B are $|(4)| = 4 > 0$, $\begin{vmatrix} 4 & 1+i \\ 1-i & 2 \end{vmatrix} = 6 > 0$,

$$|B| = 4 \begin{vmatrix} 2 & 1-i \\ 1+i & 3 \end{vmatrix} - (1+i) \begin{vmatrix} 1-i & 1-i \\ -i & 3 \end{vmatrix} + i \begin{vmatrix} 1-i & 2 \\ -i & 1+i \end{vmatrix} = 8 > 0,$$

and hence $B > O$. The normalizing constant is then

$$\frac{|\det(A)|^q |\det(B)|^p}{\pi^{pq}} = \frac{(4^3)(8^2)}{\pi^6}.$$

Let the two rows of \tilde{X} be \tilde{X}_1 and \tilde{X}_2 . Let $(\tilde{X} - \tilde{M}) = \tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{bmatrix}$,

$$\tilde{Y}_1 = (\tilde{y}_{11}, \tilde{y}_{12}, \tilde{y}_{13}) = (\tilde{x}_{11} - i, \tilde{x}_{12} + i, \tilde{x}_{13} - (1+i))$$

$$\tilde{Y}_2 = (\tilde{y}_{21}, \tilde{y}_{22}, \tilde{y}_{23}) = (\tilde{x}_{21}, \tilde{x}_{22} - (1-i), \tilde{x}_{23} - 1).$$

$$\begin{aligned}
(\tilde{X} - \tilde{M})B(\tilde{X} - \tilde{M})^* &= \tilde{Y}B\tilde{Y}^* = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{bmatrix} B[\tilde{Y}_1^*, \tilde{Y}_2^*] \\
&= \begin{bmatrix} \tilde{Y}_1 B \tilde{Y}_1^* & \tilde{Y}_1 B \tilde{Y}_2^* \\ \tilde{Y}_2 B \tilde{Y}_1^* & \tilde{Y}_2 B \tilde{Y}_2^* \end{bmatrix}.
\end{aligned}$$

Then,

$$\begin{aligned}
\text{tr}[A(\tilde{X} - \tilde{M})B(\tilde{X} - \tilde{M})^*] &= \text{tr} \left\{ \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} \begin{bmatrix} \tilde{Y}_1 B \tilde{Y}_1^* & \tilde{Y}_1 B \tilde{Y}_2^* \\ \tilde{Y}_2 B \tilde{Y}_1^* & \tilde{Y}_2 B \tilde{Y}_2^* \end{bmatrix} \right\} \\
&= 3\tilde{Y}_1 B \tilde{Y}_1^* + (1+i)(\tilde{Y}_2 B \tilde{Y}_1^*) + (1-i)(\tilde{Y}_1 B \tilde{Y}_2^*) + 2\tilde{Y}_2 B \tilde{Y}_2^* \\
&\equiv Q \tag{i}
\end{aligned}$$

where

$$\begin{aligned}
\tilde{Y}_1 B \tilde{Y}_1^* &= 4\tilde{y}_{11}\tilde{y}_{11}^* + 2\tilde{y}_{12}\tilde{y}_{12}^* + 3\tilde{y}_{13}\tilde{y}_{13}^* \\
&\quad + (1+i)\tilde{y}_{11}\tilde{y}_{12}^* + i\tilde{y}_{11}\tilde{y}_{13}^* + (1-i)\tilde{y}_{12}\tilde{y}_{11}^* \\
&\quad + (1-i)\tilde{y}_{12}\tilde{y}_{13}^* - i\tilde{y}_{13}\tilde{y}_{11}^* + (1+i)\tilde{y}_{13}\tilde{y}_{12}^* \tag{ii}
\end{aligned}$$

$$\begin{aligned}
\tilde{Y}_2 B \tilde{Y}_2^* &= 4\tilde{y}_{21}\tilde{y}_{21}^* + 2\tilde{y}_{22}\tilde{y}_{22}^* + 3\tilde{y}_{23}\tilde{y}_{23}^* \\
&\quad + (1+i)\tilde{y}_{21}\tilde{y}_{22}^* + i\tilde{y}_{21}\tilde{y}_{23}^* + (1-i)\tilde{y}_{22}\tilde{y}_{21}^* \\
&\quad + (1-i)\tilde{y}_{22}\tilde{y}_{23}^* - i\tilde{y}_{23}\tilde{y}_{21}^* + (1+i)\tilde{y}_{23}\tilde{y}_{22}^* \tag{iii}
\end{aligned}$$

$$\begin{aligned}
\tilde{Y}_1 B \tilde{Y}_2^* &= 4\tilde{y}_{11}\tilde{y}_{21}^* + 2\tilde{y}_{12}\tilde{y}_{22}^* + 3\tilde{y}_{13}\tilde{y}_{23}^* \\
&\quad + (1+i)\tilde{y}_{11}\tilde{y}_{22}^* + i\tilde{y}_{11}\tilde{y}_{23}^* + (1-i)\tilde{y}_{12}\tilde{y}_{21}^* \\
&\quad + (1-i)\tilde{y}_{12}\tilde{y}_{23}^* - i\tilde{y}_{13}\tilde{y}_{21}^* + (1+i)\tilde{y}_{13}\tilde{y}_{22}^* \tag{iv}
\end{aligned}$$

$$\tilde{Y}_2 B \tilde{Y}_1^* = (iv) \text{ with } \tilde{y}_{1j} \text{ and } \tilde{y}_{2j} \text{ interchanged.} \tag{v}$$

Hence, the density of \tilde{X} is given by

$$\tilde{f}_{2,3}(\tilde{X}) = \frac{(4^3)(8^2)}{\pi^6} e^{-Q}$$

where Q is given explicitly in (i)-(v) above. This completes the computations.

4.2.1. Some properties of a real matrix-variate Gaussian density

In order to derive certain properties, we will need some more Jacobians of matrix transformations, in addition to those provided in Chap. 1. These will be listed in this section as basic results without proofs. The derivations as well as other related Jacobians are available from Mathai (1997).

Theorem 4.2.1. *Let X be a $p \times q$, $q \geq p$, real matrix of rank p , that is, X has full rank, where the pq elements of X are distinct real scalar variables. Let $X = TU_1$ where T is a $p \times p$ real lower triangular matrix whose diagonal elements are positive and U_1 is a semi-orthonormal matrix such that $U_1U_1' = I_p$. Then*

$$dX = \left\{ \prod_{j=1}^p t_{jj}^{q-j} \right\} dT h(U_1) \quad (4.2.6)$$

where $h(U_1)$ is the differential element corresponding to U_1 .

Theorem 4.2.2. *For the differential elements $h(U_1)$ in (4.2.6), the integral is over the Stiefel manifold $V_{p,q}$ or over the space of $p \times q$, $q \geq p$, semi-orthonormal matrices and the integral over the full orthogonal group O_p when $q = p$ are respectively*

$$\int_{V_{p,q}} h(U_1) = \frac{2^p \pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} \text{ and } \int_{O_p} h(U_1) = \frac{2^p \pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \quad (4.2.7)$$

where $\Gamma_p(\alpha)$ is the real matrix-variate gamma function given by

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha) \Gamma(\alpha - 1/2) \cdots \Gamma(\alpha - (p-1)/2), \quad \Re(\alpha) > \frac{p-1}{2}, \quad (4.2.8)$$

$\Re(\cdot)$ denoting the real part of (\cdot) .

For example,

$$\Gamma_3(\alpha) = \pi^{\frac{3(2)}{4}} \Gamma(\alpha) \Gamma(\alpha - 1/2) \Gamma(\alpha - 1) = \pi^{\frac{3}{2}} \Gamma(\alpha) \Gamma(\alpha - 1/2) \Gamma(\alpha - 1), \quad \Re(\alpha) > 1.$$

With the help of Theorems 4.2.1, 4.2.2 and 1.6.7 of Chap. 1, we can derive the following result:

Theorem 4.2.3. *Let X be a real $p \times q$, $q \geq p$, matrix of rank p and $S = XX'$. Then, $S > O$ (real positive definite) and*

$$dX = \frac{\pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} |S|^{\frac{q}{2} - \frac{p+1}{2}} dS, \quad (4.2.9)$$

after integrating out over the Stiefel manifold.

4.2a.1. Some properties of a complex matrix-variate Gaussian density

The corresponding results in the complex domain follow.

Theorem 4.2a.1. *Let \tilde{X} be a $p \times q$, $q \geq p$, matrix of rank p in the complex domain and \tilde{T} be a $p \times p$ lower triangular matrix in the complex domain whose diagonal elements $t_{jj} > 0$, $j = 1, \dots, p$, are real and positive. Then, letting \tilde{U}_1 be a semi-unitary matrix such that $\tilde{U}_1 \tilde{U}_1^* = I_p$,*

$$\tilde{X} = \tilde{T} \tilde{U}_1 \Rightarrow d\tilde{X} = \left\{ \prod_{j=1}^p t_{jj}^{2(q-j)+1} \right\} d\tilde{T} \tilde{h}(\tilde{U}_1) \quad (4.2a.5)$$

where $\tilde{h}(\tilde{U}_1)$ is the differential element corresponding to \tilde{U}_1 .

When integrating out $\tilde{h}(\tilde{U}_1)$, there are three situations to be considered. One of the cases is $q > p$. When $q = p$, the integration is done over the full unitary group \tilde{O}_p ; however, there are two cases to be considered in this instance. One case occurs where all the elements of the unitary matrix \tilde{U}_1 , including the diagonal ones, are complex, in which case \tilde{O}_p will be denoted by $\tilde{O}_p^{(1)}$, and the other one, wherein the diagonal elements of \tilde{U}_1 are real, in which instance the unitary group will be denoted by $\tilde{O}_p^{(2)}$. When unitary transformations are applied to Hermitian matrices, this is our usual situations when Hermitian matrices are involved, then the diagonal elements of the unique \tilde{U}_1 are real and hence the unitary group is $\tilde{O}_p^{(2)}$. The integral of $\tilde{h}(\tilde{U}_1)$ under these three cases are given in the next theorem.

Theorem 4.2a.2. *Let $\tilde{h}(\tilde{U}_1)$ be as defined in equation (4.2a.5). Then, the integral of $\tilde{h}(\tilde{U}_1)$, over the Stiefel manifold $\tilde{V}_{p,q}$ of semi-unitary matrices for $q > p$, and when $q = p$, the integrals over the unitary groups $\tilde{O}_p^{(1)}$ and $\tilde{O}_p^{(2)}$ are the following:*

$$\begin{aligned} \int_{\tilde{V}_{p,q}} \tilde{h}(\tilde{U}_1) &= \frac{2^p \pi^{pq}}{\tilde{\Gamma}_p(q)}, \quad q > p; \\ \int_{\tilde{O}_p^{(1)}} \tilde{h}(\tilde{U}_1) &= \frac{2^p \pi^{p^2}}{\tilde{\Gamma}_p(p)}, \quad \int_{\tilde{O}_p^{(2)}} \tilde{h}(\tilde{U}_1) = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)}, \end{aligned} \quad (4.2a.6)$$

the factor 2^p being omitted when \tilde{U}_1 is uniquely specified; $\tilde{O}_p^{(1)}$ is the case of a general \tilde{X} , $\tilde{O}_p^{(2)}$ is the case corresponding to \tilde{X} Hermitian, and $\tilde{\Gamma}_p(\alpha)$ is the complex matrix-variate gamma, given by

$$\tilde{\Gamma}_p(\alpha) = \pi^{\frac{p(p-1)}{2}} \Gamma(\alpha) \Gamma(\alpha - 1) \cdots \Gamma(\alpha - p + 1), \quad \Re(\alpha) > p - 1. \quad (4.2a.7)$$

For example,

$$\tilde{\Gamma}_3(\alpha) = \pi^{\frac{3(2)}{2}} \Gamma(\alpha) \Gamma(\alpha - 1) \Gamma(\alpha - 2) = \pi^3 \Gamma(\alpha) \Gamma(\alpha - 1) \Gamma(\alpha - 2), \Re(\alpha) > 2.$$

Theorem 4.2a.3. *Let \tilde{X} be $p \times q$, $q \geq p$, matrix of rank p in the complex domain and $\tilde{S} = \tilde{X}\tilde{X}^* > O$. Then after integrating out over the Stiefel manifold,*

$$d\tilde{X} = \frac{\pi^{pq}}{\tilde{\Gamma}_p(q)} |\det(\tilde{S})|^{q-p} d\tilde{S}. \quad (4.2a.8)$$

4.2.2. Additional properties in the real and complex cases

On making use of the above results, we will establish a few results in this section as well as additional ones later on. Let us consider the matrix-variate Gaussian densities corresponding to (4.2.1) and (4.2a.2) with location matrices M and \tilde{M} , respectively, and let the densities be again denoted by $f_{p,q}(X)$ and $\tilde{f}_{p,q}(\tilde{X})$ respectively, where

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2} \text{tr}[A(X-M)B(X-M)']} \quad (4.2.10)$$

and

$$\tilde{f}_{p,q}(\tilde{X}) = \frac{|\det(A)|^q |\det(B)|^p}{\pi^{pq}} e^{-\text{tr}[A(\tilde{X}-\tilde{M})B(\tilde{X}-\tilde{M})^*]}. \quad (4.2a.9)$$

Then, in the real case the expected value of X or the mean value of X , denoted by $E(X)$, is given by

$$E(X) = \int_X X f_{p,q}(X) dX = \int_X (X - M) f_{p,q}(X) dX + M \int_X f_{p,q}(X) dX. \quad (i)$$

The second integral in (i) is the total integral in a density, which is 1, and hence the second integral gives M . On making the transformation $Y = A^{\frac{1}{2}}(X - M)B^{\frac{1}{2}}$, we have

$$E[X] = M + A^{-\frac{1}{2}} \frac{1}{(2\pi)^{\frac{np}{2}}} \int_Y Y e^{-\frac{1}{2} \text{tr}(YY')} dY B^{-\frac{1}{2}}. \quad (ii)$$

But $\text{tr}(YY')$ is the sum of squares of all elements in Y . Hence $Y e^{-\frac{1}{2} \text{tr}(YY')}$ is an odd function and the integral over each element in Y is convergent, so that each integral is zero. Thus, the integral over Y gives a null matrix. Therefore $E(X) = M$. It can be shown in a similar manner that $E(\tilde{X}) = \tilde{M}$.

Theorem 4.2.4, 4.2a.4. For the densities specified in (4.2.10) and (4.2a.9),

$$E(X) = M \text{ and } E(\tilde{X}) = \tilde{M}. \quad (4.2.11)$$

Theorem 4.2.5, 4.2a.5. For the densities given in (4.2.10), (4.2a.9)

$$E[(X - M)B(X - M)'] = qA^{-1}, \quad E[(X - M)'A(X - M)] = pB^{-1} \quad (4.2.12)$$

and

$$E[(\tilde{X} - \tilde{M})B(\tilde{X} - \tilde{M})^*] = qA^{-1}, \quad E[(\tilde{X} - \tilde{M})^*A(\tilde{X} - \tilde{M})] = pB^{-1}. \quad (4.2a.10)$$

Proof: Consider the real case first. Let $Y = A^{\frac{1}{2}}(X - M)B^{\frac{1}{2}} \Rightarrow A^{-\frac{1}{2}}Y = (X - M)B^{\frac{1}{2}}$. Then

$$E[(X - M)B(X - M)'] = \frac{A^{-\frac{1}{2}}}{(2\pi)^{\frac{pq}{2}}} \int_Y YY' e^{-\frac{1}{2}\text{tr}(YY')} dY A^{-\frac{1}{2}}. \quad (i)$$

Note that Y is $p \times q$ and YY' is $p \times p$. The non-diagonal elements in YY' are dot products of the distinct row vectors in Y and hence linear functions of the elements of Y . The diagonal elements in YY' are sums of squares of elements in the rows of Y . The exponent has all sum of squares and hence the convergent integrals corresponding to all the non-diagonal elements in YY' are zeros. Hence, only the diagonal elements need be considered. Each diagonal element is a sum of squares of q elements of Y . For example, the first diagonal element in YY' is $y_{11}^2 + y_{12}^2 + \cdots + y_{1q}^2$ where $Y = (y_{ij})$. Let $Y_1 = (y_{11}, \dots, y_{1q})$ be the first row of Y and let $s = Y_1 Y_1' = y_{11}^2 + \cdots + y_{1q}^2$. It follows from Theorem 4.2.3 that when $p = 1$,

$$dY_1 = \frac{\pi^{\frac{q}{2}}}{\Gamma(\frac{q}{2})} s^{\frac{q}{2}-1} ds. \quad (ii)$$

Then

$$\int_{Y_1} Y_1 Y_1' e^{-\frac{1}{2}Y_1 Y_1'} dY_1 = \int_{s=0}^{\infty} s \frac{\pi^{\frac{q}{2}}}{\Gamma(\frac{q}{2})} s^{\frac{q}{2}-1} e^{-\frac{1}{2}s} ds. \quad (iii)$$

The integral part over s is $2^{\frac{q}{2}+1} \Gamma(\frac{q}{2} + 1) = 2^{\frac{q}{2}+1} \frac{q}{2} \Gamma(\frac{q}{2}) = 2^{\frac{q}{2}} q \Gamma(\frac{q}{2})$. Thus $\Gamma(\frac{q}{2})$ is canceled and $(2\pi)^{\frac{q}{2}}$ cancels with $(2\pi)^{\frac{pq}{2}}$ leaving $(2\pi)^{\frac{(p-1)q}{2}}$ in the denominator and q in the numerator. We still have $p - 1$ such sets of q, y_{ij}^2 's in the exponent in (i) and each such

integrals is of the form $\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \sqrt{(2\pi)}$ which gives $(2\pi)^{\frac{(p-1)q}{2}}$ and thus the factor containing π is also canceled leaving only q at each diagonal position in YY' . Hence the integral $\frac{1}{(2\pi)^{\frac{pq}{2}}} \int_Y YY' e^{-\frac{1}{2}\text{tr}(YY')} dY = qI$ where I is the identity matrix, which establishes one of the results in (4.2.12). Now, write

$$\text{tr}[A(X - M)B(X - M)'] = \text{tr}[(X - M)'A(X - M)B] = \text{tr}[B(X - M)'A(X - M)].$$

This is the same structure as in the previous case where B occupies the place of A and the order is now q in place of p in the previous case. Then, proceeding as in the derivations from (i) to (iii), the second result in (4.2.12) follows. The results in (4.2a.10) are established in a similar manner.

From (4.2.10), it is clear that the density of Y , denoted by $g(Y)$, is of the form

$$g(Y) = \frac{1}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\text{tr}(YY')}, \quad Y = (y_{ij}), \quad -\infty < y_{ij} < \infty, \quad (4.2.13)$$

for all i and j . The individual y_{ij} 's are independently distributed and each y_{ij} has the density

$$g_{ij}(y_{ij}) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y_{ij}^2}, \quad -\infty < y_{ij} < \infty. \quad (iv)$$

Thus, we have a real standard normal density for y_{ij} . The complex case corresponding to (4.2.13), denoted by $\tilde{g}(\tilde{Y})$, is given by

$$\tilde{g}(\tilde{Y}) = \frac{1}{\pi^{pq}} e^{-\text{tr}(\tilde{Y}\tilde{Y}^*)}. \quad (4.2a.11)$$

In this case, the exponent is $\text{tr}(\tilde{Y}\tilde{Y}^*) = \sum_{i=1}^p \sum_{j=1}^q |\tilde{y}_{ij}|^2$ where $\tilde{y}_{rs} = y_{rs1} + iy_{rs2}$, y_{rs1} , y_{rs2} real, $i = \sqrt{-1}$ and $|\tilde{y}_{rs}|^2 = y_{rs1}^2 + y_{rs2}^2$.

For the real case, consider the probability that $y_{ij} \leq t_{ij}$ for some given t_{ij} and this is the distribution function of y_{ij} , which is denoted by $F_{y_{ij}}(t_{ij})$. Then, let us compute the density of y_{ij}^2 . Consider the probability that $y_{ij}^2 \leq u$, $u > 0$ for some u . Let $u_{ij} = y_{ij}^2$. Then, $\text{Pr}\{u_{ij} \leq v_{ij}\}$ for some v_{ij} is the distribution function of u_{ij} evaluated at v_{ij} , denoted by $F_{u_{ij}}(v_{ij})$. Consider

$$\text{Pr}\{y_{ij}^2 \leq t, t > 0\} = \text{Pr}\{|y_{ij}| \leq \sqrt{t}\} = \text{Pr}\{-\sqrt{t} \leq y_{ij} \leq \sqrt{t}\} = F_{y_{ij}}(\sqrt{t}) - F_{y_{ij}}(-\sqrt{t}). \quad (v)$$

Differentiate throughout with respect to t . When $\text{Pr}\{y_{ij}^2 \leq t\}$ is differentiated with respect to t , we obtain the density of $u_{ij} = y_{ij}^2$, evaluated at t . This density, denoted by $h_{ij}(u_{ij})$, is given by

$$\begin{aligned}
h_{ij}(u_{ij})|_{u_{ij}=t} &= \frac{d}{dt} F_{y_{ij}}(\sqrt{t}) - \frac{d}{dt} F(-\sqrt{t}) \\
&= g_{ij}(y_{ij} = t) \frac{1}{2} t^{\frac{1}{2}-1} - g_{ij}(y_{ij} = t) (-\frac{1}{2} t^{\frac{1}{2}-1}) \\
&= \frac{1}{\sqrt{(2\pi)}} [t^{\frac{1}{2}-1} e^{-\frac{1}{2}t}] = \frac{1}{\sqrt{(2\pi)}} [u_{ij}^{\frac{1}{2}-1} e^{-\frac{1}{2}u_{ij}}] \quad (vi)
\end{aligned}$$

evaluated at $u_{ij} = t$ for $0 \leq t < \infty$. Hence we have the following result:

Theorem 4.2.6. Consider the density $f_{p,q}(X)$ in (4.2.1) and the transformation $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$. Letting $Y = (y_{ij})$, the y_{ij} 's are mutually independently distributed as in (iv) above and each y_{ij}^2 is distributed as a real chi-square random variable having one degree of freedom or equivalently a real gamma with parameters $\alpha = \frac{1}{2}$ and $\beta = 2$ where the usual real scalar gamma density is given by

$$f(z) = \frac{1}{\beta^\alpha \Gamma(\alpha)} z^{\alpha-1} e^{-\frac{z}{\beta}}, \quad (vii)$$

for $0 \leq z < \infty$, $\Re(\alpha) > 0$, $\Re(\beta) > 0$ and $f(z) = 0$ elsewhere.

As a consequence of the y_{ij}^2 's being independently gamma distributed, $\sum_{j=1}^q y_{ij}^2$ is real gamma distributed with the parameters $\alpha = \frac{q}{2}$ and $\beta = 2$. Then $\text{tr}(YY')$ is real gamma distributed with the parameters $\alpha = \frac{pq}{2}$ and $\beta = 2$ and each diagonal element in YY' is real gamma distributed with parameters $\frac{q}{2}$ and $\beta = 2$ or a real chi-square variable with q degrees of freedom and an expected value $2\frac{q}{2} = q$. This is an alternative way of proving (4.2.12). Proofs for the other results in (4.2.12) and (4.2a.10) are parallel and hence are omitted.

4.2.3. Some special cases

Consider the real $p \times q$ matrix-variate Gaussian case where the exponent in the density is $-\frac{1}{2}\text{tr}(AXBX')$. On making the transformation $A^{\frac{1}{2}}X = Z \Rightarrow dZ = |A|^{\frac{q}{2}}dX$, Z has a $p \times q$ matrix-variate Gaussian density of the form

$$f_{p,q}(Z) = \frac{|B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\text{tr}(ZBZ')}. \quad (4.2.14)$$

If the distribution has a $p \times q$ constant matrix M as location parameter, then replace Z by $Z - M$ in (4.2.14), which does not affect the normalizing constant. Letting Z_1, Z_2, \dots, Z_p denote the rows of Z , we observe that Z_j has a q -variate multinormal distribution with the

null vector as its mean value and B^{-1} as its covariance matrix for each $j = 1, \dots, p$. This can be seen from the considerations that follow. Let us consider the transformation $Y = ZB^{\frac{1}{2}} \Rightarrow dZ = |B|^{-\frac{p}{2}} dY$. The density in (4.2.14) then reduces to the following, denoted by $f_{p,q}(Y)$:

$$f_{p,q}(Y) = \frac{1}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\text{tr}(YY')}. \quad (4.2.15)$$

This means that each element y_{ij} in $Y = (y_{ij})$ is a real univariate standard normal variable, $y_{ij} \sim N_1(0, 1)$ as per the usual notation, and all the y_{ij} 's are mutually independently distributed. Letting the p rows of Y be Y_1, \dots, Y_p , then each Y_j is a q -variate standard normal vector for $j = 1, \dots, p$. Letting the density of Y_j be denoted by $f_{Y_j}(Y_j)$, we have

$$f_{Y_j}(Y_j) = \frac{1}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(Y_j Y_j')}.$$

Now, consider the transformation $Z_j = Y_j B^{-\frac{1}{2}} \Rightarrow dY_j = |B|^{\frac{1}{2}} dZ_j$ and $Y_j = Z_j B^{\frac{1}{2}}$. That is, $Y_j Y_j' = Z_j B Z_j'$ and the density of Z_j denoted by $f_{Z_j}(Z_j)$ is as follows:

$$f_{Z_j}(Z_j) = \frac{|B|^{\frac{1}{2}}}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}(Z_j B Z_j')}, \quad B > O, \quad (4.2.16)$$

which is a q -variate real multinormal density with the covariance matrix of Z_j given by B^{-1} , for each $j = 1, \dots, p$, and the Z_j 's, $j = 1, \dots, p$, are mutually independently distributed. Thus, the following result:

Theorem 4.2.7. *Let Z_1, \dots, Z_p be the p rows of the $p \times q$ matrix Z in (4.2.14). Then each Z_j has a q -variate real multinormal distribution with the covariance matrix B^{-1} , for $j = 1, \dots, p$, and Z_1, \dots, Z_p are mutually independently distributed.*

Observe that the exponent in the original real $p \times q$ matrix-variate Gaussian density can also be rewritten in the following format:

$$\begin{aligned} -\frac{1}{2}\text{tr}(AXBX') &= -\frac{1}{2}\text{tr}(X'AXB) = -\frac{1}{2}\text{tr}(BX'AX) \\ &= -\frac{1}{2}\text{tr}(U'AU) = -\frac{1}{2}\text{tr}(ZBZ'), \quad A^{\frac{1}{2}}X = Z, \quad XB^{\frac{1}{2}} = U. \end{aligned}$$

Now, on making the transformation $U = XB^{\frac{1}{2}} \Rightarrow dX = |B|^{-\frac{p}{2}} dU$, the density of U , denoted by $f_{p,q}(U)$, is given by

$$f_{p,q}(U) = \frac{|A|^{\frac{q}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\text{tr}(U'AU)}. \quad (4.2.17)$$

Proceeding as in the derivation of Theorem 4.2.7, we have the following result:

Theorem 4.2.8. Consider the $p \times q$ real matrix U in (4.2.17). Let U_1, \dots, U_q be the columns of U . Then, U_1, \dots, U_q are mutually independently distributed with U_j having a p -variate multinormal density, denoted by $f_{U_j}(U_j)$, given as

$$f_{U_j}(U_j) = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(U_j' A U_j)}. \quad (4.2.18)$$

The corresponding results in the $p \times q$ complex Gaussian case are the following:

Theorem 4.2a.6. Consider the $p \times q$ complex Gaussian matrix \tilde{X} . Let $A^{\frac{1}{2}} \tilde{X} = \tilde{Z}$ and $\tilde{Z}_1, \dots, \tilde{Z}_p$ be the rows of \tilde{Z} . Then, $\tilde{Z}_1, \dots, \tilde{Z}_p$ are mutually independently distributed with \tilde{Z}_j having a q -variate complex multinormal density, denoted by $\tilde{f}_{\tilde{Z}_j}(\tilde{Z}_j)$, given by

$$\tilde{f}_{\tilde{Z}_j}(\tilde{Z}_j) = \frac{|\det(B)|}{\pi^q} e^{-(\tilde{Z}_j B \tilde{Z}_j^*)}. \quad (4.2a.12)$$

Theorem 4.2a.7. Let the $p \times q$ matrix \tilde{X} have a complex matrix-variate distribution. Let $\tilde{U} = \tilde{X} B^{\frac{1}{2}}$ and $\tilde{U}_1, \dots, \tilde{U}_q$ be the columns of \tilde{U} . Then $\tilde{U}_1, \dots, \tilde{U}_q$ are mutually independently distributed as p -variate complex multinormal with covariance matrix A^{-1} each, the density of \tilde{U}_j , denoted by $\tilde{f}_{\tilde{U}_j}(\tilde{U}_j)$, being given as

$$\tilde{f}_{\tilde{U}_j}(\tilde{U}_j) = \frac{|\det(A)|}{\pi^p} e^{-(\tilde{U}_j^* A \tilde{U}_j)}. \quad (4.2a.13)$$

Exercises 4.2

- 4.2.1. Prove the second result in equation (4.2.12) and prove both results in (4.2a.10).
- 4.2.2. Obtain (4.2.12) by establishing first the distribution of the row sum of squares and column sum of squares in Y , and then taking the expected values in those variables.
- 4.2.3. Prove (4.2a.10) by establishing first the distributions of row and column sum of squares of the absolute values in \tilde{Y} and then taking the expected values.
- 4.2.4. Establish 4.2.12 and 4.2a.10 by using the general polar coordinate transformations.
- 4.2.5. First prove that $\sum_{j=1}^q |\tilde{y}_{ij}|^2$ is a $2q$ -variate real gamma random variable. Then establish the results in (4.2a.10) by using the those on real gamma variables, where $\tilde{Y} = (\tilde{y}_{ij})$, the \tilde{y}_{ij} 's in (4.2a.11) being in the complex domain and $|\tilde{y}_{ij}|$ denoting the absolute value or modulus of \tilde{y}_{ij} .

4.2.6. Let the real matrix $A > O$ be 2×2 with its first row being $(1, 1)$ and let $B > O$ be 3×3 with its first row being $(1, 1, -1)$. Then complete the other rows in A and B so that $A > O$, $B > O$. Obtain the corresponding 2×3 real matrix-variate Gaussian density when (1): $M = O$, (2): $M \neq O$ with a matrix M of your own choice.

4.2.7. Let the complex matrix $A > O$ be 2×2 with its first row being $(1, 1 + i)$ and let $B > O$ be 3×3 with its first row being $(1, 1 + i, -i)$. Complete the other rows with numbers in the complex domain of your own choice so that $A = A^* > O$, $B = B^* > O$. Obtain the corresponding 2×3 complex matrix-variate Gaussian density with (1): $\tilde{M} = O$, (2): $\tilde{M} \neq O$ with a matrix \tilde{M} of your own choice.

4.2.8. Evaluate the covariance matrix in (4.2.16), which is $E(Z'_j Z_j)$, and show that it is B^{-1} .

4.2.9. Evaluate the covariance matrix in (4.2.18), which is $E(U_j U'_j)$, and show that it is A^{-1} .

4.2.10. Repeat Exercises 4.2.8 and 4.2.9 for the complex case in (4.2a.12) and (4.2a.13).

4.3. Moment Generating Function and Characteristic Function, Real Case

Let $T = (t_{ij})$ be a $p \times q$ parameter matrix. The matrix random variable $X = (x_{ij})$ is $p \times q$ and it is assumed that all of its elements x_{ij} 's are real and distinct scalar variables. Then

$$\text{tr}(T X') = \sum_{i=1}^p \sum_{j=1}^q t_{ij} x_{ij} = \text{tr}(X' T) = \text{tr}(X T'). \quad (i)$$

Note that each t_{ij} and x_{ij} appear once in (i) and thus, we can define the moment generating function (mgf) in the real matrix-variate case, denoted by $M_f(T)$ or $M_X(T)$, as follows:

$$M_f(T) = E[e^{\text{tr}(T X')}] = \int_X e^{\text{tr}(T X')} f_{p,q}(X) dX = M_X(T) \quad (ii)$$

whenever the integral is convergent, where E denotes the expected value. Thus, for the $p \times q$ matrix-variate real Gaussian density,

$$M_X(T) = M_f(T) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} \int_X e^{\text{tr}(T X') - \frac{1}{2} \text{tr}(A^{\frac{1}{2}} X B X' A^{\frac{1}{2}})} dX$$

where A is $p \times p$, B is $q \times q$ and A and B are constant real positive definite matrices so that $A^{\frac{1}{2}}$ and $B^{\frac{1}{2}}$ are uniquely defined. Consider the transformation $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}} \Rightarrow dY = |A|^{\frac{q}{2}} |B|^{\frac{p}{2}} dX$ by Theorem 1.6.4. Thus, $X = A^{-\frac{1}{2}} Y B^{-\frac{1}{2}}$ and

$$\text{tr}(T X') = \text{tr}(T B^{-\frac{1}{2}} Y' A^{-\frac{1}{2}}) = \text{tr}(A^{-\frac{1}{2}} T B^{-\frac{1}{2}} Y') = \text{tr}(T_{(1)} Y')$$

where $T_{(1)} = A^{-\frac{1}{2}}TB^{-\frac{1}{2}}$. Then

$$M_X(T) = \frac{1}{(2\pi)^{\frac{pq}{2}}} \int_Y e^{\text{tr}(T_{(1)}Y') - \frac{1}{2}\text{tr}(YY')} dY.$$

Note that $T_{(1)}Y'$ and YY' are $p \times p$. Consider $-2\text{tr}(T_{(1)}Y') + \text{tr}(YY')$, which can be written as

$$-2\text{tr}(T_{(1)}Y') + \text{tr}(YY') = -\text{tr}(T_{(1)}T'_{(1)}) + \text{tr}[(Y - T_{(1)})(Y - T_{(1)})'].$$

Therefore

$$\begin{aligned} M_X(T) &= e^{\frac{1}{2}\text{tr}(T_{(1)}T'_{(1)})} \frac{1}{(2\pi)^{\frac{pq}{2}}} \int_Y e^{-\frac{1}{2}\text{tr}[(Y - T_{(1)})(Y - T_{(1)})']} dY \\ &= e^{\frac{1}{2}\text{tr}(T_{(1)}T'_{(1)})} = e^{\frac{1}{2}\text{tr}(A^{-\frac{1}{2}}TB^{-1}T'A^{-\frac{1}{2}})} = e^{\frac{1}{2}\text{tr}(A^{-1}TB^{-1}T')} \end{aligned} \quad (4.3.1)$$

since the integral is 1 from the total integral of a matrix-variate Gaussian density.

In the presence of a location parameter matrix M , the matrix-variate Gaussian density is given by

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}}|B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2}\text{tr}(A^{\frac{1}{2}}(X-M)B(X-M)'A^{\frac{1}{2}})} \quad (4.3.2)$$

where M is a constant $p \times q$ matrix. In this case, $TX' = T(X - M + M)' = T(X - M)' + TM'$, and

$$\begin{aligned} M_X(T) &= M_f(T) = E[e^{\text{tr}(TX')}] = e^{\text{tr}(TM')} E[e^{\text{tr}(T(X-M)')}] \\ &= e^{\text{tr}(TM')} e^{\frac{1}{2}\text{tr}(A^{-1}TB^{-1}T')} = e^{\text{tr}(TM') + \frac{1}{2}\text{tr}(A^{-1}TB^{-1}T')}. \end{aligned} \quad (4.3.3)$$

When $p = 1$, we have the usual q -variate multinormal density. In this case, A is 1×1 and taken to be 1. Then the mgf is given by

$$M_X(T) = e^{TM' + \frac{1}{2}TB^{-1}T'} \quad (4.3.4)$$

where T , M and X are $1 \times q$ and $B > O$ is $q \times q$. The corresponding characteristic function when $p = 1$ is given by

$$\phi(T) = e^{iTM' - \frac{1}{2}TB^{-1}T'}. \quad (4.3.5)$$

Example 4.3.1. Let X have a 2×3 real matrix-variate Gaussian density with the following parameters:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}, \quad E[X] = M = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},$$

$$B = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

Consider the density $f_{2,3}(X)$ with the exponent preceded by $\frac{1}{2}$ to be consistent with p -variate real Gaussian density. Verify whether A and B are positive definite. Then compute the moment generating function (mgf) of X or that associated with $f_{2,3}(X)$ and write down the exponent explicitly.

Solution 4.3.1. Consider a 2×3 parameter matrix $T = (t_{ij})$. Let us compute the various quantities in the mgf. First,

$$TM' = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} t_{11} - t_{13} & -t_{11} + t_{12} \\ t_{21} - t_{23} & -t_{21} + t_{22} \end{bmatrix},$$

so that

$$\text{tr}(TM') = t_{11} - t_{13} - t_{21} + t_{22}. \quad (i)$$

Consider the leading minors in A and B . Note that $|(1)| = 1 > 0$, $|A| = 1 > 0$, $|(3)| = 3 > 0$, $\begin{vmatrix} 3 & -1 \\ -1 & 2 \end{vmatrix} = 5 > 0$, $|B| = 8 > 0$; thus both A and B are positive definite. The inverses of A and B are obtained by making use of the formula $C^{-1} = \frac{1}{|C|}(\text{Cof}(C))'$; they are

$$A^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \quad B^{-1} = \frac{1}{8} \begin{bmatrix} 5 & 4 & -3 \\ 4 & 8 & -4 \\ -3 & -4 & 5 \end{bmatrix}.$$

For determining the exponent in the mgf, we need $A^{-1}T$ and $B^{-1}T'$, which are

$$\begin{aligned} A^{-1}T &= \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \end{bmatrix} \\ &= \begin{bmatrix} 2t_{11} - t_{21} & 2t_{12} - t_{22} & 2t_{13} - t_{23} \\ -t_{11} + t_{21} & -t_{12} + t_{22} & -t_{13} + t_{23} \end{bmatrix} \\ B^{-1}T' &= \frac{1}{8} \begin{bmatrix} 5 & 4 & -3 \\ 4 & 8 & -4 \\ -3 & -4 & 5 \end{bmatrix} \begin{bmatrix} t_{11} & t_{21} \\ t_{12} & t_{22} \\ t_{13} & t_{23} \end{bmatrix} \\ &= \frac{1}{8} \begin{bmatrix} 5t_{11} + 4t_{12} - 3t_{13} & 5t_{21} + 4t_{22} - 3t_{23} \\ 4t_{11} + 8t_{12} - 4t_{13} & 4t_{21} + 8t_{22} - 4t_{23} \\ -3t_{11} - 4t_{12} + 5t_{13} & -3t_{21} - 4t_{22} + 5t_{23} \end{bmatrix}. \end{aligned}$$

Hence,

$$\begin{aligned} \frac{1}{2}\text{tr}[A^{-1}TB^{-1}T'] &= \frac{1}{16}[(2t_{11} - t_{21})(5t_{11} + 4t_{12} - 3t_{13}) \\ &\quad + (2t_{12} - t_{22})(4t_{11} + 8t_{12} - 4t_{13}) + (2t_{13} - t_{23})(-3t_{11} - 4t_{12} + 5t_{13}) \\ &\quad + (-t_{11} + t_{21})(5t_{21} + 4t_{22} - 3t_{23}) + (-t_{12} + t_{22})(4t_{21} + 8t_{22} - 4t_{23}) \\ &\quad + (-t_{13} + t_{23})(-3t_{21} - 4t_{22} + 5t_{23})]. \end{aligned} \quad (ii)$$

Thus, the mgf is $M_X(T) = e^{Q(T)}$ where

$$Q(T) = \text{tr}(TM') + \frac{1}{2}\text{tr}(A^{-1}TB^{-1}T'),$$

these quantities being given in (i) and (ii). This completes the computations.

4.3a. Moment Generating and Characteristic Functions, Complex Case

Let $\tilde{X} = (\tilde{x}_{ij})$ be a $p \times q$ matrix where the \tilde{x}_{ij} 's are distinct scalar complex variables. We may write $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, X_1, X_2 being real $p \times q$ matrices. Let \tilde{T} be a $p \times q$ parameter matrix and $\tilde{T} = T_1 + iT_2$, T_1, T_2 being real $p \times q$ matrices. The conjugate transposes of \tilde{X} and \tilde{T} are denoted by \tilde{X}^* and \tilde{T}^* , respectively. Then,

$$\begin{aligned} \text{tr}(\tilde{T}\tilde{X}^*) &= \text{tr}[(T_1 + iT_2)(X_1' - iX_2')] \\ &= \text{tr}[T_1X_1' + T_2X_2' + i(T_2X_1' - T_1X_2')] \\ &= \text{tr}(T_1X_1') + \text{tr}(T_2X_2') + i \text{tr}(T_2X_1' - T_1X_2'). \end{aligned}$$

If $T_1 = (t_{ij}^{(1)})$, $X_1 = (x_{ij}^{(1)})$, $X_2 = (x_{ij}^{(2)})$, $T_2 = (t_{ij}^{(2)})$, $\text{tr}(T_1 X_1') = \sum_{i=1}^p \sum_{j=1}^q t_{ij}^{(1)} x_{ij}^{(1)}$, $\text{tr}(T_2 X_2') = \sum_{i=1}^p \sum_{j=1}^q t_{ij}^{(2)} x_{ij}^{(2)}$. In other words, $\text{tr}(T_1 X_1') + \text{tr}(T_2 X_2')$ gives all the x_{ij} 's in the real and complex parts of \tilde{X} multiplied by the corresponding t_{ij} 's in the real and complex parts of \tilde{T} . That is, $E[e^{\text{tr}(T_1 X_1') + \text{tr}(T_2 X_2')}]$ gives a moment generating function (mgf) associated with the complex matrix-variate Gaussian density that is consistent with real multivariate mgf. However, $[\text{tr}(T_1 X_1') + \text{tr}(T_2 X_2')] = \Re(\text{tr}[\tilde{T} \tilde{X}^*])$, $\Re(\cdot)$ denoting the real part of (\cdot) . Thus, in the complex case, the mgf for any real-valued scalar function $g(\tilde{X})$ of the complex matrix argument \tilde{X} , where $g(\tilde{X})$ is a density, is defined as

$$\tilde{M}_{\tilde{X}}(\tilde{T}) = \int_{\tilde{X}} e^{\Re[\text{tr}(\tilde{T} \tilde{X}^*)]} g(\tilde{X}) d\tilde{X} \quad (4.3a.1)$$

whenever the expected value exists. On replacing \tilde{T} by $i\tilde{T}$, $i = \sqrt{-1}$, we obtain the characteristic function of \tilde{X} or that associated with \tilde{f} , denoted by $\phi_{\tilde{X}}(\tilde{T}) = \phi_{\tilde{f}}(\tilde{T})$. That is,

$$\phi_{\tilde{X}}(\tilde{T}) = \int_{\tilde{X}} e^{\Re[\text{tr}(i\tilde{T} \tilde{X}^*)]} g(\tilde{X}) d\tilde{X}. \quad (4.3a.2)$$

Then, the mgf of the matrix-variate Gaussian density in the complex domain is available by paralleling the derivation in the real case and making use of Lemma 3.2a.1:

$$\begin{aligned} \tilde{M}_{\tilde{X}}(\tilde{T}) &= E[e^{\Re[\text{tr}(\tilde{T} \tilde{X}^*)]}] \\ &= e^{\Re[\text{tr}(\tilde{T} \tilde{M}^*)] + \frac{1}{4} \Re[\text{tr}(A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}})]}. \end{aligned} \quad (4.3a.3)$$

The corresponding characteristic function is given by

$$\phi_{\tilde{X}}(\tilde{T}) = e^{\Re[\text{tr}(i\tilde{T} \tilde{M}^*)] - \frac{1}{4} \Re[\text{tr}(A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}})]}. \quad (4.3a.4)$$

Note that when $A = A^* > O$ and $B = B^* > O$ (Hermitian positive definite),

$$(A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}})^* = A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}},$$

that is, this matrix is Hermitian. Thus, letting $\tilde{U} = A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}} = U_1 + iU_2$ where U_1 and U_2 are real matrices, $U_1 = U_1'$ and $U_2 = -U_2'$, that is, U_1 and U_2 are respectively symmetric and skew symmetric real matrices. Accordingly, $\text{tr}(\tilde{U}) = \text{tr}(U_1) + i\text{tr}(U_2) = \text{tr}(U_1)$ as the trace of a real skew symmetric matrix is zero. Therefore, $\Re[\text{tr}(A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}})] = \text{tr}(A^{-\frac{1}{2}} \tilde{T} B^{-1} \tilde{T}^* A^{-\frac{1}{2}})$, the diagonal elements of a Hermitian matrix being real.

When $p = 1$, we have the usual q -variate complex multivariate normal density and taking the 1×1 matrix A to be 1, the mgf is as follows:

$$\tilde{M}_{\tilde{X}}(\tilde{T}) = e^{\Re(\tilde{T}\tilde{M}^*) + \frac{1}{4}(\tilde{T}B^{-1}\tilde{T}^*)} \quad (4.3a.5)$$

where \tilde{T} , \tilde{M} are $1 \times q$ vectors and $B = B^* > O$ (Hermitian positive definite), the corresponding characteristic function being given by

$$\phi_{\tilde{X}}(\tilde{T}) = e^{\Re(i\tilde{T}\tilde{M}^*) - \frac{1}{4}(\tilde{T}B^{-1}\tilde{T}^*)}. \quad (4.3a.6)$$

Example 4.3a.1. Consider a 2×2 matrix \tilde{X} in the complex domain having a complex matrix-variate density with the following parameters:

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{21} & \tilde{x}_{22} \end{bmatrix}, \quad E[\tilde{X}] = \tilde{M} = \begin{bmatrix} 1+i & i \\ 2-i & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} 2 & i \\ -i & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix}.$$

Determine whether A and B are Hermitian positive definite; then, obtain the mgf of this distribution and provide the exponential part explicitly.

Solution 4.3a.1. Clearly, $A > O$ and $B > O$. We first determine A^{-1} , B^{-1} , $A^{-1}\tilde{T}$, $B^{-1}\tilde{T}^*$:

$$A^{-1} = \frac{1}{5} \begin{bmatrix} 3 & -i \\ i & 2 \end{bmatrix}, \quad A^{-1}\tilde{T} = \frac{1}{5} \begin{bmatrix} 3 & -i \\ i & 2 \end{bmatrix} \begin{bmatrix} \tilde{t}_{11} & \tilde{t}_{12} \\ \tilde{t}_{21} & \tilde{t}_{22} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3\tilde{t}_{11} - i\tilde{t}_{21} & 3\tilde{t}_{12} - i\tilde{t}_{22} \\ i\tilde{t}_{11} + 2\tilde{t}_{21} & i\tilde{t}_{12} + 2\tilde{t}_{22} \end{bmatrix},$$

$$B^{-1} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}, \quad B^{-1}\tilde{T}^* = \begin{bmatrix} \tilde{t}_{11}^* + i\tilde{t}_{12}^* & \tilde{t}_{21}^* + i\tilde{t}_{22}^* \\ -i\tilde{t}_{11}^* + 2\tilde{t}_{12}^* & -i\tilde{t}_{21}^* + 2\tilde{t}_{22}^* \end{bmatrix}.$$

Letting $\delta = \frac{1}{2}\text{tr}(A^{-1}\tilde{T}B^{-1}\tilde{T}^*)$,

$$10\delta = \{(3\tilde{t}_{11} - i\tilde{t}_{21})(\tilde{t}_{11}^* + i\tilde{t}_{12}^*) + (3\tilde{t}_{12} - i\tilde{t}_{22})(-i\tilde{t}_{11}^* + 2\tilde{t}_{12}^*) \\ + (i\tilde{t}_{11} + 2\tilde{t}_{21})(\tilde{t}_{21}^* + i\tilde{t}_{22}^*) + (i\tilde{t}_{12} + 2\tilde{t}_{22})(-i\tilde{t}_{21}^* + 2\tilde{t}_{22}^*)\},$$

$$10\delta = \{3\tilde{t}_{11}\tilde{t}_{11}^* + 3i\tilde{t}_{11}\tilde{t}_{12}^* - i\tilde{t}_{21}\tilde{t}_{11}^* + \tilde{t}_{21}\tilde{t}_{12}^* \\ + 6\tilde{t}_{12}\tilde{t}_{12}^* - \tilde{t}_{22}\tilde{t}_{11}^* - 2i\tilde{t}_{22}\tilde{t}_{12}^* - 3i\tilde{t}_{12}\tilde{t}_{11}^* \\ + i\tilde{t}_{11}\tilde{t}_{21}^* - \tilde{t}_{11}\tilde{t}_{22}^* + 2\tilde{t}_{21}\tilde{t}_{21}^* + 2i\tilde{t}_{21}\tilde{t}_{22}^* \\ + \tilde{t}_{12}\tilde{t}_{21}^* + 2i\tilde{t}_{12}\tilde{t}_{22}^* - 2i\tilde{t}_{22}\tilde{t}_{21}^* + 4\tilde{t}_{22}\tilde{t}_{22}^*\},$$

$$\begin{aligned}
10\delta &= 3\tilde{t}_{11}\tilde{t}_{11}^* + 6\tilde{t}_{12}\tilde{t}_{12}^* + 2\tilde{t}_{21}\tilde{t}_{21}^* + 4\tilde{t}_{22}\tilde{t}_{22}^* \\
&\quad + 3i[\tilde{t}_{11}\tilde{t}_{12}^* - \tilde{t}_{12}\tilde{t}_{11}^*] - [\tilde{t}_{22}\tilde{t}_{11}^* + \tilde{t}_{11}\tilde{t}_{22}^*] \\
&\quad + i[\tilde{t}_{11}\tilde{t}_{21}^* - \tilde{t}_{11}^*\tilde{t}_{21}] + [\tilde{t}_{12}\tilde{t}_{21}^* + \tilde{t}_{12}^*\tilde{t}_{21}] \\
&\quad + 2i[\tilde{t}_{21}\tilde{t}_{22}^* - \tilde{t}_{21}^*\tilde{t}_{22}] + 2i[\tilde{t}_{12}\tilde{t}_{22}^* - \tilde{t}_{12}^*\tilde{t}_{22}].
\end{aligned}$$

Letting $\tilde{t}_{rs} = t_{rs1} + it_{rs2}$, $i = \sqrt{-1}$, t_{rs1} , t_{rs2} being real, for all r and s , then δ , the exponent in the mgf, can be expressed as follows:

$$\begin{aligned}
\delta &= \frac{1}{10}\{3(t_{111}^2 + t_{112}^2) + 6(t_{121}^2 + t_{122}^2) + 2(t_{211}^2 + t_{212}^2) + 4(t_{221}^2 + t_{222}^2) \\
&\quad - 6(t_{112}t_{121} - t_{111}t_{122}) - 2(t_{111}t_{221} - t_{112}t_{222}) - 2(t_{112}t_{211} - t_{111}t_{212}) \\
&\quad + 2(t_{121}t_{211} + t_{122}t_{212}) - 4(t_{212}t_{221} - t_{211}t_{222}) - 4(t_{122}t_{221} - t_{121}t_{222})\}.
\end{aligned}$$

This completes the computations.

4.3.1. Distribution of the exponent, real case

Let us determine the distribution of the exponent in the $p \times q$ real matrix-variate Gaussian density. Letting $u = \text{tr}(AXBX')$, its density can be obtained by evaluating its associated mgf. Then, taking t as its scalar parameter since u is scalar, we have

$$M_u(t) = E[e^{tu}] = E[e^{t \text{tr}(AXBX')}].$$

Since this expected value depends on X , we can integrate out over the density of X :

$$\begin{aligned}
M_u(t) &= C \int_X e^{t \text{tr}(AXBX') - \frac{1}{2} \text{tr}(AXBX')} dX \\
&= C \int_X e^{-\frac{1}{2}(1-2t)(\text{tr}(AXBX'))} dX \quad \text{for } 1 - 2t > 0 \tag{i}
\end{aligned}$$

where

$$C = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}}.$$

The integral in (i) is convergent only when $1 - 2t > 0$. Then distributing $\sqrt{(1 - 2t)}$ to each element in X and X' , and denoting the new matrix by X_t , we have $X_t = \sqrt{(1 - 2t)}X \Rightarrow dX_t = (\sqrt{(1 - 2t)})^{pq} dX = (1 - 2t)^{\frac{pq}{2}} dX$. Integral over X_t , together with C , yields 1 and hence

$$M_u(t) = (1 - 2t)^{-\frac{pq}{2}}, \quad \text{provided } 1 - 2t > 0. \tag{4.3.6}$$

The corresponding density is a real chi-square having pq degrees of freedom or a real gamma density with parameters $\alpha = \frac{pq}{2}$ and $\beta = 2$. Thus, the resulting density, denoted by $f_{u_1}(u_1)$, is given by

$$f_{u_1}(u_1) = [2^{\frac{pq}{2}} \Gamma(pq/2)]^{-1} u_1^{\frac{pq}{2}-1} e^{-\frac{u_1}{2}}, \quad 0 \leq u_1 < \infty, \quad p, q = 1, 2, \dots, \quad (4.3.7)$$

and $f_{u_1}(u_1) = 0$ elsewhere.

4.3a.1. Distribution of the exponent, complex case

In the complex case, letting $\tilde{u} = \text{tr}(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})$, we note that $\tilde{u} = \tilde{u}^*$ and \tilde{u} is a scalar, so that \tilde{u} is real. Hence, the mgf of \tilde{u} , with real parameter t , is given by

$$M_{\tilde{u}}(t) = E[e^{t \text{tr}(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})}] = C_1 \int_{\tilde{X}} e^{-(1-t)\text{tr}(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})} d\tilde{X}, \quad 1 - t > 0, \quad \text{with}$$

$$C_1 = \frac{|\det(A)|^q |\det(B)|^p}{\pi^{pq}}.$$

On making the transformation $\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}}$, we have

$$M_{\tilde{u}}(t) = \frac{1}{\pi^{pq}} \int_{\tilde{Y}} e^{-(1-t)\text{tr}(\tilde{Y} \tilde{Y}^*)} d\tilde{Y}.$$

However,

$$\text{tr}(\tilde{Y} \tilde{Y}^*) = \sum_{r=1}^p \sum_{s=1}^q |\tilde{y}_{rs}|^2 = \sum_{r=1}^p \sum_{s=1}^q (y_{rs1}^2 + y_{rs2}^2)$$

where $\tilde{y}_{rs} = y_{rs1} + iy_{rs2}$, $i = \sqrt{-1}$, y_{rs1} , y_{rs2} being real. Hence

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1-t)(y_{rs1}^2 + y_{rs2}^2)} dy_{rs1} \wedge dy_{rs2} = \frac{1}{1-t}, \quad 1 - t > 0.$$

Therefore,

$$M_{\tilde{u}}(t) = (1 - t)^{-pq}, \quad 1 - t > 0, \quad (4.3a.7)$$

and $\tilde{u} = v$ has a real gamma density with parameters $\alpha = pq$, $\beta = 1$, or a chi-square density in the complex domain with pq degrees of freedom, that is,

$$f_v(v) = \frac{1}{\Gamma(pq)} v^{pq-1} e^{-v}, \quad 0 \leq v < \infty, \quad (4.3a.8)$$

and $f_v(v) = 0$ elsewhere.

4.3.2. Linear functions in the real case

Let the $p \times q$ real matrix $X = (x_{ij})$ of the real scalar random variables x_{ij} 's have the density in (4.2.2), namely

$$f_{p,q}(X) = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}}}{(2\pi)^{\frac{pq}{2}}} e^{-\frac{1}{2} \text{tr}(A(X-M)B(X-M)')} \quad (4.3.8)$$

for $A > O$, $B > O$, where M is a $p \times q$ location parameter matrix. Let L_1 be a $p \times 1$ vector of constants. Consider the linear function $Z_1 = L_1' X$ where Z_1 is $1 \times q$. Let T be a $1 \times q$ parameter vector. Then the mgf of the $1 \times q$ vector Z_1 is

$$\begin{aligned} M_{Z_1}(T) &= E[e^{(TZ_1)'}] = E[e^{(TX'L_1)'}] = E[e^{\text{tr}(TX'L_1)'}] \\ &= E[e^{\text{tr}((L_1 T)X')'}]. \end{aligned} \quad (i)$$

This can be evaluated by replacing T by $L_1 T$ in (4.3.4). Then

$$\begin{aligned} M_{Z_1}(T) &= e^{\text{tr}((L_1 T)M') + \frac{1}{2} \text{tr}(A^{-1} L_1 T B^{-1} (L_1 T)')} \\ &= e^{\text{tr}(TM'L_1) + \frac{1}{2} \text{tr}[(L_1' A^{-1} L_1) T B^{-1} T']}. \end{aligned} \quad (ii)$$

Since $L_1' A^{-1} L_1$ is a scalar,

$$(L_1' A^{-1} L_1) T B^{-1} T' = T (L_1' A^{-1} L_1) B^{-1} T'.$$

On comparing the resulting expression with the mgf of a q -variate real normal distribution, we observe that Z_1 is a q -variate real Gaussian vector with mean value vector $L_1' M$ and covariance matrix $[L_1' A^{-1} L_1] B^{-1}$. Hence the following result:

Theorem 4.3.1. *Let the real $p \times q$ matrix X have the density specified in (4.3.8) and L_1 be a $p \times 1$ constant vector. Let Z_1 be the linear function of X , $Z_1 = L_1' X$. Then Z_1 , which is $1 \times q$, has the mgf given in (ii) and thereby Z_1 has a q -variate real Gaussian density with the mean value vector $L_1' M$ and covariance matrix $[L_1' A^{-1} L_1] B^{-1}$.*

Theorem 4.3.2. *Let L_2 be a $q \times 1$ constant vector. Consider the linear function $Z_2 = X L_2$ where the $p \times q$ real matrix X has the density specified in (4.3.8). Then Z_2 , which is $p \times 1$, is a p -variate real Gaussian vector with mean value vector $M L_2$ and covariance matrix $[L_2' B^{-1} L_2] A^{-1}$.*

The proof of Theorem 4.3.2 is parallel to the derivation of that of Theorem 4.3.1. Theorems 4.3.1 and 4.3.2 establish that when the $p \times q$ matrix X has a $p \times q$ -variate real Gaussian density with parameters M , $A > O$, $B > O$, then all linear functions of the

form $L_1'X$ where L_1 is $p \times 1$ are q -variate real Gaussian and all linear functions of the type XL_2 where L_2 is $q \times 1$ are p -variate real Gaussian, the parameters in these Gaussian densities being given in Theorems 4.3.1 and 4.3.2.

By retracing the steps, we can obtain characterizations of the density of the $p \times q$ real matrix X through linear transformations. Consider all possible $p \times 1$ constant vectors L_1 or, equivalently, let L_1 be arbitrary. Let T be a $1 \times q$ parameter vector. Then the $p \times q$ matrix L_1T , denoted by $T_{(1)}$, contains pq free parameters. In this case the mgf in (ii) can be written as

$$M(T_{(1)}) = e^{\text{tr}(T_{(1)}M') + \frac{1}{2}\text{tr}(A^{-1}T_{(1)}B^{-1}T_{(1)}'), \quad (\text{iii})$$

which has the same structure of the mgf of a $p \times q$ real matrix-variate Gaussian density as given in (4.3.8), whose the mean value matrix is M and parameter matrices are $A > O$ and $B > O$. Hence, the following result can be obtained:

Theorem 4.3.3. *Let L_1 be a constant $p \times 1$ vector, X be a $p \times q$ matrix whose elements are real scalar variables and $A > O$ be $p \times p$ and $B > O$ be $q \times q$ constant real positive definite matrices. If for an arbitrary vector L_1 , $L_1'X$ is a q -variate real Gaussian vector as specified in Theorem 4.3.1, then X has a $p \times q$ real matrix-variate Gaussian density as given in (4.3.8).*

As well, a result parallel to this one follows from Theorem 4.3.2:

Theorem 4.3.4. *Let L_2 be a $q \times 1$ constant vector, X be a $p \times q$ matrix whose elements are real scalar variables and $A > O$ be $p \times p$ and $B > O$ be $q \times q$ constant real positive definite matrices. If for an arbitrary constant vector L_2 , XL_2 is a p -variate real Gaussian vector as specified in Theorem 4.3.2, then X is $p \times q$ real matrix-variate Gaussian distributed as in (4.3.8).*

Example 4.3.2. Consider a 2×2 matrix-variate real Gaussian density with the parameters

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = E[X], \quad X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}.$$

Letting $U_1 = L_1'X$, $U_2 = XL_2$, $U_3 = L_1'XL_2$, evaluate the densities of U_1 , U_2 , U_3 by applying Theorems 4.3.1 and 4.3.2 where $L_1' = [1, 1]$, $L_2' = [1, -1]$; as well, obtain those densities without resorting to these theorems.

Solution 4.3.2. Let us first compute the following quantities:

$$A^{-1}, \quad B^{-1}, \quad L_1'A^{-1}L_1, \quad L_2'B^{-1}L_2, \quad L_1'M, \quad ML_2, \quad L_1'ML_2.$$

They are

$$\begin{aligned}
 A^{-1} &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, \quad B^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \\
 L'_1 M &= [1, 1] \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = [1, 0], \quad ML_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \\
 L'_1 A^{-1} L_1 &= [1, 1] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1, \quad L'_2 B^{-1} L_2 = \frac{1}{3} [1, -1] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2, \\
 L'_1 ML_2 &= [1, 0] \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 2.
 \end{aligned}$$

Let $U_1 = L'_1 X$, $U_2 = XL_2$, $U_3 = L'_1 XL_2$. Then by making use of Theorems 4.3.1 and 4.3.2 and then, results from Chap. 2 on q -variate real Gaussian vectors, we have the following:

$$U_1 \sim N_2((1, 0), (1)B^{-1}), \quad U_2 \sim N_2(ML_2, 2A^{-1}), \quad U_3 \sim N_1(1, (1)(2)) = N_1(1, 2).$$

Let us evaluate the densities without resorting to these theorems. Note that $U_1 = [x_{11} + x_{21}, x_{12} + x_{22}]$. Then U_1 has a bivariate real distribution. Let us compute the mgf of U_1 . Letting t_1 and t_2 be real parameters, the mgf of U_1 is

$$M_{U_1}(t_1, t_2) = E[e^{t_1(x_{11}+x_{21})+t_2(x_{12}+x_{22})}] = E[e^{t_1 x_{11}+t_1 x_{21}+t_2 x_{12}+t_2 x_{22}}],$$

which is available from the mgf of X by letting $t_{11} = t_1$, $t_{21} = t_1$, $t_{12} = t_2$, $t_{22} = t_2$. Thus,

$$\begin{aligned}
 A^{-1}T &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} t_1 & t_2 \\ t_1 & t_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ t_1 & t_2 \end{bmatrix} \\
 B^{-1}T' &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} t_1 & t_1 \\ t_2 & t_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2t_1 - t_2 & 2t_1 - t_2 \\ -t_1 + 2t_2 & -t_1 + 2t_2 \end{bmatrix},
 \end{aligned}$$

so that

$$\frac{1}{2} \text{tr}(A^{-1}TB^{-1}T') = \frac{1}{2} \left\{ \frac{1}{3} [2t_1^2 + 2t_2^2 - 2t_1 t_2] \right\} = \frac{1}{2} [t_1, t_2] B^{-1} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}. \quad (i)$$

Since

$$U_2 = XL_2 = \begin{bmatrix} x_{11} - x_{12} \\ x_{21} - x_{22} \end{bmatrix},$$

we let $t_{11} = t_1$, $t_{12} = -t_1$, $t_{21} = t_2$, $t_{22} = -t_2$. With these substitutions, we have the following:

$$A^{-1}T = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} t_1 & -t_1 \\ t_2 & -t_2 \end{bmatrix} = \begin{bmatrix} t_1 - t_2 & -t_1 + t_2 \\ -t_1 + 2t_2 & t_1 - 2t_2 \end{bmatrix}$$

$$B^{-1}T' = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} t_1 & t_2 \\ -t_1 & -t_2 \end{bmatrix} = \begin{bmatrix} t_1 & t_2 \\ -t_1 & -t_2 \end{bmatrix}.$$

Hence,

$$\begin{aligned} \text{tr}(A^{-1}TB^{-1}T') &= t_1(t_1 - t_2) - t_1(-t_1 + t_2) + t_2(-t_1 + 2t_2) - t_2(t_1 - 2t_2) \\ &= 2[t_1, t_2] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}. \end{aligned}$$

Therefore, U_2 is a 2-variate real Gaussian with covariance matrix $2A^{-1}$ and mean value vector $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. That is, $U_2 \sim N_2(ML_2, 2A^{-1})$. For determining the distribution of U_3 , observe that $L'_1XL_2 = L'_1U_2$. Then, L'_1U_2 is univariate real Gaussian with mean value $E[L'_1U_2] = L'_1ML_2 = [1, 1] \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 1$ and variance $L'_1\text{Cov}(U_2)L_1 = L'_12A^{-1}L_1 = 2$. That is, $U_3 = u_3 \sim N_1(1, 2)$. This completes the solution.

The results stated in Theorems 4.3.1 and 4.3.2 are now generalized by taking sets of linear functions of X :

Theorem 4.3.5. *Let C' be a $r \times p$, $r \leq p$, real constant matrix of full rank r and G be a $q \times s$ matrix, $s \leq q$, of rank s . Let $Z = C'X$ and $W = XG$ where X has the density specified in (4.3.8). Then, Z has a $r \times q$ real matrix-variate Gaussian density with M replaced by $C'M$ and A^{-1} replaced by $C'A^{-1}C$, B^{-1} remaining unchanged, and $W = XG$ has a $p \times s$ real matrix-variate Gaussian distribution with B^{-1} replaced by $G'B^{-1}G$ and M replaced by MG , A^{-1} remaining unchanged.*

Example 4.3.3. Let the 2×2 real $X = (x_{ij})$ have a real matrix-variate Gaussian distribution with the parameters M , A and B . Consider the set of linear functions $U = C'X$ where

$$M = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}, \quad A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad C' = \begin{bmatrix} \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Show that the rows of U are independently distributed real q -variate Gaussian vectors with common covariance matrix B^{-1} and the rows of M as the mean value vectors.

Solution 4.3.3. Let us compute A^{-1} and $C'A^{-1}C$:

$$A^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$C'A^{-1}C = \begin{bmatrix} \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

In the mgf of $U = C'X$, A^{-1} is replaced by $C'A^{-1}C = I_2$ and B^{-1} remains the same. Then, the exponent in the mgf of U , excluding $\text{tr}(TM')$ is $\frac{1}{2}\text{tr}(TB^{-1}T') = \frac{1}{2}\sum_{j=1}^p T_j B^{-1} T_j'$ where T_j is the j -th row of T . Hence the p rows of U are independently distributed q -variate real Gaussian with the common covariance matrix B^{-1} . This completes the computations.

The previous example entails a general result that now is stated as a corollary.

Corollary 4.3.1. *Let X be a $p \times q$ -variate real Gaussian matrix with the usual parameters M , A and B , whose density is as given in (4.3.8). Consider the set of linear functions $U = C'X$ where C is a $p \times p$ constant matrix of full rank p and C is such that $A = CC'$. Then $C'A^{-1}C = C'(CC')^{-1}C = C'(C')^{-1}C^{-1}C = I_p$. Consequently, the rows of U , denoted by U_1, \dots, U_p , are independently distributed as real q -variate Gaussian vectors having the common covariance matrix B^{-1} .*

It is easy to construct such a C . Since $A = (a_{ij})$ is real positive definite, set it as $A = CC'$ where C is a lower triangular matrix with positive diagonal elements. The first row, first column element in $C = (c_{ij})$ is $c_{11} = +\sqrt{a_{11}}$. Note that since $A > O$, all the diagonal elements are real positive. The first column of C is readily available from the first column of A and c_{11} . Now, given a_{22} and the first column in C , c_{22} can be determined, and so on.

Theorem 4.3.6. *Let C , G and X be as defined in Theorem 4.3.5. Consider the $r \times s$ real matrix $Z = C'XG$. Then, when X has the distribution specified in (4.3.8), Z has an $r \times s$ real matrix-variate Gaussian density with M replaced by $C'MG$, A^{-1} replaced by $C'A^{-1}C$ and B^{-1} replaced by $G'B^{-1}G$.*

Example 4.3.4. Let the 2×2 matrix $X = (x_{ij})$ have a real matrix-variate Gaussian density with the parameters M , A and B , and consider the set of linear functions $Z = C'XG$ where C' is a $p \times p$ constant matrix of rank p and G is a $q \times q$ constant matrix of rank q , where

$$M = \begin{bmatrix} 2 & -1 \\ 1 & 5 \end{bmatrix}, \quad A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix},$$

$$C' = \begin{bmatrix} \sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad G = \begin{bmatrix} \sqrt{2} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{5}{2}} \end{bmatrix}.$$

Show that all the elements z_{ij} 's in $Z = (z_{ij})$ are mutually independently distributed real scalar standard Gaussian random variables when $M = O$.

Solution 4.3.4. We have already shown in Example 4.3.3 that $C'A^{-1}C = I$. Let us verify that $GG' = B$ and compute $G'B^{-1}G$:

$$GG' = \begin{bmatrix} \sqrt{2} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{5}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{5}{2}} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = B;$$

$$B^{-1} = \frac{1}{5} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix},$$

$$G'B^{-1}G = \frac{1}{5} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{5}{2}} \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} G = \frac{1}{5} \begin{bmatrix} 3\sqrt{2} - \frac{1}{\sqrt{2}} & 0 \\ -\sqrt{\frac{5}{2}} & 2\sqrt{\frac{5}{2}} \end{bmatrix} G$$

$$= \frac{1}{5} \begin{bmatrix} 3\sqrt{2} - \frac{1}{\sqrt{2}} & 0 \\ -\sqrt{\frac{5}{2}} & 2\sqrt{\frac{5}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{5}{2}} \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = I_2.$$

Thus, A^{-1} is replaced by $C'A^{-1}C = I_2$ and B^{-1} is replaced by $G'B^{-1}G = I_2$ in the mgf of Z , so that the exponent in the mgf, excluding $\text{tr}(TM')$, is $\frac{1}{2}\text{tr}(TT')$. It follows that all the elements in $Z = C'XG$ are mutually independently distributed real scalar standard normal variables. This completes the computations.

The previous example also suggests the following results which are stated as corollaries:

Corollary 4.3.2. *Let the $p \times q$ real matrix $X = (x_{ij})$ have a real matrix-variate Gaussian density with parameters M , A and B , as given in (4.3.8). Consider the set of linear functions $Y = XG$ where G is a $q \times q$ constant matrix of full rank q , and let $B = GG'$. Then, the columns of Y , denoted by $Y_{(1)}, \dots, Y_{(q)}$, are independently distributed p -variate real Gaussian vectors with common covariance matrix A^{-1} and mean value $(MG)_{(j)}$, $j = 1, \dots, q$, where $(MG)_{(j)}$ is the j -th column of MG .*

Corollary 4.3.3. *Let $Z = C'XG$ where C is a $p \times p$ constant matrix of rank p , G is a $q \times q$ constant matrix of rank q and X is a real $p \times q$ Gaussian matrix whose parameters are $M = O$, A and B , the constant matrices C and G being such that $A = CC'$ and $B = GG'$. Then, all the elements z_{ij} in $Z = (z_{ij})$ are mutually independently distributed real scalar standard Gaussian random variables.*

4.3a.2. Linear functions in the complex case

We can similarly obtain results parallel to Theorems 4.3.1–4.3.6 in the complex case. Let \tilde{X} be $p \times q$ matrix in the complex domain, whose elements are scalar complex variables. Assume that \tilde{X} has a complex $p \times q$ matrix-variate density as specified in (4.2a.9) whose associated moment generating function is as given in (4.3a.3). Let \tilde{C}_1 be a $p \times 1$ constant vector, \tilde{C}_2 be a $q \times 1$ constant vector, \tilde{C} be a $r \times p$, $r \leq p$, constant matrix of rank r and \tilde{G} be a $q \times s$, $s \leq q$, a constant matrix of rank s . Then, we have the following results:

Theorem 4.3a.1. *Let \tilde{C}_1 be a $p \times 1$ constant vector as defined above and let the $p \times q$ matrix \tilde{X} have the density given in (4.2a.9) whose associated mgf is as specified in (4.3a.3). Let \tilde{U} be the linear function of \tilde{X} , $\tilde{U} = \tilde{C}_1^* \tilde{X}$. Then \tilde{U} has a q -variate complex Gaussian density with the mean value vector $\tilde{C}_1^* \tilde{M}$ and covariance matrix $[\tilde{C}_1^* A^{-1} \tilde{C}_1] B^{-1}$.*

Theorem 4.3a.2. *Let \tilde{C}_2 be a $q \times 1$ constant vector. Consider the linear function $\tilde{Y} = \tilde{X} \tilde{C}_2$ where the $p \times q$ complex matrix \tilde{X} has the density (4.2a.9). Then \tilde{Y} is a p -variate complex Gaussian random vector with the mean value vector $\tilde{M} \tilde{C}_2$ and the covariance matrix $[\tilde{C}_2^* B^{-1} \tilde{C}_2] A^{-1}$.*

Note 4.3a.1. Consider the mgf's of \tilde{U} and \tilde{Y} in Theorems 4.3a.1 and 4.3a.2, namely $M_{\tilde{U}}(\tilde{T}) = E[e^{\Re(\tilde{T} \tilde{U}^*)}]$ and $M_{\tilde{Y}}(\tilde{T}) = E[e^{\Re(\tilde{Y}^* \tilde{T})}]$ with the conjugate transpose of the variable part in the linear form in the exponent; then \tilde{T} in $M_{\tilde{U}}(\tilde{T})$ has to be $1 \times q$ and \tilde{T} in $M_{\tilde{Y}}(\tilde{T})$ has to be $p \times 1$. Thus, the exponent in Theorem 4.3a.1 will be of the form $[\tilde{C}_1^* A^{-1} \tilde{C}_1] \tilde{T} B^{-1} \tilde{T}^*$ whereas the corresponding exponent in Theorem 4.3a.2 will be $[\tilde{C}_2^* B^{-1} \tilde{C}_2] \tilde{T}^* A^{-1} \tilde{T}$. Note that in one case, we have $\tilde{T} B^{-1} \tilde{T}^*$ and in the other case, \tilde{T} and \tilde{T}^* are interchanged as are A and B . This has to be kept in mind when applying these theorems.

Example 4.3a.2. Consider a 2×2 matrix \tilde{X} having a complex 2×2 matrix-variate Gaussian density with the parameters $M = O$, A and B , as well as the 2×1 vectors L_1 and L_2 and the linear functions $L_1^* \tilde{X}$ and $\tilde{X} L_2$ where

$$A = \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}, L_1 = \begin{bmatrix} -2i \\ 3i \end{bmatrix}, L_2 = \begin{bmatrix} -i \\ -2i \end{bmatrix} \text{ and } \tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{21} & \tilde{x}_{22} \end{bmatrix}.$$

Evaluate the densities of $\tilde{U} = L_1^* \tilde{X}$ and $\tilde{Y} = \tilde{X} L_2$ by applying Theorems 4.3a.1 and 4.3a.2, as well as independently.

Solution 4.3a.2. First, we compute the following quantities:

$$A^{-1} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}, B^{-1} = \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix},$$

$$L_1^* = [2i, -3i], L_2^* = [i, 2i],$$

so that

$$L_1^* A^{-1} L_1 = [2i, -3i] \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} -2i \\ 3i \end{bmatrix} = 22$$

$$L_2^* B^{-1} L_2 = [i, 2i] \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix} \begin{bmatrix} -i \\ -2i \end{bmatrix} = 6.$$

Then, as per Theorems 4.3a.1 and 4.3a.2, \tilde{U} is a q -variate complex Gaussian vector whose covariance matrix is $22 B^{-1}$ and \tilde{Y} is a p -variate complex Gaussian vector whose covariance matrix is $6 A^{-1}$, that is, $\tilde{U} \sim \tilde{N}_2(O, 22 B^{-1})$, $\tilde{Y} \sim \tilde{N}_2(O, 6 A^{-1})$. Now, let us determine the densities of \tilde{U} and \tilde{Y} without resorting to these theorems. Consider the mgf of \tilde{U} by taking the parameter vector \tilde{T} as $\tilde{T} = [\tilde{t}_1, \tilde{t}_2]$. Note that

$$\tilde{T} \tilde{U}^* = t_1(-2i\tilde{x}_{11}^* + 3i\tilde{x}_{21}^*) + t_2(-2i\tilde{x}_{12}^* + 3i\tilde{x}_{22}^*). \quad (i)$$

Then, in comparison with the corresponding part in the mgf of \tilde{X} whose associated general parameter matrix is $\tilde{T} = (\tilde{t}_{ij})$, we have

$$\tilde{t}_{11} = -2i\tilde{t}_1, \tilde{t}_{12} = -2i\tilde{t}_2, \tilde{t}_{21} = 3i\tilde{t}_1, \tilde{t}_{22} = 3i\tilde{t}_2. \quad (ii)$$

We now substitute the values of (ii) in the general mgf of \tilde{X} to obtain the mgf of \tilde{U} . Thus,

$$A^{-1} \tilde{T} = \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} -2i\tilde{t}_1 & -2i\tilde{t}_2 \\ 3i\tilde{t}_1 & 3i\tilde{t}_2 \end{bmatrix} = \begin{bmatrix} (-3-2i)\tilde{t}_1 & (-3-2i)\tilde{t}_2 \\ (-2+6i)\tilde{t}_1 & (-2+6i)\tilde{t}_2 \end{bmatrix}$$

$$B^{-1} \tilde{T}^* = \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix} \begin{bmatrix} 2i\tilde{t}_1^* & -3i\tilde{t}_1^* \\ 2i\tilde{t}_2^* & -3i\tilde{t}_2^* \end{bmatrix} = \begin{bmatrix} 4i\tilde{t}_1^* + 2\tilde{t}_2^* & -6i\tilde{t}_1^* - 3\tilde{t}_2^* \\ -2\tilde{t}_1^* + 2i\tilde{t}_2^* & 3\tilde{t}_1^* - 3i\tilde{t}_2^* \end{bmatrix}.$$

Here an asterisk only denotes the conjugate as the quantities are scalar.

$$\begin{aligned}
\text{tr}[A^{-1}\tilde{T}B^{-1}\tilde{T}^*] &= [-3 - 2i]\tilde{t}_1[4i\tilde{t}_1^* + 2\tilde{t}_2^*] + [(-3 - 2i)\tilde{t}_2[-2\tilde{t}_1^* + 2i\tilde{t}_2^*] \\
&\quad + [(-2 + 6i)\tilde{t}_1][(-6i\tilde{t}_1^* - 3\tilde{t}_2^*)] + [(-2 + 6i)\tilde{t}_2][3\tilde{t}_1^* - 3i\tilde{t}_2^*] \\
&= 22[2\tilde{t}_1\tilde{t}_1^* - i\tilde{t}_1\tilde{t}_2^* + i\tilde{t}_2\tilde{t}_1^* + \tilde{t}_2\tilde{t}_2^*] \\
&= 22[\tilde{t}_1, \tilde{t}_2] \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix} \begin{bmatrix} \tilde{t}_1^* \\ \tilde{t}_2^* \end{bmatrix} = 22\tilde{T}B^{-1}\tilde{T}^*, \quad \tilde{T} = [\tilde{t}_1, \tilde{t}_2].
\end{aligned}$$

This shows that $\tilde{U} \sim \tilde{N}_2(O, 22B^{-1})$. Now, consider

$$\tilde{Y} = \tilde{X}L_2 = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{21} & \tilde{x}_{22} \end{bmatrix} \begin{bmatrix} -i \\ -2i \end{bmatrix} = \begin{bmatrix} -i\tilde{x}_{11} - 2i\tilde{x}_{12} \\ -i\tilde{x}_{21} - 2i\tilde{x}_{22} \end{bmatrix}.$$

Then, on comparing the mgf of \tilde{Y} with that of \tilde{X} whose general parameter matrix is $\tilde{T} = (\tilde{t}_{ij})$, we have

$$\tilde{t}_{11} = i\tilde{t}_1, \quad \tilde{t}_{12} = 2i\tilde{t}_1, \quad \tilde{t}_{21} = i\tilde{t}_2, \quad \tilde{t}_{22} = 2i\tilde{t}_2.$$

On substituting these values in the mgf of \tilde{X} , we have

$$\begin{aligned}
A^{-1}\tilde{T} &= \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} i\tilde{t}_1 & 2i\tilde{t}_1 \\ i\tilde{t}_2 & 2i\tilde{t}_2 \end{bmatrix} = \begin{bmatrix} i\tilde{t}_1 - \tilde{t}_2 & 2i\tilde{t}_1 - 2\tilde{t}_2 \\ \tilde{t}_1 + 2i\tilde{t}_2 & 2\tilde{t}_1 + 4i\tilde{t}_2 \end{bmatrix} \\
B^{-1}\tilde{T}^* &= \begin{bmatrix} 2 & -i \\ i & 1 \end{bmatrix} \begin{bmatrix} -i\tilde{t}_1^* & -i\tilde{t}_2^* \\ -2i\tilde{t}_1^* & -2i\tilde{t}_2^* \end{bmatrix} = \begin{bmatrix} (-2 - 2i)\tilde{t}_1^* & (-2 - 2i)\tilde{t}_2^* \\ (1 - 2i)\tilde{t}_1^* & (1 - 2i)\tilde{t}_2^* \end{bmatrix},
\end{aligned}$$

so that

$$\begin{aligned}
\text{tr}[A^{-1}\tilde{T}B^{-1}\tilde{T}^*] &= [(\tilde{t}_1 - \tilde{t}_2)][(-2 - 2i)\tilde{t}_1^*] + [2i\tilde{t}_1 - 2\tilde{t}_2][(1 - 2i)\tilde{t}_1^*] \\
&\quad + [\tilde{t}_1 + 2i\tilde{t}_2][(-2 - 2i)\tilde{t}_2^*] + [2\tilde{t}_1 + 4i\tilde{t}_2][(1 - 2i)\tilde{t}_2^*] \\
&= 6[\tilde{t}_1\tilde{t}_1^* - i\tilde{t}_1\tilde{t}_2^* + i\tilde{t}_2\tilde{t}_1^* + 2\tilde{t}_2\tilde{t}_2^*] \\
&= 6[\tilde{t}_1^*, \tilde{t}_2^*] \begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} \tilde{t}_1 \\ \tilde{t}_2 \end{bmatrix} = 6\tilde{T}^*A^{-1}\tilde{T};
\end{aligned}$$

refer to Note 4.3a.1 regarding the representation of the quadratic forms in the two cases above. This shows that $\tilde{Y} \sim \tilde{N}_2(O, 6A^{-1})$. This completes the computations.

Theorem 4.3a.3. *Let \tilde{C}_1 be a constant $p \times 1$ vector, \tilde{X} be a $p \times q$ matrix whose elements are complex scalar variables and let $A = A^* > O$ be $p \times p$ and $B = B^* > O$ be $q \times q$ constant Hermitian positive definite matrices, where an asterisk denotes the conjugate*

transpose. If, for an arbitrary $p \times 1$ constant vector \tilde{C}_1 , $\tilde{C}_1^* \tilde{X}$ is a q -variate complex Gaussian vector as specified in Theorem 4.3a.1, then \tilde{X} has the $p \times q$ complex matrix-variate Gaussian density given in (4.2a.9).

As well, a result parallel to this one follows from Theorem 4.3a.2:

Theorem 4.3a.4. Let \tilde{C}_2 be a $q \times 1$ constant vector, \tilde{X} be a $p \times q$ matrix whose elements are complex scalar variables and let $A > O$ be $p \times p$ and $B > O$ be $q \times q$ Hermitian positive definite constant matrices. If, for an arbitrary constant vector \tilde{C}_2 , $\tilde{X} \tilde{C}_2$ is a p -variate complex Gaussian vector as specified in Theorem 4.3a.2, then \tilde{X} is $p \times q$ complex matrix-variate Gaussian matrix which is distributed as in (4.2a.9).

Theorem 4.3a.5. Let \tilde{C}^* be a $r \times p$, $r \leq p$, complex constant matrix of full rank r and \tilde{G} be a $q \times s$, $s \leq q$, constant complex matrix of full rank s . Let $\tilde{U} = \tilde{C}^* \tilde{X}$ and $\tilde{W} = \tilde{X} \tilde{G}$ where \tilde{X} has the density given in (4.2a.9). Then, \tilde{U} has a $r \times q$ complex matrix-variate density with \tilde{M} replaced by $\tilde{C}^* \tilde{M}$, A^{-1} replaced by $\tilde{C}^* A^{-1} \tilde{C}$ and B^{-1} remaining the same, and \tilde{W} has a $p \times s$ complex matrix-variate distribution with B^{-1} replaced by $\tilde{G}^* B^{-1} \tilde{G}$, \tilde{M} replaced by $\tilde{M} \tilde{G}$ and A^{-1} remaining the same.

Theorem 4.3a.6. Let \tilde{C}^* , \tilde{G} and \tilde{X} be as defined in Theorem 4.3a.5. Consider the $r \times s$ complex matrix $\tilde{Z} = \tilde{C}^* \tilde{X} \tilde{G}$. Then when \tilde{X} has the distribution specified by (4.2a.9), \tilde{Z} has an $r \times s$ complex matrix-variate density with \tilde{M} replaced by $\tilde{C}^* \tilde{M} \tilde{G}$, A^{-1} replaced by $\tilde{C}^* A^{-1} \tilde{C}$ and B^{-1} replaced by $\tilde{G}^* B^{-1} \tilde{G}$.

Example 4.3a.3. Consider a 2×3 matrix \tilde{X} having a complex matrix-variate Gaussian distribution with the parameter matrices $\tilde{M} = O$, A and B where

$$A = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & i & 0 \\ -i & 1 & i \\ 0 & -i & 2 \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix}.$$

Consider the linear forms

$$C^* \tilde{X} = \begin{bmatrix} i\tilde{x}_{11} - i\tilde{x}_{21} & i\tilde{x}_{12} - i\tilde{x}_{22} & i\tilde{x}_{13} - i\tilde{x}_{23} \\ \tilde{x}_{11} + 2\tilde{x}_{21} & \tilde{x}_{12} + 2\tilde{x}_{22} & \tilde{x}_{13} + 2\tilde{x}_{23} \end{bmatrix}$$

$$\tilde{X} G = \begin{bmatrix} \tilde{x}_{11} + i\tilde{x}_{12} + 2\tilde{x}_{13} & \tilde{x}_{12} & i\tilde{x}_{11} - i\tilde{x}_{12} + i\tilde{x}_{13} \\ \tilde{x}_{21} + i\tilde{x}_{22} + 2\tilde{x}_{23} & \tilde{x}_{22} & i\tilde{x}_{21} - i\tilde{x}_{22} + i\tilde{x}_{23} \end{bmatrix}.$$

(1): Compute the distribution of $\tilde{Z} = C^* \tilde{X} G$; (2): Compute the distribution of $\tilde{Z} = C^* \tilde{X} G$ if A remains the same and G is equal to

$$\begin{bmatrix} \sqrt{3} & 0 & 0 \\ -\frac{i}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 \\ 0 & -i\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

and study the properties of this distribution.

Solution 4.3a.3. Note that $A = A^*$ and $B = B^*$ and hence both A and B are Hermitian. Moreover, $|A| = 1$ and $|B| = 1$ and since all the leading minors of A and B are positive, A and B are both Hermitian positive definite. Then, the inverses of A and B in terms of the cofactors of their elements are

$$A^{-1} = \begin{bmatrix} 1 & -i \\ i & 2 \end{bmatrix}, [\text{Cof}(B)]' = \begin{bmatrix} 1 & -2i & -1 \\ 2i & 6 & -3i \\ -1 & 3i & 2 \end{bmatrix} = B^{-1}.$$

The linear forms provided above in connection with part (1) of this exercise can be respectively expressed in terms of the following matrices:

$$C^* = \begin{bmatrix} i & -i \\ 1 & 2 \end{bmatrix}, G = \begin{bmatrix} 1 & 0 & i \\ i & 1 & -i \\ 2 & 0 & i \end{bmatrix}.$$

Let us now compute $C^* A^{-1} C$ and $G^* B^{-1} G$:

$$\begin{aligned} C^* A^{-1} C &= \begin{bmatrix} i & -i \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -i \\ i & 2 \end{bmatrix} \begin{bmatrix} -i & 1 \\ i & 2 \end{bmatrix} = 3 \begin{bmatrix} 1 & 1-i \\ 1+i & 3 \end{bmatrix} \\ G^* B^{-1} G &= \begin{bmatrix} 1 & -i & 2 \\ 0 & 1 & 0 \\ -i & i & -i \end{bmatrix} \begin{bmatrix} 1 & -2i & -1 \\ 2i & 6 & -3i \\ -1 & 3i & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & i \\ i & 1 & -i \\ 2 & 0 & i \end{bmatrix} \\ &= \begin{bmatrix} 3 & -2i & -2+i \\ 2i & 6 & 1-6i \\ -2-i & 1+6i & 7 \end{bmatrix}. \end{aligned}$$

Then in (1), $C^* \tilde{X} G$ is a 2×3 complex matrix-variate Gaussian with A^{-1} replaced by $C^* A^{-1} C$ and B^{-1} replaced by $G^* B^{-1} G$ where $C^* A^{-1} C$ and $G^* B^{-1} G$ are given above. For answering (2), let us evaluate $G^* B^{-1} G$:

$$\begin{aligned}
G^* B^{-1} G &= \begin{bmatrix} \sqrt{3} & \frac{i}{\sqrt{3}} & 0 \\ 0 & \sqrt{\frac{2}{3}} & i\sqrt{\frac{3}{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & -2i & -1 \\ 2i & 6 & -3i \\ -1 & 3i & 2 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ -\frac{i}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 \\ 0 & -i\sqrt{\frac{3}{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.
\end{aligned}$$

Observe that this $q \times q$ matrix G which is such that $GG^* = B$, is nonsingular; thus, $G^* B^{-1} G = G^*(GG^*)^{-1}G = G^*(G^*)^{-1}G^{-1}G = I$. Letting $\tilde{Y} = \tilde{X}G$, $\tilde{X} = \tilde{Y}G^{-1}$, and the exponent in the density of \tilde{X} becomes

$$\text{tr}(A^{-1}\tilde{X}B\tilde{X}^*) = \text{tr}(A^{-1}\tilde{Y}G^{-1}B(G^*)^{-1}\tilde{Y}^*) = \text{tr}(Y^*A\tilde{Y}) = \sum_{j=1}^p \tilde{Y}_{(j)}^* A \tilde{Y}_{(j)}$$

where the $\tilde{Y}_{(j)}$'s are the columns of \tilde{Y} , which are independently distributed complex p -variate Gaussian vectors with common covariance matrix A^{-1} . This completes the computations.

The conclusions obtained in the solution to Example 4.3a.1 suggest the corollaries that follow.

Corollary 4.3a.1. *Let the $p \times q$ matrix \tilde{X} have a matrix-variate complex Gaussian distribution with the parameters $M = O$, $A > O$ and $B > O$. Consider the transformation $\tilde{U} = C^*\tilde{X}$ where C is a $p \times p$ nonsingular matrix such that $CC^* = A$ so that $C^*A^{-1}C = I$. Then the rows of \tilde{U} , namely $\tilde{U}_1, \dots, \tilde{U}_p$, are mutually independently distributed q -variate complex Gaussian vectors with common covariance matrix B^{-1} .*

Corollary 4.3a.2. *Let the $p \times q$ matrix \tilde{X} have a matrix-variate complex Gaussian distribution with the parameters $M = O$, $A > O$ and $B > O$. Consider the transformation $\tilde{Y} = \tilde{X}G$ where G is a $q \times q$ nonsingular matrix such that $GG^* = B$ so that $G^*B^{-1}G = I$. Then the columns of \tilde{Y} , namely $\tilde{Y}_{(1)}, \dots, \tilde{Y}_{(q)}$, are independently distributed p -variate complex Gaussian vectors with common covariance matrix A^{-1} .*

Corollary 4.3a.3. *Let the $p \times q$ matrix \tilde{X} have a matrix-variate complex Gaussian distribution with the parameters $M = O$, $A > O$ and $B > O$. Consider the transformation $\tilde{Z} = C^*\tilde{X}G$ where C is a $p \times p$ nonsingular matrix such that $CC^* = A$ and G is a $q \times q$ nonsingular matrix such that $GG^* = B$. Then, the elements \tilde{z}_{ij} 's of $\tilde{Z} = (\tilde{z}_{ij})$ are mutually independently distributed complex standard Gaussian variables.*

4.3.3. Partitioning of the parameter matrix

Suppose that in the $p \times q$ real matrix-variate case, we partition T as $\begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$ where T_1 is $p_1 \times q$ and T_2 is $p_2 \times q$, so that $p_1 + p_2 = p$. Let $T_2 = O$ (a null matrix). Then,

$$TB^{-1}T' = \begin{pmatrix} T_1 \\ O \end{pmatrix} B^{-1} (T_1' \ O) = \begin{bmatrix} T_1 B^{-1} T_1' & O_1 \\ O_2 & O_3 \end{bmatrix}$$

where $T_1 B^{-1} T_1'$ is a $p_1 \times p_1$ matrix, O_1 is a $p_1 \times p_2$ null matrix, O_2 is a $p_2 \times p_1$ null matrix and O_3 is a $p_2 \times p_2$ null matrix. Let us similarly partition A^{-1} into sub-matrices:

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix},$$

where A^{11} is $p_1 \times p_1$ and A^{22} is $p_2 \times p_2$. Then,

$$\text{tr}(A^{-1}TB^{-1}T') = \text{tr} \begin{bmatrix} A^{11}T_1B^{-1}T_1' & O \\ O & O \end{bmatrix} = \text{tr}(A^{11}T_1B^{-1}T_1').$$

If A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is $p_1 \times p_1$ and A_{22} is $p_2 \times p_2$, then, as established in Sect. 1.3, we have

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

Therefore, under this special case of T , the mgf is given by

$$E[e^{\text{tr}(T_1 X_1)}] = e^{\frac{1}{2}\text{tr}((A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}T_1 B^{-1}T_1')}, \quad (4.3.9)$$

which is also the mgf of the $p_1 \times q$ sub-matrix of X . Note that the mgf's in (4.3.9) and (4.3.1) share an identical structure. Hence, due to the uniqueness of the mgf, X_1 has a real $p_1 \times q$ matrix-variate Gaussian density wherein the parameter B remains unchanged and A is replaced by $A_{11} - A_{12}A_{22}^{-1}A_{21}$, the A_{ij} 's denoting the sub-matrices of A as described earlier.

4.3.4. Distributions of quadratic and bilinear forms

Consider the real $p \times q$ Gaussian matrix U defined in (4.2.17) whose mean value matrix is $E[X] = M = O$ and let $U = XB^{\frac{1}{2}}$. Then,

$$U'AU = \begin{bmatrix} U_1'AU_1 & U_1'AU_2 & \dots & U_1'AU_q \\ U_2'AU_1 & U_2'AU_2 & \dots & U_2'AU_q \\ \vdots & \vdots & \ddots & \vdots \\ U_q'AU_1 & U_q'AU_2 & \dots & U_q'AU_q \end{bmatrix} \quad (4.3.10)$$

where the $p \times 1$ column vectors of U , namely, U_1, \dots, U_q , are independently distributed as $N_p(O, A^{-1})$ vectors, that is, the U_j 's are independently distributed real p -variate Gaussian (normal) vectors whose covariance matrix is $A^{-1} = E[UU']$, with density

$$f_{U_j}(U_j) = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(U_j'AU_j)}, \quad A > O. \quad (4.3.11)$$

What is then the distribution of $U_j'AU_j$ for any particular j and what are the distributions of $U_i'AU_j$, $i \neq j = 1, \dots, q$? Let $z_j = U_j'AU_j$ and $z_{ij} = U_i'AU_j$, $i \neq j$. Letting t be a scalar parameter, consider the mgf of z_j :

$$\begin{aligned} M_{z_j}(t) &= E[e^{tz_j}] = \int_{U_j} e^{tU_j'AU_j} f_{U_j}(U_j) dU_j \\ &= \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \int_{U_j} e^{-\frac{1}{2}(1-2t)U_j'AU_j} dU_j \\ &= (1-2t)^{-\frac{p}{2}} \quad \text{for } 1-2t > 0, \end{aligned}$$

which is the mgf of a real gamma random variable with parameters $\alpha = \frac{p}{2}$, $\beta = 2$ or a real chi-square random variable with p degrees of freedom for $p = 1, 2, \dots$. That is,

$$U_j'AU_j \sim \chi_p^2 \quad (\text{a real chi-square random variable having } p \text{ degrees of freedom}). \quad (4.3.12)$$

In the complex case, observe that $\tilde{U}_j^* A \tilde{U}_j$ is real when $A = A^* > O$ and hence, the parameter in the mgf is real. On making the transformation $A^{\frac{1}{2}} \tilde{U}_j = \tilde{V}_j$, $|\det(A)|$ is canceled. Then, the exponent can be expressed in terms of

$$-(1-t)\tilde{Y}^*\tilde{Y} = -(1-t) \sum_{j=1}^p |\tilde{y}_j|^2 = -(1-t) \sum_{j=1}^p (y_{j1}^2 + y_{j2}^2),$$

where $\tilde{y}_j = y_{j1} + iy_{j2}$, $i = \sqrt{-1}$. The integral gives $(1 - t)^{-p}$ for $1 - t > 0$. Hence, $\tilde{V}_j = \tilde{U}_j^* A \tilde{U}_j$ has a real gamma distribution with the parameters $\alpha = p$, $\beta = 1$, that is, a chi-square distribution with p degrees of freedom in the complex domain. Thus, $2\tilde{V}_j$ is a real chi-square random variable with $2p$ degrees of freedom, that is,

$$2\tilde{V}_j = 2\tilde{U}_j^* A \tilde{U}_j \sim \chi_{2p}^2. \quad (4.3a.9)$$

What is then the distribution of $U_i' A U_j$, $i \neq j$? Let us evaluate the mgf of $U_i' A U_j = z_{ij}$. As z_{ij} is a function of U_i and U_j , we can integrate out over the joint density of U_i and U_j where U_i and U_j are independently distributed p -variate real Gaussian random variables:

$$\begin{aligned} M_{z_{ij}}(t) &= E[e^{t z_{ij}}] = \int_{U_i} \int_{U_j} e^{t U_i' A U_j} f_{U_i}(U_i) f_{U_j}(U_j) dU_i \wedge dU_j \\ &= \frac{|A|}{(2\pi)^p} \int \int e^{t U_i' A U_j - \frac{1}{2} U_j' A U_j - \frac{1}{2} U_i' A U_i} dU_i \wedge dU_j. \end{aligned}$$

Let us first integrate out U_j . The relevant terms in the exponent are

$$-\frac{1}{2}(U_j' A U_j) + \frac{1}{2}(2t)(U_i' A U_j) = -\frac{1}{2}(U_j - C)' A (U_j - C) + \frac{1}{2}t^2 U_i' A U_i, \quad C = t U_i.$$

But the integral over U_j which is the integral over $U_j - C$ will result in the following representation:

$$\begin{aligned} M_{z_{ij}}(t) &= \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \int_{U_i} e^{\frac{t^2}{2} U_i' A U_i - \frac{1}{2} U_i' A U_i} dU_i \\ &= (1 - t^2)^{-\frac{p}{2}} \quad \text{for } 1 - t^2 > 0. \end{aligned} \quad (4.3.13)$$

What is the density corresponding to the mgf $(1 - t^2)^{-\frac{p}{2}}$? This is the mgf of a real scalar random variable u of the form $u = x - y$ where x and y are independently distributed real scalar chi-square random variables. For $p = 2$, x and y will be exponential variables so that u will have a double exponential distribution or a real Laplace distribution. In the general case, the density of u can also be worked out when x and y are independently distributed real gamma random variables with different parameters whereas chi-squares with equal degrees of freedom constitutes a special case. For the exact distribution of covariance structures such as the z_{ij} 's, see Mathai and Sebastian (2022).

Exercises 4.3

4.3.1. In the moment generating function (mgf) (4.3.3), partition the $p \times q$ parameter matrix T into column sub-matrices such that $T = (T_1, T_2)$ where T_1 is $p \times q_1$ and T_2 is $p \times q_2$ with $q_1 + q_2 = q$. Take $T_2 = O$ (the null matrix). Simplify and show that if X is similarly partitioned as $X = (Y_1, Y_2)$, then Y_1 has a real $p \times q_1$ matrix-variate Gaussian density. As well, show that Y_2 has a real $p \times q_2$ matrix-variate Gaussian density.

4.3.2. Referring to Exercises 4.3.1, write down the densities of Y_1 and Y_2 .

4.3.3. If T is the parameter matrix in (4.3.3), then what type of partitioning of T is required so that the densities of (1): the first row of X , (2): the first column of X can be determined, and write down these densities explicitly.

4.3.4. Repeat Exercises 4.3.1–4.3.3 by taking the mgf in (4.3a.3) for the corresponding complex case.

4.3.5. Write down the mgf explicitly for $p = 2$ and $q = 2$ corresponding to (4.3.3) and (4.3a.3), assuming general $A > O$ and $B > O$.

4.3.6. Partition the mgf in the complex $p \times q$ matrix-variate Gaussian case, corresponding to the partition in Sect. 4.3.1 and write down the complex matrix-variate density corresponding to \tilde{T}_1 in the complex case.

4.3.7. In the real $p \times q$ matrix-variate Gaussian case, partition the mgf parameter matrix into $T = (T_{(1)}, T_{(2)})$ where $T_{(1)}$ is $p \times q_1$ and $T_{(2)}$ is $p \times q_2$ with $q_1 + q_2 = q$. Obtain the density corresponding to $T_{(1)}$ by letting $T_{(2)} = O$.

4.3.8. Repeat Exercise 4.3.7 for the complex $p \times q$ matrix-variate Gaussian case.

4.3.9. Consider $v = \tilde{U}_j^* A \tilde{U}_j$. Provide the details of the steps for obtaining (4.3a.9).

4.3.10. Derive the mgf of $\tilde{U}_i^* A \tilde{U}_j$, $i \neq j$, in the complex $p \times q$ matrix-variate Gaussian case, corresponding to (4.3.13).

4.4. Marginal Densities in the Real Matrix-variate Gaussian Case

On partitioning the real $p \times q$ Gaussian matrix into X_1 of order $p_1 \times q$ and X_2 of order $p_2 \times q$ so that $p_1 + p_2 = p$, it was determined by applying the mgf technique that X_1 has a $p_1 \times q$ matrix-variate Gaussian distribution with the parameter matrices B remaining unchanged while A was replaced by $A_{11} - A_{12}A_{22}^{-1}A_{21}$ where the A_{ij} 's are the sub-matrices of A . This density is then the marginal density of the sub-matrix X_1 with respect to the joint density of X . Let us see whether the same result is available by direct

integration of the remaining variables, namely by integrating out X_2 . We first consider the real case. Note that

$$\begin{aligned}\operatorname{tr}(AXBX') &= \operatorname{tr} \left[A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} B \begin{pmatrix} X_1' & X_2' \end{pmatrix} \right] \\ &= \operatorname{tr} \left[A \begin{pmatrix} X_1 B X_1' & X_1 B X_2' \\ X_2 B X_1' & X_2 B X_2' \end{pmatrix} \right].\end{aligned}$$

Now, letting A be similarly partitioned, we have

$$\begin{aligned}\operatorname{tr}(AXBX') &= \operatorname{tr} \left[\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_1 B X_1' & X_1 B X_2' \\ X_2 B X_1' & X_2 B X_2' \end{pmatrix} \right] \\ &= \operatorname{tr}(A_{11} X_1 B X_1') + \operatorname{tr}(A_{12} X_2 B X_1') \\ &\quad + \operatorname{tr}(A_{21} X_1 B X_2') + \operatorname{tr}(A_{22} X_2 B X_2'),\end{aligned}$$

as the remaining terms do not appear in the trace. However, $(A_{12} X_2 B X_1')' = X_1 B X_2' A_{21}$, and since $\operatorname{tr}(PQ) = \operatorname{tr}(QP)$ and $\operatorname{tr}(S) = \operatorname{tr}(S')$ whenever S , PQ and QP are square matrices, we have

$$\operatorname{tr}(AXBX') = \operatorname{tr}(A_{11} X_1 B X_1') + 2\operatorname{tr}(A_{21} X_1 B X_2') + \operatorname{tr}(A_{22} X_2 B X_2').$$

We may now complete the quadratic form in $\operatorname{tr}(A_{22} X_2 B X_2') + 2\operatorname{tr}(A_{21} X_1 B X_2')$ by taking a matrix $C = A_{22}^{-1} A_{21} X_1$ and replacing X_2 by $X_2 + C$. Note that when $A > O$, $A_{11} > O$ and $A_{22} > O$. Thus,

$$\begin{aligned}\operatorname{tr}(AXBX') &= \operatorname{tr}(A_{22}(X_2 + C)B(X_2 + C)') + \operatorname{tr}(A_{11} X_1 B X_1') - \operatorname{tr}(A_{12} A_{22}^{-1} A_{21} X_1 B X_1') \\ &= \operatorname{tr}(A_{22}(X_2 + C)B(X_2 + C)') + \operatorname{tr}((A_{11} - A_{12} A_{22}^{-1} A_{21}) X_1 B X_1').\end{aligned}$$

On applying a result on partitioned matrices from Sect. 1.3, we have

$$|A| = |A_{22}| |A_{11} - A_{12} A_{22}^{-1} A_{21}|,$$

and clearly, $(2\pi)^{\frac{pq}{2}} = (2\pi)^{\frac{p_1 q}{2}} (2\pi)^{\frac{p_2 q}{2}}$. When integrating out X_2 , $|A_{22}|^{\frac{q}{2}}$ and $(2\pi)^{\frac{p_2 q}{2}}$ are getting canceled. Hence, the marginal density of X_1 , the $p_1 \times q$ sub-matrix of X , denoted by $f_{p_1, q}(X_1)$, is given by

$$f_{p_1, q}(X_1) = \frac{|B|^{\frac{p_1}{2}} |A_{11} - A_{12} A_{22}^{-1} A_{21}|^{\frac{q}{2}}}{(2\pi)^{\frac{p_1 q}{2}}} e^{-\frac{1}{2} \operatorname{tr}((A_{11} - A_{12} A_{22}^{-1} A_{21}) X_1 B X_1')}. \quad (4.4.1)$$

When $p_1 = 1$, $p_2 = 0$ and $p = 1$, we have the usual multivariate Gaussian density. When $p = 1$, the 1×1 matrix A will be taken as 1 without any loss of generality.

Then, from (4.4.1), the multivariate (q -variate) Gaussian density corresponding to (4.2.3) is given by

$$f_{1,q}(X_1) = \frac{(1/2)^{\frac{q}{2}} |B|^{\frac{1}{2}}}{\pi^{\frac{q}{2}}} e^{-\frac{1}{2} \text{tr}(X_1 B X_1')} = \frac{|B|^{\frac{1}{2}}}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2} X_1 B X_1'}$$

since the 1×1 quadratic form $X_1 B X_1'$ is equal to its trace. It is usually expressed in terms of $B = V^{-1}$, $V > O$. When $q = 1$, X is reducing to a $p \times 1$ vector, say Y . Thus, for a $p \times 1$ column vector Y with a location parameter μ , the density, denoted by $f_{p,1}(Y)$, is the following:

$$f_{p,1}(Y) = \frac{1}{|V|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2} (Y-\mu)' V^{-1} (Y-\mu)}, \quad (4.4.2)$$

where $Y' = (y_1, \dots, y_p)$, $\mu' = (\mu_1, \dots, \mu_p)$, $-\infty < y_j < \infty$, $-\infty < \mu_j < \infty$, $j = 1, \dots, p$, $V > O$. Observe that when Y is $p \times 1$, $\text{tr}(Y - \mu)' V^{-1} (Y - \mu) = (Y - \mu)' V^{-1} (Y - \mu)$. From symmetry, we can write down the density of the sub-matrix X_2 of X from the density given in (4.4.1). Let us denote the density of X_2 by $f_{p_2,q}(X_2)$. Then,

$$f_{p_2,q}(X_2) = \frac{|B|^{\frac{p_2}{2}} |A_{22} - A_{21} A_{11}^{-1} A_{12}|^{\frac{q}{2}}}{(2\pi)^{\frac{p_2 q}{2}}} e^{-\frac{1}{2} \text{tr}((A_{22} - A_{21} A_{11}^{-1} A_{12}) X_2 B X_2')}. \quad (4.4.3)$$

Note that $A_{22} - A_{21} A_{11}^{-1} A_{12} > O$ as $A > O$, our initial assumptions being that $A > O$ and $B > O$.

Theorem 4.4.1. *Let the $p \times q$ real matrix X have a real matrix-variate Gaussian density with the parameter matrices $A > O$ and $B > O$ where A is $p \times p$ and B is $q \times q$. Let X be partitioned into sub-matrices as $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $p_1 \times q$ and X_2 is $p_2 \times q$, with*

$p_1 + p_2 = p$. Let A be partitioned into sub-matrices as $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ where A_{11} is $p_1 \times p_1$. Then X_1 has a $p_1 \times q$ real matrix-variate Gaussian density with the parameter matrices $A_{11} - A_{12} A_{22}^{-1} A_{21} > O$ and $B > O$, as given in (4.4.1) and X_2 has a $p_2 \times q$ real matrix-variate Gaussian density with the parameter matrices $A_{22} - A_{21} A_{11}^{-1} A_{12} > O$ and $B > O$, as given in (4.4.3).

Observe that the p_1 rows taken in X_1 need not be the first p_1 rows. They can be any set of p_1 rows. In that instance, it suffices to make the corresponding permutations in the rows and columns of A and B so that the new set of p_1 rows can be taken as the first p_1 rows, and similarly for X_2 .

Can a similar result be obtained in connection with a matrix-variate Gaussian distribution if we take a set of column vectors and form a sub-matrix of X ? Let us partition the

$p \times q$ matrix X into sub-matrices of columns as $X = (Y_1 Y_2)$ where Y_1 is $p \times q_1$ and Y_2 is $p \times q_2$ such that $q_1 + q_2 = q$. The variables Y_1, Y_2 are used in order to avoid any confusion with X_1, X_2 utilized in the discussions so far. Let us partition B as follows:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad B_{11} \text{ being } q_1 \times q_1, \quad B_{22} \text{ being } q_2 \times q_2.$$

Then,

$$\begin{aligned} \text{tr}(AXBX') &= \text{tr}\left[A(Y_1 Y_2) \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{pmatrix} Y_1' \\ Y_2' \end{pmatrix}\right] \\ &= \text{tr}(AY_1 B_{11} Y_1') + \text{tr}(AY_2 B_{21} Y_1') + \text{tr}(AY_1 B_{12} Y_2') + \text{tr}(AY_2 B_{22} Y_2') \\ &= \text{tr}(AY_1 B_{11} Y_1') + 2\text{tr}(AY_1 B_{12} Y_2') + \text{tr}(AY_2 B_{22} Y_2'). \end{aligned}$$

As in the previous case of row sub-matrices, we complete the quadratic form:

$$\begin{aligned} \text{tr}(AXBX') &= \text{tr}(AY_1 B_{11} Y_1') - \text{tr}(AY_1 (B_{12} B_{22}^{-1} B_{21} Y_1') + \text{tr}(A(Y_2 + C) B_{22} (Y_2 + C)') \\ &= \text{tr}(AY_1 (B_{11} - B_{12} B_{22}^{-1} B_{21}) Y_1') + \text{tr}(A(Y_2 + C) B_{22} (Y_2 + C)'). \end{aligned}$$

Now, by integrating out Y_2 , we have the result, observing that $A > O$, $B > O$, $B_{11} - B_{12} B_{22}^{-1} B_{21} > O$ and $|B| = |B_{22}| |B_{11} - B_{12} B_{22}^{-1} B_{21}|$. A similar result follows for the marginal density of Y_2 . These results will be stated as the next theorem.

Theorem 4.4.2. *Let the $p \times q$ real matrix X have a real matrix-variate Gaussian density with the parameter matrices $M = O$, $A > O$ and $B > O$ where A is $p \times p$ and B is $q \times q$. Let X be partitioned into column sub-matrices as $X = (Y_1 Y_2)$ where Y_1 is $p \times q_1$ and Y_2 is $p \times q_2$ with $q_1 + q_2 = q$. Then Y_1 has a $p \times q_1$ real matrix-variate Gaussian density with the parameter matrices $A > O$ and $B_{11} - B_{12} B_{22}^{-1} B_{21} > O$, denoted by $f_{p,q_1}(Y_1)$, and Y_2 has a $p \times q_2$ real matrix-variate Gaussian density denoted by $f_{p,q_2}(Y_2)$ where*

$$f_{p,q_1}(Y_1) = \frac{|A|^{\frac{q_1}{2}} |B_{11} - B_{12} B_{22}^{-1} B_{21}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_1}{2}}} e^{-\frac{1}{2} \text{tr}[AY_1 (B_{11} - B_{12} B_{22}^{-1} B_{21}) Y_1']} \quad (4.4.4)$$

$$f_{p,q_2}(Y_2) = \frac{|A|^{\frac{q_2}{2}} |B_{22} - B_{21} B_{11}^{-1} B_{12}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_2}{2}}} e^{-\frac{1}{2} \text{tr}[AY_2 (B_{22} - B_{21} B_{11}^{-1} B_{12}) Y_2']} \quad (4.4.5)$$

If $q = 1$ and $q_2 = 0$ in (4.4.4), $q_1 = 1$. When $q = 1$, the 1×1 matrix B is taken to be 1. Then Y_1 in (4.4.4) is $p \times 1$ or a column vector of p real scalar variables. Let it still be denoted by Y_1 . Then the corresponding density, which is a real p -variate Gaussian

(normal) density, available from (4.4.4) or from the basic matrix-variate density, is the following:

$$\begin{aligned} f_{p,1}(Y_1) &= \frac{|A|^{\frac{1}{2}}(1/2)^{\frac{p}{2}}}{\pi^{\frac{p}{2}}} e^{-\frac{1}{2}\text{tr}(AY_1Y_1')} \\ &= \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\text{tr}(Y_1'AY_1)} = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}Y_1'AY_1}, \end{aligned} \quad (4.4.6)$$

observing that $\text{tr}(Y_1'AY_1) = Y_1'AY_1$ since Y_1 is $p \times 1$ and then, $Y_1'AY_1$ is 1×1 . In the usual representation of a multivariate Gaussian density, A replaced by $A = V^{-1}$, V being positive definite.

Example 4.4.1. Let the 2×3 matrix $X = (x_{ij})$ have a real matrix-variate distribution with the parameter matrices $M = O$, $A > O$, $B > O$ where

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}, \quad A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

Let us partition X , A and B as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = [Y_1, Y_2], \quad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where $A_{11} = (2)$, $A_{12} = (1)$, $A_{21} = (1)$, $A_{22} = (1)$, $X_1 = [x_{11}, x_{12}, x_{13}]$, $X_2 = [x_{21}, x_{22}, x_{23}]$,

$$Y_1 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \quad Y_2 = \begin{bmatrix} x_{13} \\ x_{23} \end{bmatrix}, \quad B_{11} = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$B_{21} = [0, 1]$, $B_{22} = (2)$. Compute the densities of X_1 , X_2 , Y_1 and Y_2 .

Solution 4.4.1. We need the following quantities: $A_{11} - A_{12}A_{22}^{-1}A_{21} = 2 - 1 = 1$, $A_{22} - A_{21}A_{11}^{-1}A_{12} = 1 - \frac{1}{2} = \frac{1}{2}$, $|B| = 1$,

$$\begin{aligned} B_{22} - B_{21}B_{11}^{-1}B_{12} &= 2 - [0, 1] \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2 - \frac{3}{2} = \frac{1}{2} \\ B_{11} - B_{12}B_{22}^{-1}B_{21} &= \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} [0, 1] = \begin{bmatrix} 3 & -1 \\ -1 & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

Let us compute the constant parts or normalizing constants in the various densities. With our usual notations, the normalizing constants in $f_{p_1,q}(X_1)$ and $f_{p_2,q}(X_2)$ are

$$\frac{|B|^{\frac{p_1}{2}} |A_{11} - A_{12}A_{22}^{-1}A_{21}|^{\frac{q}{2}}}{(2\pi)^{\frac{p_1q}{2}}} = \frac{|B|^{\frac{1}{2}}(1)^{\frac{3}{2}}}{(2\pi)^{\frac{3}{2}}}$$

$$\frac{|B|^{\frac{p_2}{2}} |A_{22} - A_{21}A_{11}^{-1}A_{12}|^{\frac{q}{2}}}{(2\pi)^{\frac{p_2q}{2}}} = \frac{|B|^{\frac{1}{2}}(\frac{1}{2})^{\frac{3}{2}}}{(2\pi)^{\frac{3}{2}}}.$$

Hence, the corresponding densities of X_1 and X_2 are the following:

$$f_{1,3}(X_1) = \frac{|B|^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}} e^{-\frac{1}{2}X_1 B X_1'}, \quad -\infty < x_{1j} < \infty, \quad j = 1, 2, 3,$$

$$f_{1,3}(X_2) = \frac{|B|^{\frac{1}{2}}}{2^{\frac{3}{2}}(2\pi)^{\frac{3}{2}}} e^{-\frac{1}{4}(X_2 B X_2')}, \quad -\infty < x_{2j} < \infty, \quad j = 1, 2, 3.$$

Let us now evaluate the normalizing constants in the densities $f_{p,q_1}(Y_1)$, $f_{p,q_2}(Y_2)$:

$$\frac{|A|^{\frac{q_1}{2}} |B_{11} - B_{12}B_{22}^{-1}B_{21}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_1}{2}}} = \frac{|A|^{\frac{1}{2}}(\frac{1}{2})^1}{4\pi^2} = \frac{1}{8\pi^2},$$

$$\frac{|A|^{\frac{q_2}{2}} |B_{22} - B_{21}B_{11}^{-1}B_{12}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_2}{2}}} = \frac{|A|^{\frac{1}{2}}(\frac{1}{2})^1}{2\pi} = \frac{1}{4\pi}.$$

Thus, the density of Y_1 is

$$f_{2,2}(Y_1) = \frac{1}{8\pi^2} e^{-\frac{1}{2}\text{tr}\{AY_1(B_{11}-B_{12}B_{22}^{-1}B_{21})Y_1'\}}$$

$$= \frac{1}{8\pi^2} e^{-\frac{1}{2}Q}, \quad -\infty < x_{ij} < \infty, \quad i, j = 1, 2,$$

where

$$Q = \text{tr} \left\{ \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \right\}$$

$$= 6x_{11}^2 + x_{12}^2 + 3x_{21}^2 + \frac{1}{2}x_{22}^2 - 4x_{11}x_{12} - 2x_{11}x_{22}$$

$$+ 3x_{11}x_{21} + x_{12}x_{22} - 2x_{12}x_{21} - x_{22}x_{21},$$

the density of Y_2 being

$$\begin{aligned} f_{2,1}(Y_2) &= \frac{1}{4\pi} e^{-\frac{1}{2}\text{tr}\{AY_2(B_{22}-B_{21}B_{11}^{-1}B_{12})Y_2'\}} \\ &= \frac{1}{4\pi} e^{-\frac{1}{4}[2x_{13}^2+x_{23}^2+2x_{13}x_{23}]}, \quad -\infty < x_{i3} < \infty, \quad i = 1, 2. \end{aligned}$$

This completes the computations.

4.4a. Marginal Densities in the Complex Matrix-variate Gaussian Case

The derivations of the results are parallel to those provided in the real case. Accordingly, we will state the corresponding results.

Theorem 4.4a.1. *Let the $p \times q$ matrix \tilde{X} have a complex matrix-variate Gaussian density with the parameter matrices $M = O$, $A > O$, $B > O$ where A is $p \times p$ and B is $q \times q$. Consider a row partitioning of \tilde{X} into sub-matrices \tilde{X}_1 and \tilde{X}_2 where \tilde{X}_1 is $p_1 \times q$ and \tilde{X}_2 is $p_2 \times q$, with $p_1 + p_2 = p$. Then, \tilde{X}_1 and \tilde{X}_2 have $p_1 \times q$ complex matrix-variate and $p_2 \times q$ complex matrix-variate Gaussian densities with parameter matrices $A_{11} - A_{12}A_{22}^{-1}A_{21}$ and B , and $A_{22} - A_{21}A_{11}^{-1}A_{12}$ and B , respectively, denoted by $\tilde{f}_{p_1,q}(\tilde{X}_1)$ and $\tilde{f}_{p_2,q}(\tilde{X}_2)$. The density of \tilde{X}_1 is given by*

$$\tilde{f}_{p_1,q}(\tilde{X}_1) = \frac{|\det(B)|^{p_1} |\det(A_{11} - A_{12}A_{22}^{-1}A_{21})|^q}{\pi^{p_1 q}} e^{-\text{tr}((A_{11} - A_{12}A_{22}^{-1}A_{21})\tilde{X}_1 B \tilde{X}_1^*)}, \quad (4.4a.1)$$

the corresponding vector case for $p = 1$ being available from (4.4a.1) for $p_1 = 1$, $p_2 = 0$ and $p = 1$; in this case, the density is

$$\tilde{f}_{1,q}(\tilde{X}_1) = \frac{|\det(B)|}{\pi^q} e^{-(\tilde{X}_1 - \mu)B(\tilde{X}_1 - \mu)^*} \quad (4.4a.2)$$

where \tilde{X}_1 and μ are $1 \times q$ and μ is a location parameter vector. The density of \tilde{X}_2 is the following:

$$\tilde{f}_{p_2,q}(\tilde{X}_2) = \frac{|\det(B)|^{p_2} |\det(A_{22} - A_{21}A_{11}^{-1}A_{12})|^q}{\pi^{p_2 q}} e^{-\text{tr}((A_{22} - A_{21}A_{11}^{-1}A_{12})\tilde{X}_2 B \tilde{X}_2^*)}. \quad (4.4a.3)$$

Theorem 4.4a.2. *Let \tilde{X} , A and B be as defined in Theorem 4.2a.1 and let \tilde{X} be partitioned into column sub-matrices $\tilde{X} = (\tilde{Y}_1 \tilde{Y}_2)$ where \tilde{Y}_1 is $p \times q_1$ and \tilde{Y}_2 is $p \times q_2$, so that $q_1 + q_2 = q$. Then \tilde{Y}_1 and \tilde{Y}_2 have $p \times q_1$ complex matrix-variate and $p \times q_2$ complex matrix-variate Gaussian densities given by*

$$\begin{aligned} \tilde{f}_{p,q_1}(\tilde{Y}_1) &= \frac{|\det(A)|^{q_1} |\det(B_{11} - B_{12} B_{22}^{-1} B_{21})|^p}{\pi^{pq_1}} \\ &\times e^{-\text{tr}(A\tilde{Y}_1(B_{11} - B_{12} B_{22}^{-1} B_{21})\tilde{Y}_1^*)} \end{aligned} \quad (4.4a.4)$$

$$\begin{aligned} \tilde{f}_{p,q_2}(\tilde{Y}_2) &= \frac{|\det(A)|^{q_2} |\det(B_{22} - B_{21} B_{11}^{-1} B_{12})|^p}{\pi^{pq_2}} \\ &\times e^{-\text{tr}(A\tilde{Y}_2(B_{22} - B_{21} B_{11}^{-1} B_{12})\tilde{Y}_2^*)}. \end{aligned} \quad (4.4a.5)$$

When $q = 1$, we have the usual complex multivariate case. In this case, it will be a p -variate complex Gaussian density. This is available from (4.4a.4) by taking $q_1 = 1$, $q_2 = 0$ and $q = 1$. Now, \tilde{Y}_1 is a $p \times 1$ column vector. Let μ be a $p \times 1$ location parameter vector. Then the density is

$$\tilde{f}_{p,1}(\tilde{Y}_1) = \frac{|\det(A)|}{\pi^p} e^{-(\tilde{Y}_1 - \mu)^* A (\tilde{Y}_1 - \mu)} \quad (4.4a.6)$$

where $A > O$ (Hermitian positive definite), $\tilde{Y}_1 - \mu$ is $p \times 1$ and its $1 \times p$ conjugate transpose is $(\tilde{Y}_1 - \mu)^*$.

Example 4.4a.1. Consider a 2×3 complex matrix-variate Gaussian distribution with the parameters $M = O$, $A > O$, $B > O$ where

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{21} & \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix}, \quad A = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -i & i \\ i & 2 & -i \\ -i & i & 2 \end{bmatrix}.$$

Consider the partitioning

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} = [\tilde{Y}_1, \tilde{Y}_2]$$

where

$$\tilde{Y}_1 = \begin{bmatrix} \tilde{x}_{11} \\ \tilde{x}_{21} \end{bmatrix}, \quad \tilde{Y}_2 = \begin{bmatrix} \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 2 & -i \\ i & 2 \end{bmatrix}, \quad B_{21} = \begin{bmatrix} i \\ -i \end{bmatrix},$$

$\tilde{X}_1 = [\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{13}]$, $\tilde{X}_2 = [\tilde{x}_{21}, \tilde{x}_{22}, \tilde{x}_{23}]$, $A_{11} = (2)$, $A_{12} = (i)$, $A_{21} = (-i)$, $A_{22} = 2$; $B_{11} = (2)$, $B_{12} = [-i, i]$. Compute the densities of \tilde{X}_1 , \tilde{X}_2 , \tilde{Y}_1 , \tilde{Y}_2 .

Solution 4.4a.1. It is easy to ascertain that $A = A^*$ and $B = B^*$; hence both matrices are Hermitian. As well, all the leading minors of A and B are positive so that $A > O$ and $B > O$. We need the following numerical results: $|A| = 3$, $|B| = 2$,

$$\begin{aligned} A_{11} - A_{12} A_{22}^{-1} A_{21} &= 2 - (i)(1/2)(-i) = 2 - \frac{1}{2} = \frac{3}{2} \\ A_{22} - A_{21} A_{11}^{-1} A_{12} &= 2 - (-i)(1/2)(i) = 2 - \frac{1}{2} = \frac{3}{2} \end{aligned}$$

$$\begin{aligned}
B_{11} - B_{12}B_{22}^{-1}B_{21} &= 2 - [-i, i] \left(\frac{1}{3}\right) \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} i \\ -i \end{bmatrix} = \frac{2}{3} \\
B_{22} - B_{21}B_{11}^{-1}B_{12} &= \begin{bmatrix} 2 & -i \\ i & 2 \end{bmatrix} - \begin{bmatrix} i \\ -i \end{bmatrix} \left(\frac{1}{2}\right) [-i, i] = \begin{bmatrix} 2 & -i \\ i & 2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 3 & 1-2i \\ 1+2i & 3 \end{bmatrix}.
\end{aligned}$$

With these preliminary calculations, we can obtain the required densities with our usual notations:

$$\begin{aligned}
\tilde{f}_{p_1, q}(\tilde{X}_1) &= \frac{|\det(B)|^{p_1} |\det(A_{11} - A_{12}A_{22}^{-1}A_{21})|^q}{\pi^{p_1 q}} \\
&\quad \times e^{-\text{tr}[(A_{11} - A_{12}A_{22}^{-1}A_{21})\tilde{X}_1 B \tilde{X}_1^*]}, \text{ that is,} \\
\tilde{f}_{1,3}(\tilde{X}_1) &= \frac{2(3/2)^3}{\pi^3} e^{-\frac{3}{2}\tilde{X}_1 B \tilde{X}_1^*}
\end{aligned}$$

where

$$\begin{aligned}
Q_1 &= \tilde{X}_1 B \tilde{X}_1^* = [\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{13}] \begin{bmatrix} 2 & -i & i \\ i & 2 & -i \\ -i & i & 2 \end{bmatrix} \begin{bmatrix} \tilde{x}_{11}^* \\ \tilde{x}_{12}^* \\ \tilde{x}_{13}^* \end{bmatrix} \\
&= 2\tilde{x}_{11}\tilde{x}_{11}^* + 2\tilde{x}_{12}\tilde{x}_{12}^* + 2\tilde{x}_{13}\tilde{x}_{13}^* - i\tilde{x}_{11}\tilde{x}_{12}^* + i\tilde{x}_{11}\tilde{x}_{13}^* \\
&\quad + i\tilde{x}_{12}\tilde{x}_{11}^* - i\tilde{x}_{12}\tilde{x}_{13}^* - i\tilde{x}_{13}\tilde{x}_{11}^* + i\tilde{x}_{13}\tilde{x}_{12}^*;
\end{aligned}$$

$$\begin{aligned}
\tilde{f}_{p_2, q}(\tilde{X}_2) &= \frac{|\det(A)|^{p_2} |\det(A_{22} - A_{21}A_{11}^{-1}A_{12})|^q}{\pi^{p_2 q}} \\
&\quad \times e^{-\text{tr}[(A_{22} - A_{21}A_{11}^{-1}A_{12})\tilde{X}_2 B \tilde{X}_2^*]}, \text{ that is,} \\
\tilde{f}_{1,3}(\tilde{X}_2) &= \frac{2(3/2)^3}{\pi^3} e^{-\frac{3}{2}\tilde{X}_2 B \tilde{X}_2^*}
\end{aligned}$$

where let $Q_2 = \tilde{X}_2 B \tilde{X}_2^*$, Q_2 being obtained by replacing \tilde{X}_1 in Q_1 by \tilde{X}_2 ;

$$\begin{aligned}
\tilde{f}_{p, q_1}(\tilde{Y}_1) &= \frac{|\det(A)|^{q_1} |\det(B_{11} - B_{12}B_{22}^{-1}B_{21})|^p}{\pi^{p q_1}} \\
&\quad \times e^{-\text{tr}[A\tilde{Y}_1(B_{11} - B_{12}B_{22}^{-1}B_{21})\tilde{Y}_1^*]}, \text{ that is,} \\
\tilde{f}_{2,1}(\tilde{Y}_1) &= \frac{3(2/3)^2}{\pi^2} e^{-\text{tr}(\frac{2}{3}Q_3)}
\end{aligned}$$

where

$$\begin{aligned} Q_3 &= \tilde{Y}_1^* A \tilde{Y}_1 = [\tilde{x}_{11}^*, \tilde{x}_{21}^*] \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} \tilde{x}_{11} \\ \tilde{x}_{21} \end{bmatrix} \\ &= 2\tilde{x}_{11}\tilde{x}_{11}^* + 2\tilde{x}_{21}\tilde{x}_{21}^* + i\tilde{x}_{21}\tilde{x}_{11}^* - i\tilde{x}_{21}^*\tilde{x}_{11}; \end{aligned}$$

$$\begin{aligned} \tilde{f}_{p,q_2}(\tilde{Y}_2) &= \frac{|\det(A)|^{q_2} |\det(B_{22} - B_{21} B_{11}^{-1} B_{12})|^p}{\pi^{p q_2}} \\ &\quad \times e^{-\text{tr}[A \tilde{Y}_2 (B_{22} - B_{21} B_{11}^{-1} B_{12}) \tilde{Y}_2^*]}, \text{ that is,} \end{aligned}$$

$$\tilde{f}_{2,2}(\tilde{Y}_2) = \frac{3^2}{\pi^4} e^{-\frac{1}{2} Q}$$

where

$$\begin{aligned} 2Q &= \text{tr} \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix} \begin{bmatrix} \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix} \begin{bmatrix} 3 & 1-2i \\ 1+2i & 3 \end{bmatrix} \begin{bmatrix} \tilde{x}_{12}^* & \tilde{x}_{22}^* \\ \tilde{x}_{13}^* & \tilde{x}_{23}^* \end{bmatrix} \\ &= 6\tilde{x}_{12}\tilde{x}_{12}^* + 6\tilde{x}_{13}\tilde{x}_{13}^* + 6\tilde{x}_{22}\tilde{x}_{22}^* + 6\tilde{x}_{23}\tilde{x}_{23}^* \\ &\quad + [2(1-2i)(\tilde{x}_{12}\tilde{x}_{13}^* + \tilde{x}_{22}\tilde{x}_{23}^*) + 2(1+2i)(\tilde{x}_{23}\tilde{x}_{22}^* + \tilde{x}_{13}\tilde{x}_{12}^*)] \\ &\quad + [i(1-2i)(\tilde{x}_{22}\tilde{x}_{13}^* - \tilde{x}_{12}\tilde{x}_{23}^*) - i(1+2i)(\tilde{x}_{13}\tilde{x}_{22}^* - \tilde{x}_{23}\tilde{x}_{12}^*)] \\ &\quad + 3i[\tilde{x}_{22}\tilde{x}_{12}^* + \tilde{x}_{23}\tilde{x}_{13}^* - \tilde{x}_{12}\tilde{x}_{22}^* - \tilde{x}_{13}\tilde{x}_{23}^*]. \end{aligned}$$

This completes the computations.

Exercises 4.4

4.4.1. Write down explicitly the density of a $p \times q$ matrix-variate Gaussian for $p = 3, q = 3$. Then by integrating out the other variables, obtain the density for the case (1): $p = 2, q = 2$; (2): $p = 2, q = 1$; (3): $p = 1, q = 2$; (4): $p = 1, q = 1$. Take the location matrix $M = O$. Let A and B to be general positive definite parameter matrices.

4.4.2. Repeat Exercise 4.4.1 for the complex case.

4.4.3. Write down the densities obtained in Exercises 4.4.1 and 4.4.2. Then evaluate the marginal densities for $p = 2, q = 2$ in both the real and complex domains by partitioning matrices and integrating out by using matrix methods.

4.4.4. Let the 2×2 real matrix $A > O$ where the first row is $(1, 1)$. Let the real $B > O$ be 3×3 where the first row is $(1, -1, 1)$. Complete A and B with numbers of your choosing

so that $A > O, B > O$. Consider a real 2×3 matrix-variate Gaussian density with these A and B as the parameter matrices. Take your own non-null location matrix. Write down the matrix-variate Gaussian density explicitly. Then by integrating out the other variables, either directly or by matrix methods, obtain (1): the 1×3 matrix-variate Gaussian density; (2): the 2×2 matrix-variate Gaussian density from your 2×3 matrix-variate Gaussian density.

4.4.5. Repeat Exercise 4.4.4 for the complex case if the first row of A is $(1, 1 + i)$ and the first row of B is $(2, 1 + i, 1 - i)$ where $A = A^* > O$ and $B = B^* > O$.

4.5. Conditional Densities in the Real Matrix-variate Gaussian Case

Consider a real $p \times q$ matrix-variate Gaussian density with the parameters $M = O, A > O, B > O$. Let us consider the partition of the $p \times q$ real Gaussian matrix X into row sub-matrices as $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $p_1 \times q$ and X_2 is $p_2 \times q$ with $p_1 + p_2 = p$. We have already established that the marginal density of X_2 is

$$f_{p_2,q}(X_2) = \frac{|A_{22} - A_{21}A_{11}^{-1}A_{12}|^{\frac{q}{2}}|B|^{\frac{p_2}{2}}}{(2\pi)^{\frac{p_2q}{2}}} e^{-\frac{1}{2}\text{tr}[(A_{22}-A_{21}A_{11}^{-1}A_{12})X_2BX_2']}$$

Thus, the conditional density of X_1 given X_2 is obtained as

$$f_{p_1,q}(X_1|X_2) = \frac{f_{p,q}(X)}{f_{p_2,q}(X_2)} = \frac{|A|^{\frac{q}{2}}|B|^{\frac{p}{2}}}{|A_{22} - A_{21}A_{11}^{-1}A_{12}|^{\frac{q}{2}}|B|^{\frac{p_2}{2}}} \frac{(2\pi)^{\frac{p_2q}{2}}}{(2\pi)^{\frac{pq}{2}}} \times e^{-\frac{1}{2}[\text{tr}(AXBX') + \text{tr}[(A_{22}-A_{21}A_{11}^{-1}A_{12})X_2BX_2']}$$

Note that

$$\begin{aligned} AXBX' &= A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} B \begin{pmatrix} X_1' & X_2' \end{pmatrix} = A \begin{bmatrix} X_1BX_1' & X_1BX_2' \\ X_2BX_1' & X_2BX_2' \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X_1BX_1' & X_1BX_2' \\ X_2BX_1' & X_2BX_2' \end{bmatrix} = \begin{bmatrix} \alpha & * \\ * & \beta \end{bmatrix} \end{aligned}$$

where $\alpha = A_{11}X_1BX_1' + A_{12}X_2BX_1', \beta = A_{21}X_1BX_2' + A_{22}X_2BX_2'$ and the asterisks designate elements that are not utilized in the determination of the trace. Then

$$\text{tr}(AXBX') = \text{tr}(A_{11}X_1BX_1' + A_{12}X_2BX_1') + \text{tr}(A_{21}X_1BX_2' + A_{22}X_2BX_2')$$

Thus the exponent in the conditional density simplifies to

$$\begin{aligned} & \text{tr}(A_{11}X_1BX_1') + 2\text{tr}(A_{12}X_2BX_1') + \text{tr}(A_{22}X_2BX_2') - \text{tr}[(A_{22} - A_{21}A_{11}^{-1}A_{12})X_2BX_2'] \\ &= \text{tr}(A_{11}X_1BX_1') + 2\text{tr}(A_{12}X_2BX_1') + \text{tr}[A_{21}A_{11}^{-1}A_{12}X_2BX_2'] \\ &= \text{tr}[A_{11}(X_1 + C)B(X_1 + C)'], \quad C = A_{11}^{-1}A_{12}X_2. \end{aligned}$$

We note that $E(X_1|X_2) = -C = -A_{11}^{-1}A_{12}X_2$: the regression of X_1 on X_2 , the constant part being $|A_{11}|^{\frac{q}{2}}|B|^{\frac{p_1}{2}}/(2\pi)^{\frac{p_1q}{2}}$. Hence the following result:

Theorem 4.5.1. *If the $p \times q$ matrix X has a real matrix-variate Gaussian density with the parameter matrices $M = O$, $A > O$ and $B > O$ where A is $p \times p$ and B is $q \times q$ and if X is partitioned into row sub-matrices $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $p_1 \times q$ and X_2 is $p_2 \times q$, so that $p_1 + p_2 = p$, then the conditional density of X_1 given X_2 , denoted by $f_{p_1,q}(X_1|X_2)$, is given by*

$$f_{p_1,q}(X_1|X_2) = \frac{|A_{11}|^{\frac{q}{2}}|B|^{\frac{p_1}{2}}}{(2\pi)^{\frac{p_1q}{2}}} e^{-\frac{1}{2}\text{tr}[A_{11}(X_1+C)B(X_1+C)']} \quad (4.5.1)$$

where $C = A_{11}^{-1}A_{12}X_2$ if the location parameter is a null matrix; otherwise $C = -M_1 + A_{11}^{-1}A_{12}(X_2 - M_2)$ with M partitioned into row sub-matrices M_1 and M_2 , M_1 being $p_1 \times q$ and M_2 , $p_2 \times q$.

Corollary 4.5.1. *Let X , X_1 , X_2 , M , M_1 and M_2 be as defined in Theorem 4.5.1; then, in the real Gaussian case, the conditional expectation of X_1 given X_2 , denoted by $E(X_1|X_2)$, is*

$$E(X_1|X_2) = M_1 - A_{11}^{-1}A_{12}(X_2 - M_2). \quad (4.5.2)$$

We may adopt the following general notation to represent a real matrix-variate Gaussian (or normal) density:

$$X \sim N_{p,q}(M, A, B), \quad A > O, \quad B > O, \quad (4.5.3)$$

which signifies that the $p \times q$ matrix X has a real matrix-variate Gaussian distribution with location parameter matrix M and parameter matrices $A > O$ and $B > O$ where A is $p \times p$ and B is $q \times q$. Accordingly, the usual q -variate multivariate normal density will be denoted as follows:

$$X_1 \sim N_{1,q}(\mu, B), \quad B > O \Rightarrow X_1' \sim N_q(\mu', B^{-1}), \quad B > O, \quad (4.5.4)$$

where μ is the location parameter vector, which is the first row of M , and X_1 is a $1 \times q$ row vector consisting of the first row of the matrix X . Note that $B^{-1} = \text{Cov}(X_1)$ and the covariance matrix usually appears as the second parameter in the standard notation $N_p(\cdot, \cdot)$. In this case, the 1×1 matrix A will be taken as 1 to be consistent with the usual notation in the real multivariate normal case. The corresponding column case will be denoted as follows:

$$Y_1 \sim N_{p,1}(\mu_{(1)}, A), A > O \Rightarrow Y_1 \sim N_p(\mu_{(1)}, A^{-1}), A > O, A^{-1} = \text{Cov}(Y_1) \quad (4.5.5)$$

where Y_1 is a $p \times 1$ vector consisting of the first column of X and $\mu_{(1)}$ is the first column of M . With this partitioning of X , we have the following result:

Theorem 4.5.2. *Let the real matrices X , M , A and B be as defined in Theorem 4.5.1 and X be partitioned into column sub-matrices as $X = (Y_1 \ Y_2)$ where Y_1 is $p \times q_1$ and Y_2 is $p \times q_2$ with $q_1 + q_2 = q$. Let the density of X , the marginal densities of Y_1 and Y_2 and the conditional density of Y_1 given Y_2 , be respectively denoted by $f_{p,q}(X)$, $f_{p,q_1}(Y_1)$, $f_{p,q_2}(Y_2)$ and $f_{p,q_1}(Y_1|Y_2)$. Then, the conditional density of Y_1 given Y_2 is*

$$f_{p,q_1}(Y_1|Y_2) = \frac{|A|^{\frac{q_1}{2}} |B_{11}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_1}{2}}} e^{-\frac{1}{2} \text{tr}[A(Y_1 - M_{(1)} + C_1)B_{11}(Y_1 - M_{(1)} + C_1)']} \quad (4.5.6)$$

where $A > O$, $B_{11} > O$ and $C_1 = (Y_2 - M_{(2)})B_{21}B_{11}^{-1}$, so that the conditional expectation of Y_1 given Y_2 , or the regression of Y_1 on Y_2 , is obtained as

$$E(Y_1|Y_2) = M_{(1)} - (Y_2 - M_{(2)})B_{21}B_{11}^{-1}, \quad M = (M_{(1)} \ M_{(2)}), \quad (4.5.7)$$

where $M_{(1)}$ is $p \times q_1$ and $M_{(2)}$ is $p \times q_2$ with $q_1 + q_2 = q$. As well, the conditional density of Y_2 given Y_1 is the following:

$$f_{p,q_2}(Y_2|Y_1) = \frac{|A|^{\frac{q_2}{2}} |B_{22}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_2}{2}}} e^{-\frac{1}{2} \text{tr}[A(Y_2 - M_{(2)} + C_2)B_{22}(Y_2 - M_{(2)} + C_2)']} \quad (4.5.8)$$

where

$$M_{(2)} - C_2 = M_{(2)} - (Y_1 - M_{(1)})B_{12}B_{22}^{-1} = E[Y_2|Y_1]. \quad (4.5.9)$$

Example 4.5.1. Consider a 2×3 real matrix $X = (x_{ij})$ having a real matrix-variate Gaussian distribution with the parameters M , $A > O$ and $B > O$ where

$$M = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 0 & -1 \end{bmatrix}, \quad A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 3 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Let X be partitioned as $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = [Y_1, Y_2]$ where $X_1 = [x_{11}, x_{12}, x_{13}]$, $X_2 = [x_{21}, x_{22}, x_{23}]$, $Y_1 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$ and $Y_2 = \begin{bmatrix} x_{13} \\ x_{23} \end{bmatrix}$. Determine the conditional densities of X_1 given X_2 , X_2 given X_1 , Y_1 given Y_2 , Y_2 given Y_1 and the conditional expectations $E[X_1|X_2]$, $E[X_2|X_1]$, $E[Y_1|Y_2]$ and $E[Y_2|Y_1]$.

Solution 4.5.1. Given the specified partitions of X , A and B are partitioned accordingly as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, B_{11} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}, B_{12} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$B_{21} = [1, 0], B_{22} = (1), A_{11} = (2), A_{12} = (1), A_{21} = (1), A_{22} = (3).$$

The following numerical results are needed:

$$\begin{aligned} A_{11} - A_{12}A_{22}^{-1}A_{21} &= 2 - (1)(1/3)(1) = \frac{5}{3} \\ A_{22} - A_{21}A_{11}^{-1}A_{12} &= 3 - (1)(1/2)(1) = \frac{5}{2} \\ B_{11} - B_{12}B_{22}^{-1}B_{21} &= \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix}(1)[1, 0] = \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix} \\ B_{22} - B_{21}B_{11}^{-1}B_{12} &= 1 - [1, 0](1/5) \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{2}{5}; \end{aligned}$$

$$|A_{11}| = 2, |A_{22}| = 3, |A| = 5, |A_{11} - A_{12}A_{22}^{-1}A_{21}| = \frac{5}{3}, |A_{22} - A_{21}A_{11}^{-1}A_{12}| = \frac{5}{2},$$

$$|B_{11}| = 5, |B_{22}| = 1, |B_{11} - B_{12}B_{22}^{-1}B_{21}| = 2, |B_{22} - B_{21}B_{11}^{-1}B_{12}| = \frac{2}{5}, |B| = 2;$$

$$A_{11}^{-1}A_{12} = \frac{1}{2}, A_{22}^{-1}A_{21} = \frac{1}{3}, B_{22}^{-1}B_{21} = [1, 0]$$

$$B_{11}^{-1}B_{12} = \frac{1}{5} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 \\ 1 \end{bmatrix}, M_1 = [1, -1, 1], M_2 = [2, 0, -1]$$

$$M_{(1)} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}, M_{(2)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

All the conditional expectations can now be determined. They are

$$\begin{aligned} E[X_1|X_2] &= M_1 - A_{11}^{-1}A_{12}(X_2 - M_2) = [1, -1, 1] - \frac{1}{2}(x_{21} - 2, x_{22}, x_{23} + 1) \\ &= [1 - \frac{1}{2}(x_{21} - 2), -1 - \frac{1}{2}x_{22}, 1 - \frac{1}{2}(x_{23} + 1)] \end{aligned} \quad (i)$$

$$\begin{aligned} E[X_2|X_1] &= M_2 - A_{22}^{-1}A_{21}(X_1 - M_1) = [2, 0, -1] - \frac{1}{3}[x_{11} - 1, x_{12} + 1, x_{13} - 1] \\ &= [2 - \frac{1}{3}(x_{11} - 1), -\frac{1}{3}(x_{12} + 1), -1 - \frac{1}{3}(x_{13} - 1)]; \end{aligned} \quad (ii)$$

$$\begin{aligned} E[Y_1|Y_2] &= M_{(1)} - (Y_2 - M_{(2)})B_{21}B_{11}^{-1} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} x_{13} - 1 \\ x_{23} + 1 \end{bmatrix} [1, 0] \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 - \frac{3}{5}(x_{13} - 1) & -1 - \frac{1}{5}(x_{13} - 1) \\ 2 - \frac{3}{5}(x_{23} + 1) & -\frac{1}{5}(x_{23} + 1) \end{bmatrix} \end{aligned} \quad (iii)$$

$$\begin{aligned} E[Y_2|Y_1] &= M_{(2)} - (Y_1 - M_{(1)})B_{12}B_{22}^{-1} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} x_{11} - 1 & x_{12} + 1 \\ x_{21} - 2 & x_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} [(1)] = \begin{bmatrix} 2 - x_{11} \\ 1 - x_{21} \end{bmatrix}. \end{aligned} \quad (iv)$$

The conditional densities can now be obtained. That of X_1 given X_2 is

$$f_{p_1, q}(X_1|X_2) = \frac{|A_{11}|^{\frac{q}{2}}|B|^{\frac{p_1}{2}}}{(2\pi)^{\frac{p_1 q}{2}}} e^{-\frac{1}{2}\text{tr}[A_{11}(X_1 - M_1 + C)B(X_1 - M_1 + C)']}$$

for the matrices $A > O$ and $B > O$ previously specified; that is,

$$f_{1,3}(X_1|X_2) = \frac{4}{(2\pi)^{\frac{3}{2}}} e^{-\frac{3}{2}(X_1 - M_1 + C)B(X_1 - M_1 + C)'}$$

where $M_1 - C = E[X_1|X_2]$ is given in (i). The conditional density of $X_2|X_1$ is the following:

$$f_{p_2, q}(X_2|X_1) = \frac{|A_{22}|^{\frac{q}{2}}|B|^{\frac{p_2}{2}}}{(2\pi)^{\frac{p_2 q}{2}}} e^{-\frac{1}{2}\text{tr}[A_{22}(X_2 - M_2 + C_1)B(X_2 - M_2 + C_1)']},$$

that is,

$$f_{1,3}(X_2|X_1) = \frac{(3^{\frac{3}{2}})(2^{\frac{1}{2}})}{(2\pi)^{\frac{3}{2}}} e^{-\frac{3}{2}(X_2 - M_2 + C_1)B(X_2 - M_2 + C_2)'}$$

where $M_2 - C_2 = E[X_2|X_1]$ is given in (ii). The conditional density of Y_1 given Y_2 is

$$f_{p,q_1}(Y_1|Y_2) = \frac{|A|^{\frac{q_1}{2}} |B_{11}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_1}{2}}} e^{-\frac{1}{2}\text{tr}[A(Y_1 - M_{(1)} + C_2)B_{11}(Y_1 - M_{(1)} + C_2)']},$$

that is,

$$f_{2,2}(Y_1|Y_2) = \frac{25}{(2\pi)^2} e^{-\frac{1}{2}\text{tr}[A(Y_1 - M_{(1)} + C_2)B_{11}(Y_1 - M_{(1)} + C_2)']}$$

where $M_{(1)} - C_1 = E[Y_1|Y_2]$ is specified in (iii). Finally, the conditional density of $Y_2|Y_1$ is the following:

$$f_{p,q_2}(Y_2|Y_1) = \frac{|A|^{\frac{q_2}{2}} |B_{22}|^{\frac{p}{2}}}{(2\pi)^{\frac{pq_2}{2}}} e^{-\frac{1}{2}\text{tr}[A(Y_2 - M_{(2)} + C_3)B_{22}(Y_2 - M_{(2)} + C_3)']},$$

that is,

$$f_{2,1}(Y_2|Y_1) = \frac{\sqrt{5}}{(2\pi)} e^{-\text{tr}[A(Y_2 - M_{(2)} + C_3)B_{22}(Y_2 - M_{(2)} + C_3)']}$$

where $M_{(2)} - C_3 = E[Y_2|Y_1]$ is given in (iv). This completes the computations.

4.5a. Conditional Densities in the Matrix-variate Complex Gaussian Case

The corresponding distributions in the complex case closely parallel those obtained for the real case. A tilde will be utilized to distinguish them from the real distributions. Thus,

$$\tilde{X} \sim \tilde{N}_{p,q}(\tilde{M}, A, B), \quad A = A^* > O, \quad B = B^* > O$$

will denote a complex $p \times q$ matrix \tilde{X} having a $p \times q$ matrix-variate complex Gaussian density. For the $1 \times q$ case, that is, the q -variate multivariate normal distribution in the complex case, which is obtained from the marginal distribution of the first row of \tilde{X} , we have

$$\tilde{X}_1 \sim \tilde{N}_{1,q}(\mu, B), \quad B > O, \quad \tilde{X}_1 \sim \tilde{N}_q(\mu, B^{-1}), \quad B^{-1} = \text{Cov}(\tilde{X}_1),$$

where \tilde{X}_1 is $1 \times q$ vector having a q -variate complex normal density with $E(\tilde{X}_1) = \mu$. The case $q = 1$ corresponds to a column vector in \tilde{X} , which constitutes a $p \times 1$ column vector in the complex domain. Letting it be denoted as \tilde{Y}_1 , we have

$$\tilde{Y}_1 \sim \tilde{N}_{p,1}(\mu_{(1)}, A), \quad A > O, \quad \text{that is, } \tilde{Y}_1 \sim \tilde{N}_p(\mu_{(1)}, A^{-1}), \quad A^{-1} = \text{Cov}(\tilde{Y}_1),$$

where $\mu_{(1)}$ is the first column of M .

Theorem 4.5a.1. Let \tilde{X} be $p \times q$ matrix in the complex domain having a $p \times q$ matrix-variate complex Gaussian density denoted by $\tilde{f}_{p,q}(\tilde{X})$. Let $\tilde{X} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix}$ be a row partitioning of \tilde{X} into sub-matrices where \tilde{X}_1 is $p_1 \times q$ and \tilde{X}_2 is $p_2 \times q$, with $p_1 + p_2 = p$. Then the conditional density of \tilde{X}_1 given \tilde{X}_2 denoted by $\tilde{f}_{p_1,q}(\tilde{X}_1|\tilde{X}_2)$, is given by

$$\tilde{f}_{p_1,q}(\tilde{X}_1|\tilde{X}_2) = \frac{|\det(A_{11})|^q |\det(B)|^{p_1}}{\pi^{p_1 q}} e^{-\text{tr}[A_{11}(\tilde{X}_1 - M_1 + \tilde{C})B(\tilde{X}_1 - M_1 + \tilde{C})^*]} \quad (4.5a.1)$$

where $\tilde{C} = A_{11}^{-1}A_{12}(\tilde{X}_2 - M_2)$, $E[\tilde{X}] = M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$, and the regression of \tilde{X}_1 on \tilde{X}_2 is as follows:

$$E(\tilde{X}_1|\tilde{X}_2) = \begin{cases} M_1 - A_{11}^{-1}A_{12}(\tilde{X}_2 - M_2) & \text{if } M = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \\ -A_{11}^{-1}A_{12}\tilde{X}_2 & \text{if } M = O. \end{cases} \quad (4.5a.2)$$

Analogously, the conditional density of \tilde{X}_2 given \tilde{X}_1 is

$$\tilde{f}_{p_2,q}(\tilde{X}_2|\tilde{X}_1) = \frac{|\det(A_{22})|^q |\det(B)|^{p_2}}{\pi^{p_2 q}} e^{-\text{tr}[A_{22}(\tilde{X}_2 - M_2 + C_1)B(\tilde{X}_2 - M_2 + C_1)^*]} \quad (4.5a.3)$$

where $C_1 = A_{22}^{-1}A_{21}(\tilde{X}_1 - M_1)$, so that the conditional expectation of \tilde{X}_2 given \tilde{X}_1 or the regression of \tilde{X}_2 on \tilde{X}_1 is given by

$$E[\tilde{X}_2|\tilde{X}_1] = M_2 - A_{22}^{-1}A_{21}(\tilde{X}_1 - M_1). \quad (4.5a.4)$$

Theorem 4.5a.2. Let \tilde{X} be as defined in Theorem 4.5a.1. Let \tilde{X} be partitioned into column submatrices, that is, $\tilde{X} = (\tilde{Y}_1 \tilde{Y}_2)$ where \tilde{Y}_1 is $p \times q_1$ and \tilde{Y}_2 is $p \times q_2$, with $q_1 + q_2 = q$. Then the conditional density of \tilde{Y}_1 given \tilde{Y}_2 , denoted by $\tilde{f}_{p,q_1}(\tilde{Y}_1|\tilde{Y}_2)$ is given by

$$\tilde{f}_{p,q_1}(\tilde{Y}_1|\tilde{Y}_2) = \frac{|\det(A)|^{q_1} |\det(B_{11})|^p}{\pi^{p q_1}} e^{-\text{tr}[A(\tilde{Y}_1 - M_{(1)} + \tilde{C}_{(1)})B_{11}(\tilde{Y}_1 - M_{(1)} + \tilde{C}_{(1)})]} \quad (4.5a.5)$$

where $\tilde{C}_{(1)} = (\tilde{Y}_2 - M_{(2)})B_{21}B_{11}^{-1}$, and the regression of \tilde{Y}_1 on \tilde{Y}_2 or the conditional expectation of \tilde{Y}_1 given \tilde{Y}_2 is given by

$$E(\tilde{Y}_1|\tilde{Y}_2) = M_{(1)} - (\tilde{Y}_2 - M_{(2)})B_{21}B_{11}^{-1} \quad (4.5a.6)$$

with $E[\tilde{X}] = M = [M_{(1)} \ M_{(2)}] = E[\tilde{Y}_1 \ \tilde{Y}_2]$. As well the conditional density of \tilde{Y}_2 given \tilde{Y}_1 is the following:

$$\tilde{f}_{p,q_2}(\tilde{Y}_2|\tilde{Y}_1) = \frac{|\det(A)|^{q_2} |\det(B_{22})|^p}{\pi^{pq_2}} e^{-\text{tr}[A(\tilde{Y}_2 - M_{(2)} + C_{(2)})B_{22}(\tilde{Y}_2 - M_{(2)} + C_{(2)})^*]} \quad (4.5a.7)$$

where $C_{(2)} = (\tilde{Y}_1 - M_{(1)})B_{12}B_{22}^{-1}$ and the conditional expectation of \tilde{Y}_2 given \tilde{Y}_1 is then

$$E[\tilde{Y}_2|\tilde{Y}_1] = M_{(2)} - (\tilde{Y}_1 - M_{(1)})B_{12}B_{22}^{-1}. \quad (4.5a.8)$$

Example 4.5a.1. Consider a 2×3 matrix-variate complex Gaussian distribution with the parameters

$$A = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -i & 0 \\ i & 2 & i \\ 0 & -i & 1 \end{bmatrix}, \quad M = E[\tilde{X}] = \begin{bmatrix} 1+i & i & -i \\ i & 2+i & 1-i \end{bmatrix}.$$

Consider the partitioning of $\tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix} = [\tilde{Y}_1 \ \tilde{Y}_2]$ where $\tilde{X}_1 = [\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{13}]$, $\tilde{X}_2 = [\tilde{x}_{21}, \tilde{x}_{22}, \tilde{x}_{23}]$, $\tilde{Y}_1 = \begin{bmatrix} \tilde{x}_{11} \\ \tilde{x}_{21} \end{bmatrix}$ and $\tilde{Y}_2 = \begin{bmatrix} \tilde{x}_{12} & \tilde{x}_{13} \\ \tilde{x}_{22} & \tilde{x}_{23} \end{bmatrix}$. Determine the conditional densities of $\tilde{X}_1|\tilde{X}_2$, $\tilde{X}_2|\tilde{X}_1$, $\tilde{Y}_1|\tilde{Y}_2$ and $\tilde{Y}_2|\tilde{Y}_1$ and the corresponding conditional expectations.

Solution 4.5a.1. As per the partitioning of \tilde{X} , we have the following partitions of A , B and M :

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix}, \quad B_{22}^{-1} = \begin{bmatrix} 1 & -i \\ i & 2 \end{bmatrix}, \quad B_{21} = \begin{bmatrix} i \\ 0 \end{bmatrix},$$

$$A_{11} = (2), \quad A_{12} = (i), \quad A_{21} = (-i), \quad A_{22} = (1), \quad B_{12} = [-i, 0], \quad A_{11}^{-1} = \frac{1}{2}, \quad A_{22}^{-1} = 1, \quad B_{11}^{-1} = \frac{1}{3},$$

$$A_{11} - A_{12}A_{22}^{-1}A_{21} = 2 - (i)(1)(-i) = 1, \quad |A_{11} - A_{12}A_{22}^{-1}A_{21}| = 1,$$

$$A_{22} - A_{21}A_{11}^{-1}A_{12} = \frac{1}{2}, \quad |A_{22} - A_{21}A_{11}^{-1}A_{12}| = \frac{1}{2}, \quad |A| = 1, \quad |B| = 2,$$

$$B_{11} - B_{12}B_{22}^{-1}B_{21} = 3 - [-i \ 0] \begin{bmatrix} 1 & -i \\ i & 2 \end{bmatrix} \begin{bmatrix} i \\ 0 \end{bmatrix} = 2, \quad |B_{11} - B_{12}B_{22}^{-1}B_{21}| = 2,$$

$$B_{22} - B_{21}B_{11}^{-1}B_{12} = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix} - \begin{bmatrix} i \\ 0 \end{bmatrix} (1/3)[-i \ 0] = \begin{bmatrix} \frac{5}{3} & i \\ -i & 1 \end{bmatrix}, \quad |B_{22} - B_{21}B_{11}^{-1}B_{12}| = \frac{2}{3}.$$

$$M_1 = [1+i, i, -i], \quad M_2 = [i, 2+i, 1-i], \quad M_{(1)} = \begin{bmatrix} 1+i \\ i \end{bmatrix}, \quad M_{(2)} = \begin{bmatrix} i & -i \\ 2+i & 1-i \end{bmatrix}.$$

All the conditional expectations can now be determined. They are

$$E[\tilde{X}_1|\tilde{X}_2] = M_1 - A_{11}^{-1}A_{12}(\tilde{X}_2 - M_2) = [1 + i, i, -i] - \frac{i}{2}(\tilde{X}_2 - M_2) \quad (i)$$

$$E[\tilde{X}_2|\tilde{X}_1] = M_2 - A_{22}^{-1}A_{21}(\tilde{X}_1 - M_1) = [i, 2 + i, 1 - i] + i(\tilde{X}_1 - M_1) \quad (ii)$$

$$E[\tilde{Y}_1|\tilde{Y}_2] = M_{(1)} - (\tilde{Y}_2 - M_{(2)})B_{21}B_{11}^{-1} = \frac{1}{3} \begin{bmatrix} 2 + i\tilde{x}_{12} \\ -1 + 5i - i\tilde{x}_{22} \end{bmatrix} \quad (iii)$$

$$\begin{aligned} E[\tilde{Y}_2|\tilde{Y}_1] &= M_{(2)} - (\tilde{Y}_1 - M_{(1)})B_{12}B_{22}^{-1} \\ &= \begin{bmatrix} i & -i \\ 2 + i & 1 - i \end{bmatrix} - \begin{bmatrix} \tilde{x}_{11} - (1 + i) \\ \tilde{x}_{21} - i \end{bmatrix} \begin{bmatrix} -i & 0 \end{bmatrix} \begin{bmatrix} 1 & -i \\ -i & 2 \end{bmatrix} \\ &= \begin{bmatrix} i\tilde{x}_{11} + 1 & \tilde{x}_{11} - 1 - 2i \\ -i\tilde{x}_{21} + i + 3 & \tilde{x}_{21} + 1 - 2i \end{bmatrix}. \end{aligned} \quad (iv)$$

Now, on substituting the above quantities in equations (4.5a.1), (4.5a.3), (4.5a.5) and (4.5a.7), the following densities are obtained:

$$\tilde{f}_{1,3}(\tilde{X}_1|\tilde{X}_2) = \frac{2^4}{\pi^3} e^{-2(\tilde{X}_1 - E_1)B(\tilde{X}_1 - E_1)^*}$$

where $E_1 = E[\tilde{X}_1|\tilde{X}_2]$ given in (i);

$$\tilde{f}_{1,3}(\tilde{X}_2|\tilde{X}_1) = \frac{2}{\pi^3} e^{-(\tilde{X}_2 - E_2)B(\tilde{X}_2 - E_2)^*}$$

where $E_2 = E[\tilde{X}_2|\tilde{X}_1]$ given in (ii);

$$\tilde{f}_{2,1}(\tilde{Y}_1|\tilde{Y}_2) = \frac{3^2}{\pi^2} e^{-3\text{tr}[A(\tilde{Y}_1 - E_3)(\tilde{Y}_1 - E_3)^*]}$$

where $E_3 = E[\tilde{Y}_1|\tilde{Y}_2]$ given in (iii);

$$\tilde{f}_{2,2}(\tilde{Y}_2|\tilde{Y}_1) = \frac{1}{\pi^4} e^{-\text{tr}[A(\tilde{Y}_2 - E_4)B_{22}(\tilde{Y}_2 - E_4)^*]}$$

where $E_4 = E[\tilde{Y}_2|\tilde{Y}_1]$ given in (iv). The exponent in the density of $\tilde{Y}_1|\tilde{Y}_2$ can be simplified as follows:

$$\begin{aligned} -\text{tr}[A(\tilde{Y}_1 - M_{(1)})B_{11}(\tilde{Y}_1 - M_{(1)})^*] &= -3(\tilde{Y}_1 - M_{(1)})^*A(\tilde{Y}_1 - M_{(1)}) \\ &= -3[(\tilde{x}_{11} - (1 + i))^* (\tilde{x}_{21} - i)^*] \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix} \begin{bmatrix} (\tilde{x}_{11} - (1 + i)) \\ (\tilde{x}_{21} - i) \end{bmatrix} \\ &= -6\{(x_{111}^2 + x_{112}^2) + \frac{1}{2}(x_{211}^2 + x_{212}^2) + (x_{112}x_{211} - x_{111}x_{212}) - 2x_{112} - x_{111} - x_{211} + \frac{3}{2}\} \end{aligned}$$

by writing $\tilde{x}_{k1} = x_{k11} + ix_{k12}$, $k = 1, 2$, $i = \sqrt{-1}$. This completes the computations.

4.5.1. Re-examination of the case $q = 1$

When $q = 1$, we have a $p \times 1$ vector-variate or the usual p -variate Gaussian density of the form in (4.5.5). Let us consider the real case first. Let the $p \times 1$ vector be denoted by Y_1 with

$$Y_1 = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} Y_{(1)} \\ Y_{(2)} \end{bmatrix}, \quad Y_{(1)} = \begin{bmatrix} y_1 \\ \vdots \\ y_{p_1} \end{bmatrix}, \quad Y_{(2)} = \begin{bmatrix} y_{p_1+1} \\ \vdots \\ y_p \end{bmatrix};$$

$$M_{(1)} = \begin{bmatrix} M_{(1)}^{(p_1)} \\ M_{(1)}^{(p_2)} \\ M_{(2)} \end{bmatrix}, \quad M_{(1)}^{(p_1)} = \begin{bmatrix} m_1 \\ \vdots \\ m_{p_1} \end{bmatrix}, \quad M_{(2)}^{(p_2)} = \begin{bmatrix} m_{p_1+1} \\ \vdots \\ m_p \end{bmatrix}, \quad E[Y_1] = M_{(1)}, \quad p_1 + p_2 = p.$$

Then, from (4.5.2) wherein $q = 1$, we have

$$E[Y_{(1)}|Y_{(2)}] = M_{(1)}^{(p_1)} - A_{11}^{-1}A_{12}(Y_{(2)} - M_{(2)}^{(p_2)}), \quad (4.5.10)$$

with $A = \Sigma^{-1}$, Σ being the covariance matrix of Y_1 , that is, $\text{Cov}(Y_1) = E[(Y_1 - E(Y_1))(Y_1 - E(Y_1))']$. Let

$$A^{-1} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \text{where } \Sigma_{11} \text{ is } p_1 \times p_1 \text{ and } \Sigma_{22} \text{ is } p_2 \times p_2.$$

From the partitioning of matrices presented in Sect. 1.3, we have

$$-A_{11}^{-1}A_{12} = A^{12}(A^{22})^{-1} = \Sigma_{12}\Sigma_{22}^{-1}. \quad (4.5.11)$$

Accordingly, we may rewrite (4.5.10) in terms of the sub-matrices of the covariance matrix as

$$E[Y_{(1)}|Y_{(2)}] = M_{(1)}^{(p_1)} + \Sigma_{12}\Sigma_{22}^{-1}(Y_{(2)} - M_{(2)}^{(p_2)}). \quad (4.5.12)$$

If $p_1 = 1$, then $Y_{(2)}$ will contain $p - 1$ elements, denoted by $Y'_{(2)} = (y_2, \dots, y_p)$. Letting $E[y_1] = m_1$, we have

$$E[y_1|Y_{(2)}] = m_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_{(2)} - M_{(2)}^{(p_2)}), \quad p_2 = p - 1. \quad (4.5.13)$$

The conditional expectation (4.5.13) is *the best predictor of y_1 at the preassigned values of y_2, \dots, y_p* , where $m_1 = E[y_1]$. It will now be shown that $\Sigma_{12}\Sigma_{22}^{-1}$ can be expressed in terms of variances and correlations. Let $\sigma_j^2 = \sigma_{jj} = \text{Var}(y_j)$ where $\text{Var}(\cdot)$ denotes the

variance of (\cdot) . Note that $\sigma_{ij} = \text{Cov}(y_i, y_j)$ or the covariance between y_1 and y_j . Letting ρ_{ij} be the correlation between y_i and y_j , we have

$$\begin{aligned}\Sigma_{12} &= [\text{Cov}(y_1, y_2), \dots, \text{Cov}(y_1, y_p)] \\ &= [\sigma_1\sigma_2\rho_{12}, \dots, \sigma_1\sigma_p\rho_{1p}].\end{aligned}$$

Then

$$\Sigma = \begin{bmatrix} \sigma_1\sigma_1 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2\sigma_2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_p\sigma_1\rho_{p1} & \sigma_p\sigma_2\rho_{p2} & \cdots & \sigma_p\sigma_p \end{bmatrix}, \quad \rho_{ij} = \rho_{ji}, \quad \rho_{jj} = 1,$$

for all j . Let $D = \text{diag}(\sigma_1, \dots, \sigma_p)$ be a diagonal matrix whose diagonal elements are $\sigma_1, \dots, \sigma_p$, the standard deviations of y_1, \dots, y_p , respectively. Letting $R = (\rho_{ij}) =$ denote the correlation matrix wherein ρ_{ij} is the correlation between y_i and y_j , we can express Σ as DRD , that is,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}$$

so that

$$\Sigma^{-1} = D^{-1}R^{-1}D^{-1}, \quad p = 2, 3, \dots \quad (4.5.14)$$

We can then re-express (4.5.13) in terms of variances and correlations since

$$\Sigma_{12}\Sigma_{22}^{-1} = \sigma_1 R_{12} D_{(2)} D_{(2)}^{-1} R_{22}^{-1} D_{(2)}^{-1} = \sigma_1 R_{12} R_{22}^{-1} D_{(2)}^{-1}$$

where $D_{(2)} = \text{diag}(\sigma_2, \dots, \sigma_p)$ and R is partitioned accordingly. Thus,

$$E[y_1|Y_{(2)}] = m_1 + \sigma_1 R_{12} R_{22}^{-1} D_{(2)}^{-1} (Y_{(2)} - M_{(2)}^{(p2)}). \quad (4.5.15)$$

An interesting particular case occurs when $p = 2$, as there are then only two real scalar variables y_1 and y_2 , and

$$E[y_1|y_2] = m_1 + \frac{\sigma_1}{\sigma_2} \rho_{12} (y_2 - m_2), \quad (4.5.16)$$

which is the *regression of y_1 on y_2 or the best predictor of y_1 at a given value of y_2 .*

4.6. Sampling from a Real Matrix-variate Gaussian Density

Let the $p \times q$ matrix $X_\alpha = (x_{ij\alpha})$ have a $p \times q$ real matrix-variate Gaussian density with parameter matrices M , $A > O$ and $B > O$. When n independently and identically distributed (iid) matrix random variables that are distributed as X_α are available, we say that we have a *simple random sample of size n from X_α or from the population distributed as X_α* . We will consider simple random samples from a $p \times q$ matrix-variate Gaussian population in the real and complex domains. Since the procedures are parallel to those utilized in the vector variable case, we will recall the particulars in connection with that particular case. Some of the following materials are re-examinations of those already presented Chap. 3. For $q = 1$, we have a p -vector which will be denoted by Y_1 . In our previous notations, Y_1 is the same Y_1 for $q_1 = 1$, $q_2 = 0$ and $q = 1$. Consider a sample of size n from a population distributed as Y_1 and let the $p \times n$ sample matrix be denoted by \mathbf{Y} . Then,

$$\mathbf{Y} = [Y_1, \dots, Y_n] = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pn} \end{bmatrix}, \quad Y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{p1} \end{bmatrix}.$$

In this case, the columns of \mathbf{Y} , that is, Y_j , $j = 1, \dots, n$, are iid variables, distributed as Y_1 . Let an $n \times 1$ column vector whose components are all equal to 1 be denoted by J and consider

$$\bar{\mathbf{Y}} = \frac{1}{n} \mathbf{Y} J = \frac{1}{n} \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ y_{21} & \cdots & y_{2n} \\ \vdots & \ddots & \vdots \\ y_{p1} & \cdots & y_{pn} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_p \end{bmatrix}$$

where $\bar{y}_j = \frac{\sum_{k=1}^n y_{jk}}{n}$ denotes the average of the variables, distributed as y_j . Let

$$S = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})' \text{ where the bold-faced } \bar{\mathbf{Y}} = \begin{bmatrix} \bar{y}_1 & \cdots & \bar{y}_1 \\ \bar{y}_2 & \cdots & \bar{y}_2 \\ \vdots & \ddots & \vdots \\ \bar{y}_p & \cdots & \bar{y}_p \end{bmatrix} = [\bar{Y}, \dots, \bar{Y}].$$

Then,

$$S = (s_{ij}), \quad s_{ij} = \sum_{k=1}^n (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j) \text{ for all } i \text{ and } j. \quad (4.6.1)$$

This matrix S is known as the *sample sum of products matrix or corrected sample sum of products matrix*. Here “corrected” indicates that the deviations are taken from the respec-

tive averages $\bar{y}_1, \dots, \bar{y}_p$. Note that $\frac{1}{n}s_{ij}$ is equal to the sample covariance between y_i and y_j and when $i = j$, it is the sample variance of y_i . Observing that

$$J = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \Rightarrow JJ' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \text{ and } J'J = n,$$

we have

$$\mathbf{Y}\left(\frac{1}{n}JJ'\right) = \bar{\mathbf{Y}} \Rightarrow \mathbf{Y} - \bar{\mathbf{Y}} = \mathbf{Y}\left[I - \frac{1}{n}JJ'\right].$$

Hence

$$S = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})' = \mathbf{Y}\left[I - \frac{1}{n}JJ'\right]\left[I - \frac{1}{n}JJ'\right]'\mathbf{Y}'.$$

However,

$$\begin{aligned} \left[I - \frac{1}{n}JJ'\right]\left[I - \frac{1}{n}JJ'\right]' &= I - \frac{1}{n}JJ' - \frac{1}{n}JJ' + \frac{1}{n^2}JJ'JJ' \\ &= I - \frac{1}{n}JJ' \text{ since } J'J = n. \end{aligned}$$

Thus,

$$S = \mathbf{Y}\left[I - \frac{1}{n}JJ'\right]\mathbf{Y}'. \quad (4.6.2)$$

Letting $C_1 = (I - \frac{1}{n}JJ')$, we note that $C_1^2 = C_1$ and that the rank of C_1 is $n - 1$. Accordingly, C_1 is an idempotent matrix having $n - 1$ eigenvalues equal to 1, the remaining one being equal to zero. Now, letting $C_2 = \frac{1}{n}JJ'$, it is easy to verify that $C_2^2 = C_2$ and that the rank of C_2 is one; thus, C_2 is idempotent with $n - 1$ eigenvalues equal to zero, the remaining one being equal to 1. Further, since $C_1C_2 = O$, that is, C_1 and C_2 are orthogonal to each other, $\mathbf{Y} - \bar{\mathbf{Y}} = \mathbf{Y}C_1$ and $\bar{\mathbf{Y}} = \mathbf{Y}C_2$ are independently distributed, so that $\mathbf{Y} - \bar{\mathbf{Y}}$ and $\bar{\mathbf{Y}}$ are independently distributed. Consequently, $S = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})'$ and $\bar{\mathbf{Y}}$ are independently distributed as well. This will be stated as the next result.

Theorem 4.6.1, 4.6a.1. *Let Y_1, \dots, Y_n be a simple random sample of size n from a p -variate real Gaussian population having a $N_p(\mu, \Sigma)$, $\Sigma > O$, distribution. Let \bar{Y} be the sample average and S be the sample sum of products matrix; then, \bar{Y} and S are statistically independently distributed. In the complex domain, let the \tilde{Y}_j 's be iid $N_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} =$*

$\tilde{\Sigma}^* > O$, and \tilde{Y} and \tilde{S} denote the sample average and sample sum of products matrix; then, \tilde{Y} and \tilde{S} are independently distributed.

4.6.1. The distribution of the sample sum of products matrix, real case

Reprising the notations of Sect. 4.6, let the $p \times n$ matrix \mathbf{Y} denote a sample matrix whose columns Y_1, \dots, Y_n are iid as $N_p(\mu, \Sigma)$, $\Sigma > O$, Gaussian vectors. Let the sample mean be $\bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n) = \frac{1}{n}\mathbf{Y}J$ where $J' = (1, \dots, 1)$. Let the bold-faced matrix $\bar{\mathbf{Y}} = [\bar{Y}, \dots, \bar{Y}] = \mathbf{Y}C_1$ where $C_1 = I_n - \frac{1}{n}JJ'$. Note that $C_1 = I_n - C_2 = C_1^2$ and $C_2 = \frac{1}{n}JJ' = C_2^2$, that is, C_1 and C_2 are idempotent matrices whose respective ranks are $n - 1$ and 1. Since $C_1 = C_1'$, there exists an $n \times n$ orthonormal matrix P , $PP' = I_n$, $P'P = I_n$, such that $P'C_1P = D$ where

$$D = \begin{bmatrix} I_{n-1} & O \\ O & 0 \end{bmatrix} = P'C_1P.$$

Let $\mathbf{Y} = ZP'$ where Z is $p \times n$. Then, $\mathbf{Y} = ZP' \Rightarrow \mathbf{Y}C_1 = ZP'C_1 = ZP'PDP' = ZDP'$, so that

$$\begin{aligned} S &= (\mathbf{Y}C_1)(\mathbf{Y}C_1)' = \mathbf{Y}C_1C_1'\mathbf{Y}' = Z \begin{bmatrix} I_{n-1} & O \\ O & 0 \end{bmatrix} \begin{bmatrix} I_{n-1} & O \\ O & 0 \end{bmatrix} Z' \\ &= (Z_{n-1}, O)(Z_{n-1}, O)' = Z_{n-1}Z_{n-1}' \end{aligned} \quad (4.6.3)$$

where Z_{n-1} is a $p \times (n-1)$ matrix obtained by deleting the last column of the $p \times n$ matrix Z . Thus, $S = Z_{n-1}Z_{n-1}'$ where Z_{n-1} contains $p(n-1)$ distinct real variables. Accordingly, Theorems 4.2.1, 4.2.2, 4.2.3, and the analogous results in the complex domain, are applicable to Z_{n-1} as well as to the corresponding quantity \tilde{Z}_{n-1} in the complex case. Observe that when $Y_1 \sim N_p(\mu, \Sigma)$, $\mathbf{Y} - \bar{\mathbf{Y}}$ has expected value $\mathbf{M} - \mathbf{M} = O$, $\mathbf{M} = (\mu, \dots, \mu)$. Hence, $\mathbf{Y} - \bar{\mathbf{Y}} = (\mathbf{Y} - \mathbf{M}) - (\bar{\mathbf{Y}} - \mathbf{M})$ and therefore, without any loss of generality, we can assume Y_1 to be coming from a $N_p(O, \Sigma)$, $\Sigma > O$, vector random variable whenever $\mathbf{Y} - \bar{\mathbf{Y}}$ is involved.

Theorem 4.6.2. *Let \mathbf{Y} , \bar{Y} , $\bar{\mathbf{Y}}$, J , C_1 and C_2 be as defined in this section. Then, the $p \times n$ matrix $(\mathbf{Y} - \bar{\mathbf{Y}})J = O$, which implies that there exist linear relationships among the columns of \mathbf{Y} . However, all the elements of Z_{n-1} as defined in (4.6.3) are distinct real variables. Thus, Theorems 4.2.1, 4.2.2 and 4.2.3 are applicable to Z_{n-1} .*

Note that the corresponding result for the complex Gaussian case also holds.

4.6.2. Linear functions of sample vectors

Let $Y_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, or equivalently, let the Y_j 's constitutes a simple random sample of size n from this p -variate real Gaussian population. Then, the density of the $p \times n$ sample matrix \mathbf{Y} , denoted by $L(\mathbf{Y})$, is the following:

$$L(\mathbf{Y}) = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}[\Sigma^{-1}(\mathbf{Y}-\mathbf{M})(\mathbf{Y}-\mathbf{M})']},$$

where $\mathbf{M} = (\mu, \dots, \mu)$ is $p \times n$ whose columns are all equal to the $p \times 1$ parameter vector μ . Consider a linear function of the sample values Y_1, \dots, Y_n . Let the linear function be $U = \mathbf{Y}A$ where A is an $n \times q$ constant matrix of rank q , $q \leq p \leq n$, so that U is $p \times q$. Let us consider the mgf of U . Since U is $p \times q$, we employ a $q \times p$ parameter matrix T so that $\text{tr}(TU)$ will contain all the elements in U multiplied by the corresponding parameters. The mgf of U is then

$$\begin{aligned} M_U(T) &= E[e^{\text{tr}(TU)}] = E[e^{\text{tr}(T\mathbf{Y}A)}] = E[e^{\text{tr}(AT\mathbf{Y})}] \\ &= e^{\text{tr}(AT\mathbf{M})} E[e^{\text{tr}(AT(\mathbf{Y}-\mathbf{M}))}] \end{aligned}$$

where $\mathbf{M} = (\mu, \dots, \mu)$. Letting $W = \Sigma^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{M})$, $d\mathbf{Y} = |\Sigma|^{\frac{n}{2}} dW$ and

$$\begin{aligned} M_U(T) &= e^{\text{tr}(AT\mathbf{M})} |\Sigma|^{\frac{n}{2}} E[e^{\text{tr}(AT\Sigma^{\frac{1}{2}}W)}] \\ &= \frac{e^{\text{tr}(AT\mathbf{M})}}{(2\pi)^{\frac{np}{2}}} \int_W e^{\text{tr}(AT\Sigma^{\frac{1}{2}}W) - \frac{1}{2}\text{tr}(WW')} dW. \end{aligned}$$

Now, expanding

$$\text{tr}[(W - C)(W - C)'] = \text{tr}(WW') - 2\text{tr}(WC') + \text{tr}(CC').$$

and comparing the resulting expression with the exponent in the integrand, which excluding $-\frac{1}{2}$, is $\text{tr}(WW') - 2\text{tr}(AT\Sigma^{\frac{1}{2}}W)$, we may let $C' = AT\Sigma^{\frac{1}{2}}$ so that $\text{tr}(CC') = \text{tr}(AT\Sigma T'A') = \text{tr}(T\Sigma T'A'A)$. Since $\text{tr}(AT\mathbf{M}) = \text{tr}(T\mathbf{M}A)$ and

$$\frac{1}{(2\pi)^{\frac{np}{2}}} \int_W e^{-\frac{1}{2}((W-C)(W-C)')} dW = 1,$$

we have

$$M_U(T) = M_{\mathbf{Y}A}(T) = e^{\text{tr}(T\mathbf{M}A) + \frac{1}{2}\text{tr}(T\Sigma T'A'A)}$$

where $\mathbf{M}A = E[\mathbf{Y}A]$, $\Sigma > O$, $A'A > O$, A being a full rank matrix, and $T\Sigma T'A'A$ is a $q \times q$ positive definite matrix. Hence, the $p \times q$ matrix $U = \mathbf{Y}A$ has a matrix-variate

real Gaussian density with the parameters $\mathbf{MA} = E[\mathbf{YA}]$ and $A'A > O$, $\Sigma > O$. Thus, the following result:

Theorem 4.6.3, 4.6a.2. *Let $Y_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, or equivalently, let the Y_j 's constitutes a simple random sample of size n from this p -variate real Gaussian population. Consider a set of linear functions of Y_1, \dots, Y_n , $U = \mathbf{YA}$ where $\mathbf{Y} = (Y_1, \dots, Y_n)$ is a $p \times n$ sample matrix and A is an $n \times q$ constant matrix of rank q , $q \leq p \leq n$. Then, U has a nonsingular $p \times q$ matrix-variate real Gaussian distribution with the parameters $\mathbf{MA} = E[\mathbf{YA}]$, $A'A > O$, and $\Sigma > O$. Analogously, in the complex domain, $\tilde{U} = \tilde{\mathbf{Y}}A$ is a $p \times q$ -variate complex Gaussian distribution with the corresponding parameters $E[\tilde{\mathbf{Y}}A]$, $A^*A > O$, and $\tilde{\Sigma} > O$, A^* denoting the conjugate transpose of A . In the usual format of a $p \times q$ matrix-variate $N_{p,q}(M, A, B)$ real Gaussian density, M is replaced by \mathbf{MA} , A , by $A'A$ and B , by Σ , in the real case, with corresponding changes for the complex case.*

A certain particular case turns out to be of interest. Observe that $\mathbf{MA} = \mu(J'A)$, $J' = (1, \dots, 1)$, and that when $q = 1$, we are considering only one linear combination of Y_1, \dots, Y_n in the form $U_1 = a_1Y_1 + \dots + a_nY_n$, where a_1, \dots, a_n are real scalar constants. Then $J'A = \sum_{j=1}^n a_j$, $A'A = \sum_{j=1}^n a_j^2$, and the $p \times 1$ vector U_1 has a p -variate real nonsingular Gaussian distribution with the parameters $(\sum_{j=1}^n a_j)\mu$ and $(\sum_{j=1}^n a_j^2)\Sigma$. This result was stated in Theorem 3.5.4.

Corollary 4.6.1, 4.6a.1. *Let A as defined in Theorem 4.6.3 be $n \times 1$, in which case A is a column vector whose components are a_1, \dots, a_n , and the resulting single linear function of Y_1, \dots, Y_n is $U_1 = a_1Y_1 + \dots + a_nY_n$. Let the population be p -variate real Gaussian with the parameters μ and $\Sigma > O$. Then U_1 has a p -variate nonsingular real normal distribution with the parameters $(\sum_{j=1}^n a_j)\mu$ and $(\sum_{j=1}^n a_j^2)\Sigma$. Analogously, in the complex Gaussian population case, $\tilde{U}_1 = a_1\tilde{Y}_1 + \dots + a_n\tilde{Y}_n$ is distributed as a complex Gaussian with mean value $(\sum_{j=1}^n a_j)\tilde{\mu}$ and covariance matrix $(\sum_{j=1}^n a_j^*a_j)\tilde{\Sigma}$. Taking $a_1 = \dots = a_n = \frac{1}{n}$, $U_1 = \frac{1}{n}(Y_1 + \dots + Y_n) = \bar{Y}$, the sample average, which has a p -variate real Gaussian density with the parameters μ and $\frac{1}{n}\Sigma$. Correspondingly, in the complex Gaussian case, the sample average \tilde{Y} is a p -variate complex Gaussian vector with the parameters $\tilde{\mu}$ and $\frac{1}{n}\tilde{\Sigma}$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$.*

4.6.3. The general real matrix-variate case

In order to avoid a multiplicity of symbols, we will denote the $p \times q$ real matrix-variate random variable by $X_\alpha = (x_{ij\alpha})$ and the corresponding complex matrix by $\tilde{X}_\alpha = (\tilde{x}_{ij\alpha})$. Consider a simple random sample of size n from the population represented by the real

$p \times q$ matrix $X_\alpha = (x_{ij\alpha})$. Let $X_\alpha = (x_{ij\alpha})$ be the α -th sample value, so that the X_α 's, $\alpha = 1, \dots, n$, are iid as X_1 . Let the $p \times nq$ sample matrix be denoted by the bold-faced $\mathbf{X} = [X_1, X_2, \dots, X_n]$ where each X_j is $p \times q$. Let the sample average be denoted by $\bar{X} = (\bar{x}_{ij})$, $\bar{x}_{ij} = \frac{1}{n} \sum_{\alpha=1}^n x_{ij\alpha}$. Let \mathbf{X}_d be the sample deviation matrix which is the $p \times qn$ matrix

$$\mathbf{X}_d = [X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}], \quad X_\alpha - \bar{X} = (x_{ij\alpha} - \bar{x}_{ij}), \quad (4.6.4)$$

wherein the corresponding sample average is subtracted from each element. For example,

$$\begin{aligned} X_\alpha - \bar{X} &= \begin{bmatrix} x_{11\alpha} - \bar{x}_{11} & x_{12\alpha} - \bar{x}_{12} & \cdots & x_{1q\alpha} - \bar{x}_{1q} \\ x_{21\alpha} - \bar{x}_{21} & x_{22\alpha} - \bar{x}_{22} & \cdots & x_{2q\alpha} - \bar{x}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1\alpha} - \bar{x}_{p1} & x_{p2\alpha} - \bar{x}_{p2} & \cdots & x_{pq\alpha} - \bar{x}_{pq} \end{bmatrix} \\ &= [C_{1\alpha} \quad C_{2\alpha} \quad \cdots \quad C_{q\alpha}] \end{aligned} \quad (i)$$

where $C_{j\alpha}$ is the j -th column in the α -th sample deviation matrix $X_\alpha - \bar{X}$. In this notation, the $p \times qn$ sample deviation matrix can be expressed as follows:

$$\mathbf{X}_d = [C_{11}, C_{21}, \dots, C_{q1}, C_{12}, C_{22}, \dots, C_{q2}, \dots, C_{1n}, C_{2n}, \dots, C_{qn}] \quad (ii)$$

where, for example, $C_{\gamma\alpha}$ denotes the γ -th column in the α -th $p \times q$ matrix, $X_\alpha - \bar{X}$, that is,

$$C_{\gamma\alpha} = \begin{bmatrix} x_{1\gamma\alpha} - \bar{x}_{1\gamma} \\ x_{2\gamma\alpha} - \bar{x}_{2\gamma} \\ \vdots \\ x_{p\gamma\alpha} - \bar{x}_{p\gamma} \end{bmatrix}. \quad (iii)$$

Then, the sample sum of products matrix, denoted by S , is given by

$$\begin{aligned} S = \mathbf{X}_d \mathbf{X}_d' &= C_{11}C_{11}' + C_{21}C_{21}' + \cdots + C_{q1}C_{q1}' \\ &\quad + C_{12}C_{12}' + C_{22}C_{22}' + \cdots + C_{q2}C_{q2}' \\ &\quad \vdots \\ &\quad + C_{1n}C_{1n}' + C_{2n}C_{2n}' + \cdots + C_{qn}C_{qn}'. \end{aligned} \quad (iv)$$

Let us rearrange these matrices by collecting the terms relevant to each column of \mathbf{X} which are

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{p1} \end{bmatrix}, \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{p2} \end{bmatrix}, \dots, \begin{bmatrix} x_{1q} \\ x_{2q} \\ \vdots \\ x_{pq} \end{bmatrix}.$$

Then, the terms relevant to these columns are the following:

$$\begin{aligned}
 S &= \mathbf{X}_d \mathbf{X}_d' = C_{11} C_{11}' + C_{21} C_{21}' + \cdots + C_{q1} C_{q1}' \\
 &\quad + C_{12} C_{12}' + C_{22} C_{22}' + \cdots + C_{q2} C_{q2}' \\
 &\quad \vdots \\
 &\quad + C_{1n} C_{1n}' + C_{2n} C_{2n}' + \cdots + C_{qn} C_{qn}' \\
 &\equiv S_1 + S_2 + \cdots + S_q
 \end{aligned} \tag{v}$$

where S_1 denotes the $p \times p$ sample sum of products matrix in the first column of \mathbf{X} , S_2 , the $p \times p$ sample sum of products matrix corresponding to the second column of \mathbf{X} , and so on, S_q being equal to the $p \times p$ sample sum of products matrix corresponding to the q -th column of \mathbf{X} .

Theorem 4.6.4. Let $X_\alpha = (x_{ij\alpha})$ be a real $p \times q$ matrix of distinct real scalar variables $x_{ij\alpha}$'s. Letting X_α , \bar{X} , \mathbf{X} , \mathbf{X}_d , S , and S_1, \dots, S_q be as previously defined, the sample sum of products matrix in the $p \times nq$ sample matrix \mathbf{X} , denoted by S , is given by

$$S = S_1 + \cdots + S_q. \tag{4.6.5}$$

Example 4.6.1. Consider a 2×2 real matrix-variate $N_{2,2}(O, A, B)$ distribution with the parameters

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}.$$

Let X_α , $\alpha = 1, \dots, 5$, be a simple random sample of size 5 from this real Gaussian population. Suppose that the following observations on X_α , $\alpha = 1, \dots, 5$, were obtained:

$$\begin{aligned}
 X_1 &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \\
 X_4 &= \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \quad X_5 = \begin{bmatrix} -4 & 1 \\ -1 & -2 \end{bmatrix}.
 \end{aligned}$$

Compute the sample matrix, the sample average, the sample deviation matrix and the sample sum of products matrix.

Solution 4.6.1. The sample average is available as

$$\begin{aligned}\bar{X} &= \frac{1}{5}[X_1 + \cdots + X_5] \\ &= \frac{1}{5} \begin{bmatrix} 1 + (-1) + 0 + (-1) + (-4) & 1 + 1 + 1 + 1 + 1 \\ 1 + (-2) + 1 + 1 + (-1) & 2 + 1 + 2 + 2 + (-2) \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix}.\end{aligned}$$

The deviations are then

$$\begin{aligned}X_{1d} &= X_1 - \bar{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}, \quad X_{2d} = \begin{bmatrix} 0 & 0 \\ -2 & 0 \end{bmatrix} \\ X_{3d} &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad X_{4d} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad X_{5d} = \begin{bmatrix} -3 & 0 \\ -1 & -3 \end{bmatrix}.\end{aligned}$$

Thus, the sample matrix, the sample average matrix and the sample deviation matrix, denoted by bold-faced letters, are the following:

$$\mathbf{X} = [X_1, X_2, X_3, X_4, X_5], \quad \bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}] \text{ and } \mathbf{X}_d = [X_{1d}, X_{2d}, X_{3d}, X_{4d}, X_{5d}].$$

The sample sum of products matrix is then

$$S = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]' = [\mathbf{X}_d][\mathbf{X}_d]' = S_1 + S_2$$

where S_1 is obtained from the first columns of each of $X_{\alpha d}$, $\alpha = 1, \dots, 5$, and S_2 is evaluated from the second columns of $X_{\alpha d}$, $\alpha = 1, \dots, 5$. That is,

$$\begin{aligned}S_1 &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} [2 \ 1] + \begin{bmatrix} 0 \\ -2 \end{bmatrix} [0 \ -2] + \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1 \ 1] + \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \ 1] + \begin{bmatrix} -3 \\ -1 \end{bmatrix} [-3 \ -1] \\ &= \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 8 \end{bmatrix}; \\ S_2 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \ 1] + \begin{bmatrix} 0 \\ 0 \end{bmatrix} [0 \ 0] + \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \ 1] + \begin{bmatrix} 0 \\ 1 \end{bmatrix} [0 \ 1] + \begin{bmatrix} 0 \\ -3 \end{bmatrix} [0 \ -3] \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \mathcal{O} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 12 \end{bmatrix}; \\ S &= S_1 + S_2 = \begin{bmatrix} 14 & 6 \\ 6 & 20 \end{bmatrix}.\end{aligned}$$

This S can be directly verified by taking $[\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]' = [\mathbf{X}_d][\mathbf{X}_d]'$ where

$$\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}_d = \begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -3 & 0 \\ 1 & 1 & -2 & 0 & 1 & 1 & 1 & 1 & -1 & -3 \end{bmatrix}, \quad S = \mathbf{X}_d \mathbf{X}_d'.$$

4.6a. The General Complex Matrix-variate Case

The preceding analysis has its counterpart for the complex case. Let $\tilde{X}_\alpha = (\tilde{x}_{ij\alpha})$ be a $p \times q$ matrix in the complex domain with the $\tilde{x}_{ij\alpha}$'s being distinct complex scalar variables. Consider a simple random sample of size n from this population designated by \tilde{X}_1 . Let the α -th sample matrix be \tilde{X}_α , $\alpha = 1, \dots, n$, the \tilde{X}_α 's being iid as \tilde{X}_1 , and the $p \times nq$ sample matrix be denoted by the bold-faced $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_n]$. Let the sample average be denoted by $\bar{\tilde{X}} = (\bar{\tilde{x}}_{ij})$, $\bar{\tilde{x}}_{ij} = \frac{1}{n} \sum_{\alpha=1}^n \tilde{x}_{ij\alpha}$, and $\tilde{\mathbf{X}}_d$ be the sample deviation matrix:

$$\tilde{\mathbf{X}}_d = [\tilde{X}_1 - \bar{\tilde{X}}, \dots, \tilde{X}_n - \bar{\tilde{X}}].$$

Let \tilde{S} be the sample sum of products matrix, namely, $\tilde{S} = \tilde{\mathbf{X}}_d \tilde{\mathbf{X}}_d^*$ where an asterisk denotes the complex conjugate transpose and let \tilde{S}_j be the sample sum of products matrix corresponding to the j -th column of $\tilde{\mathbf{X}}$. Then we have the following result:

Theorem 4.6a.3. Let $\tilde{\mathbf{X}}$, $\bar{\tilde{X}}$, $\tilde{\mathbf{X}}_d$, \tilde{S} and \tilde{S}_j be as previously defined. Then,

$$\tilde{S} = \tilde{S}_1 + \dots + \tilde{S}_q = \tilde{\mathbf{X}}_d \tilde{\mathbf{X}}_d^*. \quad (4.6a.1)$$

Example 4.6a.1. Consider a 2×2 complex matrix-variate $\tilde{N}_{2,2}(O, A, B)$ distribution where

$$A = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & i \\ -1 & 2 \end{bmatrix}.$$

A simple random sample of size 4 from this population is available, that is, $\tilde{X}_\alpha \stackrel{iid}{\sim} \tilde{N}_{2,2}(O, A, B)$, $\alpha = 1, 2, 3, 4$. The following are one set of observations on these sample values:

$$\tilde{X}_1 = \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix}, \tilde{X}_2 = \begin{bmatrix} 3 & -i \\ i & 1 \end{bmatrix}, \tilde{X}_3 = \begin{bmatrix} 1 & 1-i \\ 1+i & 3 \end{bmatrix}, \tilde{X}_4 = \begin{bmatrix} 2 & 3+i \\ 3-i & 7 \end{bmatrix}.$$

Determine the observed sample average, the sample matrix, the sample deviation matrix and the sample sum of products matrix.

Solution 4.6a.1. The sample average is

$$\begin{aligned} \bar{\tilde{X}} &= \frac{1}{4} [\tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 + \tilde{X}_4] \\ &= \frac{1}{4} \left\{ \begin{bmatrix} 2 & i \\ -i & 1 \end{bmatrix} + \begin{bmatrix} 3 & -i \\ i & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1-i \\ 1+i & 3 \end{bmatrix} + \begin{bmatrix} 2 & 3+i \\ 3-i & 7 \end{bmatrix} \right\} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \end{aligned}$$

and the deviations are as follows:

$$\begin{aligned}\tilde{X}_{1d} &= \tilde{X}_1 - \bar{\tilde{X}} = \begin{bmatrix} 0 & -1+i \\ -1-i & -2 \end{bmatrix}, \quad \tilde{X}_{2d} = \begin{bmatrix} 1 & -1-i \\ -1+i & -2 \end{bmatrix}, \\ \tilde{X}_{3d} &= \begin{bmatrix} -1 & -i \\ i & 0 \end{bmatrix}, \quad \tilde{X}_{4d} = \begin{bmatrix} 0 & 2+i \\ 2-i & 4 \end{bmatrix}.\end{aligned}$$

The sample deviation matrix is then $\tilde{\mathbf{X}}_d = [\tilde{X}_{1d}, \tilde{X}_{2d}, \tilde{X}_{3d}, \tilde{X}_{4d}]$. If $V_{\alpha 1}$ denotes the first column of $\tilde{X}_{\alpha d}$, then with our usual notation, $\tilde{S}_1 = \sum_{j=1}^4 V_{\alpha 1} V_{\alpha 1}^*$ and similarly, if $V_{\alpha 2}$ is the second column of $\tilde{X}_{\alpha d}$, then $\tilde{S}_2 = \sum_{\alpha=1}^4 V_{\alpha 2} V_{\alpha 2}^*$, the sample sum of products matrix being $\tilde{S} = \tilde{S}_1 + \tilde{S}_2$. Let us evaluate these quantities:

$$\begin{aligned}\tilde{S}_1 &= \begin{bmatrix} 0 \\ -1-i \end{bmatrix} [0 \ -1+i] + \begin{bmatrix} 1 \\ -1+i \end{bmatrix} [1 \ -1-i] + \begin{bmatrix} -1 \\ i \end{bmatrix} [-1 \ -i] + \begin{bmatrix} 0 \\ 2-i \end{bmatrix} [0 \ 2+i] \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1-i \\ -1+i & 2 \end{bmatrix} + \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 10 \end{bmatrix},\end{aligned}$$

$$\begin{aligned}\tilde{S}_2 &= \begin{bmatrix} -1+i \\ -2 \end{bmatrix} [-1-i \ -2] + \begin{bmatrix} -1-i \\ -2 \end{bmatrix} [-1+i \ -2] + \begin{bmatrix} -i \\ 0 \end{bmatrix} [i \ 0] + \begin{bmatrix} 2+i \\ 4 \end{bmatrix} [2-i \ 4] \\ &= \begin{bmatrix} 2 & 2-2i \\ 2+2i & 4 \end{bmatrix} + \begin{bmatrix} 2 & 2+2i \\ 2-2i & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 5 & 8+4i \\ 8-4i & 16 \end{bmatrix} \\ &= \begin{bmatrix} 10 & 12+4i \\ 12-4i & 24 \end{bmatrix},\end{aligned}$$

and then,

$$\tilde{S} = \tilde{S}_1 + \tilde{S}_2 = \begin{bmatrix} 2 & -1 \\ -1 & 10 \end{bmatrix} + \begin{bmatrix} 10 & 12+4i \\ 12-4i & 24 \end{bmatrix} = \begin{bmatrix} 12 & 11+4i \\ 11-4i & 34 \end{bmatrix}.$$

This can also be verified directly as $\tilde{S} = [\tilde{\mathbf{X}}_d][\tilde{\mathbf{X}}_d]^*$ where the deviation matrix is

$$\tilde{\mathbf{X}}_d = \begin{bmatrix} 0 & -1+i & 1 & -1-i & -1 & -i & 0 & 2+i \\ -i-1 & -2 & -1+i & -2 & i & 0 & 2-i & 4 \end{bmatrix}.$$

As expected,

$$[\tilde{\mathbf{X}}_d][\tilde{\mathbf{X}}_d]^* = \begin{bmatrix} 12 & 11+4i \\ 11-4i & 34 \end{bmatrix}.$$

This completes the calculations.

Exercises 4.6

4.6.1. Let A be a 2×2 matrix whose first row is $(1, 1)$ and B be 3×3 matrix whose first row is $(1, -1, 1)$. Select your own real numbers to complete the matrices A and B so that $A > O$ and $B > O$. Then consider a 2×3 matrix X having a real matrix-variate Gaussian density with the location parameter $M = O$ and the foregoing parameter matrices A and B . Let the first row of X be X_1 and its second row be X_2 . Determine the marginal densities of X_1 and X_2 , the conditional density of X_1 given X_2 , the conditional density of X_2 given X_1 , the conditional expectation of X_1 given $X_2 = (1, 0, 1)$ and the conditional expectation of X_2 given $X_1 = (1, 2, 3)$.

4.6.2. Consider the matrix X utilized in Exercise 4.6.1. Let its first two columns be Y_1 and its last one be Y_2 . Then, obtain the marginal densities of Y_1 and Y_2 , and the conditional densities of Y_1 given Y_2 and Y_2 given Y_1 , and evaluate the conditional expectation of Y_1 given $Y_2' = (1, -1)$ as well as the conditional expectation of Y_2 given $Y_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$.

4.6.3. Let $A > O$ and $B > O$ be 2×2 and 3×3 matrices whose first rows are $(1, 1 - i)$ and $(2, i, 1 + i)$, respectively. Select your own complex numbers to complete the matrices $A = A^* > O$ and $B = B^* > O$. Now, consider a 2×3 matrix \tilde{X} having a complex matrix-variate Gaussian density with the aforementioned matrices A and B as parameter matrices. Assume that the location parameter is a null matrix. Letting the row partitioning of \tilde{X} , denoted by \tilde{X}_1, \tilde{X}_2 , be as specified in Exercise 4.6.1, answer all the questions posed in that exercise.

4.6.4. Let A, B and \tilde{X} be as given in Exercise 4.6.3. Consider the column partitioning specified in Exercise 4.6.2. Then answer all the questions posed in Exercise 4.6.2.

4.6.5. Repeat Exercise 4.6.4 with the non-null location parameter

$$\tilde{M} = \begin{bmatrix} 2 & 1 - i & i \\ 1 + i & 2 + i & -3i \end{bmatrix}.$$

4.7. The Singular Matrix-variate Gaussian Distribution

Consider the moment generating function specified in (4.3.3) for the real case, namely,

$$M_X(T) = M_f(T) = e^{\text{tr}(TM') + \frac{1}{2}\text{tr}(\Sigma_1 T \Sigma_2 T')} \quad (4.7.1)$$

where $\Sigma_1 = A^{-1} > O$ and $\Sigma_2 = B^{-1} > O$. In the complex case, the moment generating function is of the form

$$\tilde{M}_{\tilde{X}}(\tilde{T}) = e^{\Re[\text{tr}(\tilde{T}\tilde{M}^*)] + \frac{1}{4}\text{tr}(\Sigma_1 \tilde{T} \Sigma_2 \tilde{T}^*)}. \quad (4.7a.1)$$

The properties of the singular matrix-variate Gaussian distribution can be studied by making use of (4.7.1) and (4.7a.1). Suppose that we restrict Σ_1 and Σ_2 to be positive semi-definite matrices, that is, $\Sigma_1 \geq O$ and $\Sigma_2 \geq O$. In this case, one can also study many properties of the distributions represented by the mgf's given in (4.7.1) and (4.7a.1); however, the corresponding densities will not exist unless the matrices Σ_1 and Σ_2 are both strictly positive definite. The $p \times q$ real or complex matrix-variate density does not exist if at least one of A or B is singular. When either or both Σ_1 and Σ_2 are only positive semi-definite, the distributions corresponding to the mgf's specified by (4.7.1) and (4.7a.1) are respectively referred to as *real matrix-variate singular Gaussian* and *complex matrix-variate singular Gaussian*.

For instance, let

$$\Sigma_1 = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \text{ and } \Sigma_2 = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

in the mgf of a 2×3 real matrix-variate Gaussian distribution. Note that $\Sigma_1 = \Sigma_1'$ and $\Sigma_2 = \Sigma_2'$. Since the leading minors of Σ_1 are $|(4)| = 4 > 0$ and $|\Sigma_1| = 0$ and those of Σ_2 are $|(3)| = 3 > 0$, $\begin{vmatrix} 3 & -1 \\ -1 & 2 \end{vmatrix} = 5 > 0$ and $|\Sigma_2| = 2 > 0$, Σ_1 is positive semi-definite and Σ_2 is positive definite. Accordingly, the resulting Gaussian distribution does not possess a density. Fortunately, its distributional properties can nevertheless be investigated via its associated moment generating function.

References

- A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.
- A.M. Mathai and Nicy Sebastian (2022): On distributions of covariance structures, *Communications in Statistics, Theory and Methods*, <http://doi.org/10.1080/03610926.2022.2045022>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 5

Matrix-Variate Gamma and Beta Distributions



5.1. Introduction

The notations introduced in the preceding chapters will still be followed in this one. Lower-case letters such as x, y will be utilized to represent real scalar variables, whether mathematical or random. Capital letters such as X, Y will be used to denote vector/matrix random or mathematical variables. A tilde placed on top of a letter will indicate that the variables are in the complex domain. However, the tilde will be omitted in the case of constant matrices such as A, B . The determinant of a square matrix A will be denoted as $|A|$ or $\det(A)$ and, in the complex domain, the absolute value or modulus of the determinant of B will be denoted as $|\det(B)|$. Square matrices appearing in this chapter will be assumed to be of dimension $p \times p$ unless otherwise specified.

We will first define the real matrix-variate gamma function, gamma integral and gamma density, wherefrom their counterparts in the complex domain will be developed. A particular case of the real matrix-variate gamma density known as the Wishart density is widely utilized in multivariate statistical analysis. Actually, the formulation of this distribution in 1928 constituted a significant advance in the early days of the discipline. A real matrix-variate gamma function, denoted by $\Gamma_p(\alpha)$, will be defined in terms of a matrix-variate integral over a real positive definite matrix $X > O$. This integral representation of $\Gamma_p(\alpha)$ will be explicitly evaluated with the help of the transformation of a real positive definite matrix in terms of a lower triangular matrix having positive diagonal elements in the form $X = TT'$ where $T = (t_{ij})$ is a lower triangular matrix with positive diagonal elements, that is, $t_{ij} = 0, i < j$ and $t_{jj} > 0, j = 1, \dots, p$. When the diagonal elements are positive, it can be shown that the transformation $X = TT'$ is unique. Its associated Jacobian is provided in Theorem 1.6.7. This result is now restated for ready reference: For a $p \times p$ real positive definite matrix $X = (x_{ij}) > O$,

$$X = TT' \Rightarrow dX = 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT \quad (5.1.1)$$

where $T = (t_{ij})$, $t_{ij} = 0$, $i < j$ and $t_{jj} > 0$, $j = 1, \dots, p$. Consider the following integral representation of $\Gamma_p(\alpha)$ where the integral is over a real positive definite matrix X and the integrand is a real-valued scalar function of X :

$$\Gamma_p(\alpha) = \int_{X>O} |X|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(X)} dX. \quad (5.1.2)$$

Under the transformation in (5.1.1),

$$\begin{aligned} |X|^{\alpha - \frac{p+1}{2}} dX &= \left\{ \prod_{j=1}^p (t_{jj}^2)^{\alpha - \frac{p+1}{2}} \right\} 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT \\ &= 2^p \left\{ \prod_{j=1}^p (t_{jj}^2)^{\alpha - \frac{j}{2}} \right\} dT. \end{aligned}$$

Observe that $\text{tr}(X) = \text{tr}(TT')$ = the sum of the squares of all the elements in T , which is $\sum_{j=1}^p t_{jj}^2 + \sum_{i>j} t_{ij}^2$. By letting $t_{jj}^2 = y_j \Rightarrow dt_{jj} = \frac{1}{2} y_j^{\frac{1}{2}-1} dy_j$, noting that $t_{jj} > 0$, the integral over t_{jj} gives

$$2 \int_0^\infty (t_{jj}^2)^{\alpha - \frac{j}{2}} e^{-t_{jj}^2} dt_{jj} = \Gamma\left(\alpha - \frac{j-1}{2}\right), \Re\left(\alpha - \frac{j-1}{2}\right) > 0, \quad j = 1, \dots, p,$$

the final condition being $\Re(\alpha) > \frac{p-1}{2}$. Thus, we have the gamma product $\Gamma(\alpha)\Gamma(\alpha - \frac{1}{2}) \cdots \Gamma(\alpha - \frac{p-1}{2})$. Now for $i > j$, the integral over t_{ij} gives

$$\prod_{i>j} \int_{-\infty}^\infty e^{-t_{ij}^2} dt_{ij} = \prod_{i>j} \sqrt{\pi} = \pi^{\frac{p(p-1)}{4}}.$$

Therefore

$$\begin{aligned} \Gamma_p(\alpha) &= \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha)\Gamma\left(\alpha - \frac{1}{2}\right) \cdots \Gamma\left(\alpha - \frac{p-1}{2}\right), \Re(\alpha) > \frac{p-1}{2}, \\ &= \int_{X>O} |X|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(X)} dX, \Re(\alpha) > \frac{p-1}{2}. \end{aligned} \quad (5.1.3)$$

For example,

$$\Gamma_2(\alpha) = \pi^{\frac{(2)(1)}{4}} \Gamma(\alpha)\Gamma\left(\alpha - \frac{1}{2}\right) = \pi^{\frac{1}{2}} \Gamma(\alpha)\Gamma\left(\alpha - \frac{1}{2}\right), \Re(\alpha) > \frac{1}{2}.$$

This $\Gamma_p(\alpha)$ is known by different names in the literature. The first author calls it *the real matrix-variate gamma function* because of its association with a real matrix-variate gamma integral.

5.1a. The Complex Matrix-variate Gamma

In the complex case, consider a $p \times p$ Hermitian positive definite matrix $\tilde{X} = \tilde{X}^* > O$, where \tilde{X}^* denotes the conjugate transpose of \tilde{X} . Let $\tilde{T} = (\tilde{t}_{ij})$ be a lower triangular matrix with the diagonal elements being real and positive. In this case, it can be shown that the transformation $\tilde{X} = \tilde{T}\tilde{T}^*$ is one-to-one. Then, as stated in Theorem 1.6a.7, the Jacobian is

$$d\tilde{X} = 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T}. \tag{5.1a.1}$$

With the help of (5.1a.1), we can evaluate the following integral over $p \times p$ Hermitian positive definite matrices where the integrand is a real-valued scalar function of \tilde{X} . We will denote the integral by $\tilde{\Gamma}_p(\alpha)$, that is,

$$\tilde{\Gamma}_p(\alpha) = \int_{\tilde{X} > O} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})} d\tilde{X}. \tag{5.1a.2}$$

Let us evaluate the integral in (5.1a.2) by making use of (5.1a.1). Parallel to the real case, we have

$$\begin{aligned} |\det(\tilde{X})|^{\alpha-p} d\tilde{X} &= \left\{ \prod_{i=1}^p (t_{jj}^2)^{\alpha-p} \right\} 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T} \\ &= \left\{ \prod_{j=1}^p 2(t_{jj}^2)^{\alpha-j+\frac{1}{2}} \right\} d\tilde{T}. \end{aligned}$$

As well,

$$e^{-\text{tr}(\tilde{X})} = e^{-\sum_{j=1}^p t_{jj}^2 - \sum_{i>j} |\tilde{t}_{ij}|^2}.$$

Since t_{jj} is real and positive, the integral over t_{jj} gives the following:

$$2 \int_0^\infty (t_{jj}^2)^{\alpha-j+\frac{1}{2}} e^{-t_{jj}^2} dt_{jj} = \Gamma(\alpha - (j - 1)), \Re(\alpha - (j - 1)) > 0, j = 1, \dots, p,$$

the final condition being $\Re(\alpha) > p - 1$. Note that the absolute value of \tilde{t}_{ij} , namely, $|\tilde{t}_{ij}|$ is such that $|\tilde{t}_{ij}|^2 = t_{ij1}^2 + t_{ij2}^2$ where $\tilde{t}_{ij} = t_{ij1} + it_{ij2}$ with t_{ij1}, t_{ij2} real and $i = \sqrt{-1}$.

Thus,

$$\prod_{i>j} \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-(t_{ij1}^2+t_{ij2}^2)} dt_{ij1} \wedge dt_{ij2} = \prod_{i>j} \pi = \pi^{\frac{p(p-1)}{2}}.$$

Then

$$\begin{aligned}\tilde{\Gamma}_p(\alpha) &= \int_{\tilde{X} > O} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})} d\tilde{X}, \Re(\alpha) > \frac{p-1}{2}, \\ &= \pi^{\frac{p(p-1)}{2}} \Gamma(\alpha) \Gamma(\alpha-1) \cdots \Gamma(\alpha-(p-1)), \Re(\alpha) > p-1.\end{aligned}\quad (5.1a.3)$$

We will refer to $\tilde{\Gamma}_p(\alpha)$ as the complex matrix-variate gamma because of its association with a complex matrix-variate gamma integral. As an example, consider

$$\tilde{\Gamma}_2(\alpha) = \pi^{\frac{(2)(1)}{2}} \Gamma(\alpha) \Gamma(\alpha-1) = \pi \Gamma(\alpha) \Gamma(\alpha-1), \Re(\alpha) > 1.$$

5.2. The Real Matrix-variate Gamma Density

In view of (5.1.3), we can define a real matrix-variate gamma density with shape parameter α as follows, where X is $p \times p$ real positive definite matrix:

$$f_1(X) = \begin{cases} \frac{1}{\tilde{\Gamma}_p(\alpha)} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(X)}, & X > O, \Re(\alpha) > \frac{p-1}{2} \\ 0, & \text{elsewhere.} \end{cases} \quad (5.2.1)$$

Example 5.2.1. Let

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}, \tilde{X} = \begin{bmatrix} x_1 & x_2 + iy_2 \\ x_2 - iy_2 & x_3 \end{bmatrix}, X = X' > O, \tilde{X} = \tilde{X}^* > O,$$

where x_{11} , x_{12} , x_{22} , x_1 , x_2 , y_2 , x_3 are all real scalar variables, $i = \sqrt{-1}$, $x_{22} > 0$, $x_{11}x_{22} - x_{12}^2 > 0$. While these are the conditions for the positive definiteness of the real matrix X , $x_1 > 0$, $x_3 > 0$, $x_1x_3 - (x_2^2 + y_2^2) > 0$ are the conditions for the Hermitian positive definiteness of \tilde{X} . Let us evaluate the following integrals, subject to the previously specified conditions on the elements of the matrix:

$$\begin{aligned}(1) : \delta_1 &= \int_{X > O} e^{-(x_{11}+x_{22})} dx_{11} \wedge dx_{12} \wedge dx_{22} \\ (2) : \delta_2 &= \int_{\tilde{X} > O} e^{-(x_1+x_3)} dx_1 \wedge d(x_2 + iy_2) \wedge dx_3 \\ (3) : \delta_3 &= \int_{X > O} |X| e^{-(x_{11}+x_{22})} dx_{11} \wedge dx_{12} \wedge dx_{22} \\ (4) : \delta_4 &= \int_{\tilde{X} > O} |\det(\tilde{X})|^2 e^{-(x_1+x_3)} dx_1 \wedge d(x_2 + iy_2) \wedge dx_3.\end{aligned}$$

Solution 5.2.1. (1): Observe that δ_1 can be evaluated by treating the integral as a real matrix-variate integral, namely,

$$\delta_1 = \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(X)} dX \text{ with } p = 2, \frac{p+1}{2} = \frac{3}{2}, \alpha = \frac{3}{2},$$

and hence the integral is

$$\Gamma_2(3/2) = \pi^{\frac{2(1)}{4}} \Gamma(3/2)\Gamma(1) = \pi^{1/2}(1/2)\Gamma(1/2) = \frac{\pi}{2}.$$

This result can also be obtained by direct integration as a multiple integral. In this case, the integration has to be done under the conditions $x_{11} > 0, x_{22} > 0, x_{11}x_{22} - x_{12}^2 > 0$, that is, $x_{12}^2 < x_{11}x_{22}$ or $-\sqrt{x_{11}x_{22}} < x_{12} < \sqrt{x_{11}x_{22}}$. The integral over x_{12} yields $\int_{-\sqrt{x_{11}x_{22}}}^{\sqrt{x_{11}x_{22}}} dx_{12} = 2\sqrt{x_{11}x_{22}}$, that over x_{11} then gives

$$2 \int_{x_{11}=0}^{\infty} \sqrt{x_{11}} e^{-x_{11}} dx_{11} = 2 \int_0^{\infty} x_{11}^{\frac{3}{2}-1} e^{-x_{11}} dx_{11} = 2\Gamma(3/2) = \pi^{\frac{1}{2}},$$

and on integrating with respect to x_{22} , we have

$$\int_0^{\infty} \sqrt{x_{22}} e^{-x_{22}} dx_{22} = \frac{1}{2}\pi^{\frac{1}{2}},$$

so that $\delta_1 = \frac{1}{2}\sqrt{\pi}\sqrt{\pi} = \frac{\pi}{2}$.

(2): On observing that δ_2 can be viewed as a complex matrix-variate integral, it is seen that

$$\delta_2 = \int_{\tilde{X}>O} |\det(\tilde{X})|^{2-2} e^{-\text{tr}(\tilde{X})} d\tilde{X} = \tilde{\Gamma}_2(2) = \pi^{\frac{2(1)}{2}} \Gamma(2)\Gamma(1) = \pi.$$

This answer can also be obtained by evaluating the multiple integral. Since $\tilde{X} > O$, we have $x_1 > 0, x_3 > 0, x_1x_3 - (x_2^2 + y_2^2) > 0$, that is, $x_1 > \frac{(x_2^2+y_2^2)}{x_3}$. Integrating first with respect to x_1 and letting $y = x_1 - \frac{(x_2^2+y_2^2)}{x_3}$, we have

$$\int_{x_1>\frac{(x_2^2+y_2^2)}{x_3}} e^{-x_1} dx_1 = \int_{y=0}^{\infty} e^{-y-\frac{(x_2^2+y_2^2)}{x_3}} dy = e^{-\frac{(x_2^2+y_2^2)}{x_3}}.$$

Now, the integrals over x_2 and y_2 give

$$\int_{-\infty}^{\infty} e^{-\frac{x_2^2}{x_3}} dx_2 = \sqrt{x_3} \int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{x_3}\sqrt{\pi} \text{ and } \int_{-\infty}^{\infty} e^{-\frac{y_2^2}{x_3}} dy_2 = \sqrt{x_3}\sqrt{\pi},$$

that with respect to x_3 then yielding

$$\int_0^{\infty} x_3 e^{-x_3} dx_3 = \Gamma(2) = 1,$$

so that $\delta_2 = (1)\sqrt{\pi}\sqrt{\pi} = \pi$.

(3): Observe that δ_3 can be evaluated as a real matrix-variate integral. Then

$$\begin{aligned} \delta_3 &= \int_{X>O} |X| e^{-\text{tr}(X)} dX = \int_{X>O} |X|^{\frac{5}{2}-\frac{3}{2}} e^{-\text{tr}(X)} dX, \text{ with } \frac{p+1}{2} = \frac{3}{2} \text{ as } p=2 \\ &= \Gamma_2(5/2) = \pi^{\frac{2(1)}{4}} \Gamma(5/2) \Gamma(4/2) = \pi^{\frac{1}{2}} (3/2) (1/2) \pi^{1/2} (1) \\ &= \frac{3}{4} \pi. \end{aligned}$$

Let us proceed by direct integration:

$$\begin{aligned} \delta_3 &= \int_{X>O} [x_{11}x_{22} - x_{12}^2] e^{-(x_{11}+x_{22})} dx_{11} \wedge dx_{12} \wedge dx_{22} \\ &= \int_{X>O} x_{22} \left[x_{11} - \frac{x_{12}^2}{x_{22}} \right] e^{-(x_{11}+x_{22})} dx_{11} \wedge dx_{12} \wedge dx_{22}; \end{aligned}$$

letting $y = x_{11} - \frac{x_{12}^2}{x_{22}}$, the integral over x_{11} yields

$$\int_{x_{11} > \frac{x_{12}^2}{x_{22}}} \left[x_{11} - \frac{x_{12}^2}{x_{22}} \right] e^{-x_{11}} dx_{11} = \int_{y=0}^{\infty} y e^{-y - \frac{x_{12}^2}{x_{22}}} dy = e^{-\frac{x_{12}^2}{x_{22}}}.$$

Now, the integral over x_{12} gives $\sqrt{x_{22}}\sqrt{\pi}$ and finally, that over x_{22} yields

$$\int_{x_{22}>0} x_{22}^{\frac{3}{2}} e^{-x_{22}} dx_{22} = \Gamma(5/2) = (3/2)(1/2)\sqrt{\pi} = \frac{3}{4}\pi.$$

(4): Noting that we can treat δ_4 as a complex matrix-variate integral, we have

$$\begin{aligned} \delta_4 &= \int_{\tilde{X}>O} |\det(\tilde{X})|^2 e^{-\text{tr}(\tilde{X})} d\tilde{X} = \int_{\tilde{X}>O} |\det(\tilde{X})|^{4-2} e^{-\text{tr}(\tilde{X})} d\tilde{X} = \tilde{\Gamma}_2(4), \alpha=4, p=2, \\ &= \pi^{\frac{2(1)}{2}} \Gamma(4) \Gamma(3) = \pi(3!)(2!) = 12\pi. \end{aligned}$$

Direct evaluation will be challenging in this case as the integrand involves $|\det(\tilde{X})|^2$.

If a scale parameter matrix $B > O$ is to be introduced in (5.2.1), then consider $\text{tr}(BX) = \text{tr}(B^{\frac{1}{2}}XB^{\frac{1}{2}})$ where $B^{\frac{1}{2}}$ is the positive definite square root of the real positive definite constant matrix B . On applying the transformation $Y = B^{\frac{1}{2}}XB^{\frac{1}{2}} \Rightarrow dX = |B|^{-\frac{(p+1)}{2}}dY$, as stated in Theorem 1.6.5, we have

$$\begin{aligned} \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(BX)} dX &= \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(B^{\frac{1}{2}}XB^{\frac{1}{2}})} dX \\ &= |B|^{-\alpha} \int_{Y>O} |Y|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(Y)} dY \\ &= |B|^{-\alpha} \Gamma_p(\alpha). \end{aligned} \tag{5.2.2}$$

This equality brings about two results. First, the following identity which will turn out to be very handy in many of the computations:

$$|B|^{-\alpha} \equiv \frac{1}{\Gamma_p(\alpha)} \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(BX)} dX, \quad B > O, \Re(\alpha) > \frac{p-1}{2}. \tag{5.2.3}$$

As well, the following two-parameter real matrix-variate gamma density with shape parameter α and scale parameter matrix $B > O$ can be constructed from (5.2.2):

$$f(X) = \begin{cases} \frac{|B|^\alpha}{\Gamma_p(\alpha)} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(BX)}, & X > O, B > O, \Re(\alpha) > \frac{p-1}{2} \\ 0, & \text{elsewhere.} \end{cases} \tag{5.2.4}$$

5.2.1. The mgf of the real matrix-variate gamma distribution

Let us determine the mgf associated with the density given in (5.2.4), that is, the two-parameter real matrix-variate gamma density. Observing that $X = X'$, let T be a symmetric $p \times p$ real positive definite parameter matrix. Then, noting that

$$\text{tr}(TX) = \sum_{j=1}^p t_{jj}x_{jj} + 2 \sum_{i>j} t_{ij}x_{ij}, \tag{i}$$

it is seen that the non-diagonal elements in X multiplied by the corresponding parameters will have twice the weight of the diagonal elements multiplied by the corresponding parameters. For instance, consider the 2×2 case:

$$\begin{aligned} \text{tr} \left\{ \begin{bmatrix} t_{11} & t_{12} \\ t_{12} & t_{22} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix} \right\} &= \text{tr} \begin{bmatrix} t_{11}x_{11} + t_{12}x_{12} & \alpha_1 \\ \alpha_2 & t_{12}x_{12} + t_{22}x_{22} \end{bmatrix} \\ &= t_{11}x_{11} + 2t_{12}x_{12} + t_{22}x_{22} \end{aligned} \tag{ii}$$

where α_1 and α_2 represent elements that are not involved in the evaluation of the trace. Note that due to the symmetry of T and X , $t_{21} = t_{12}$ and $x_{21} = x_{12}$, so that the cross product term $t_{12}x_{12}$ in (ii) appears twice whereas each of the terms $t_{11}x_{11}$ and $t_{22}x_{22}$ appear only once.

However, in order to be consistent with the mgf in a real multivariate case, each variable need only be multiplied once by the corresponding parameter, the mgf being then obtained by taking the expected value of the resulting exponential sum. Accordingly, the parameter matrix has to be modified as follows: let $*T = (*t_{ij})$ where $*t_{jj} = t_{jj}$, $*t_{ij} = \frac{1}{2}t_{ij}$, $i \neq j$, and $t_{ij} = t_{ji}$ for all i and j or, in other words, the non-diagonal elements of the symmetric matrix T are weighted by $\frac{1}{2}$, such a matrix being denoted as $*T$. Then,

$$\text{tr}(*TX) = \sum_{i,j} t_{ij}x_{ij},$$

and the mgf in the real matrix-variate two-parameter gamma density, denoted by $M_X(*T)$, is the following:

$$\begin{aligned} M_X(*T) &= E[e^{\text{tr}(*TX)}] \\ &= \frac{|B|^\alpha}{\Gamma_p(\alpha)} \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{\text{tr}(*TX-BX)} dX. \end{aligned}$$

Now, since

$$\text{tr}(BX - *TX) = \text{tr}((B - *T)^{\frac{1}{2}} X (B - *T)^{\frac{1}{2}})$$

for $(B - *T) > O$, that is, $(B - *T)^{\frac{1}{2}} > O$, which means that $Y = (B - *T)^{\frac{1}{2}} X (B - *T)^{\frac{1}{2}} \Rightarrow dX = |B - *T|^{-\frac{p+1}{2}} dY$, we have

$$\begin{aligned} M_X(*T) &= \frac{|B|^\alpha}{\Gamma_p(\alpha)} \int_{X>O} |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}((B-*T)X)} dX \\ &= \frac{|B|^\alpha}{\Gamma_p(\alpha)} |B - *T|^{-\alpha} \int_{Y>O} |Y|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(Y)} dY \\ &= |B|^\alpha |B - *T|^{-\alpha} \\ &= |I - B^{-1} *T|^{-\alpha} \text{ for } I - B^{-1} *T > O. \end{aligned} \tag{5.2.5}$$

When $*T$ is replaced by $-*T$, (5.2.5) gives the Laplace transform of the two-parameter gamma density in the real matrix-variate case as specified by (5.2.4), which is denoted by $L_f(*T)$, that is,

$$L_f(*T) = M_X(-*T) = |I + B^{-1} *T|^{-\alpha} \text{ for } I + B^{-1} *T > O. \tag{5.2.6}$$

For example, if

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix}, B = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \text{ and } {}_{{*}}T = \begin{bmatrix} {}_{{*}}t_{11} & {}_{{*}}t_{12} \\ {}_{{*}}t_{12} & {}_{{*}}t_{22} \end{bmatrix},$$

then $|B| = 5$ and

$$\begin{aligned} M_X({}_{{*}}T) &= |B|^\alpha |B - {}_{{*}}T|^{-\alpha} = 5^\alpha \begin{vmatrix} 2 - {}_{{*}}t_{11} & -1 - {}_{{*}}t_{12} \\ -1 - {}_{{*}}t_{12} & 3 - {}_{{*}}t_{22} \end{vmatrix}^{-\alpha} \\ &= 5^\alpha \{(2 - {}_{{*}}t_{11})(3 - {}_{{*}}t_{22}) - (1 + {}_{{*}}t_{12})^2\}^{-\alpha}. \end{aligned}$$

If ${}_{{*}}T$ is partitioned into sub-matrices and X is partitioned accordingly as

$${}_{{*}}T = \begin{bmatrix} {}_{{*}}T_{11} & {}_{{*}}T_{12} \\ {}_{{*}}T_{21} & {}_{{*}}T_{22} \end{bmatrix} \text{ and } X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \tag{iii}$$

where ${}_{{*}}T_{11}$ and X_{11} are $r \times r$, $r \leq p$, then what can be said about the densities of the diagonal blocks X_{11} and X_{22} ? The mgf of X_{11} is available from the definition by letting ${}_{{*}}T_{12} = O$, ${}_{{*}}T_{21} = O$ and ${}_{{*}}T_{22} = O$, as then $E[e^{\text{tr}({}_{{*}}TX)}] = E[e^{\text{tr}({}_{{*}}T_{11}X_{11})}]$. However, $B^{-1}{}_{{*}}T$ is not positive definite since $B^{-1}{}_{{*}}T$ is not symmetric, and thereby $I - B^{-1}{}_{{*}}T$ cannot be positive definite when ${}_{{*}}T_{12} = O$, ${}_{{*}}T_{21} = O$, ${}_{{*}}T_{22} = O$. Consequently, the mgf of X_{11} cannot be determined from (5.2.6). As an alternative, we could rewrite (5.2.6) in the symmetric format and then try to evaluate the density of X_{11} . As it turns out, the densities of X_{11} and X_{22} can be readily obtained from the mgf in two situations: either when $B = I$ or B is a block diagonal matrix, that is,

$$B = \begin{bmatrix} B_{11} & O \\ O & B_{22} \end{bmatrix} \Rightarrow B^{-1} = \begin{bmatrix} B_{11}^{-1} & O \\ O & B_{22}^{-1} \end{bmatrix}. \tag{iv}$$

Hence we have the following results:

Theorem 5.2.1. *Let the $p \times p$ matrices $X > O$ and ${}_{{*}}T > O$ be partitioned as in (iii). Let X have a $p \times p$ real matrix-variate gamma density with shape parameter α and scale parameter matrix I_p . Then X_{11} has an $r \times r$ real matrix-variate gamma density and X_{22} has a $(p - r) \times (p - r)$ real matrix-variate gamma density with shape parameter α and scale parameters I_r and I_{p-r} , respectively.*

Theorem 5.2.2. *Let X be partitioned as in (iii). Let the $p \times p$ real positive definite parameter matrix $B > O$ be partitioned as in (iv). Then X_{11} has an $r \times r$ real matrix-variate gamma density with the parameters $(\alpha$ and $B_{11} > O)$ and X_{22} has a $(p - r) \times (p - r)$ real matrix-variate gamma density with the parameters $(\alpha$ and $B_{22} > O)$.*

Theorem 5.2.3. *Let X be partitioned as in (iii). Then X_{11} and X_{22} are statistically independently distributed under the restrictions specified in Theorems 5.2.1 and 5.2.2.*

In the general case of B , write the mgf as $M_X(*T) = |B|^\alpha |B - *T|^{-\alpha}$, which corresponds to a symmetric format. Then, when $*T = \begin{bmatrix} *T_{11} & O \\ O & O \end{bmatrix}$,

$$\begin{aligned} M_X(*T) &= |B|^\alpha \begin{vmatrix} B_{11} - *T_{11} & B_{12} \\ B_{21} & B_{22} \end{vmatrix}^{-\alpha} \\ &= |B|^\alpha |B_{22}|^{-\alpha} |B_{11} - *T_{11} - B_{12}B_{22}^{-1}B_{21}|^{-\alpha} \\ &= |B_{22}|^\alpha |B_{11} - B_{12}B_{22}^{-1}B_{21}|^\alpha |B_{22}|^{-\alpha} |(B_{11} - B_{12}B_{22}^{-1}B_{21}) - *T_{11}|^{-\alpha} \\ &= |B_{11} - B_{12}B_{22}^{-1}B_{21}|^\alpha |(B_{11} - B_{12}B_{22}^{-1}B_{21}) - *T_{11}|^{-\alpha}, \end{aligned}$$

which is obtained by making use of the representations of the determinant of a partitioned matrix, which are available from Sect. 1.3. Now, on comparing the last line with the first one, it is seen that X_{11} has a real matrix-variate gamma distribution with shape parameter α and scale parameter matrix $B_{11} - B_{12}B_{22}^{-1}B_{21}$. Hence, the following result:

Theorem 5.2.4. *If the $p \times p$ real positive definite matrix has a real matrix-variate gamma density with the shape parameter α and scale parameter matrix B and if X and B are partitioned as in (iii), then X_{11} has a real matrix-variate gamma density with shape parameter α and scale parameter matrix $B_{11} - B_{12}B_{22}^{-1}B_{21}$, and the sub-matrix X_{22} has a real matrix-variate gamma density with shape parameter α and scale parameter matrix $B_{22} - B_{21}B_{11}^{-1}B_{12}$.*

5.2a. The Matrix-variate Gamma Function and Density, Complex Case

Let $\tilde{X} = \tilde{X}^* > O$ be a $p \times p$ Hermitian positive definite matrix. When \tilde{X} is Hermitian, all its diagonal elements are real and hence $\text{tr}(\tilde{X})$ is real. Let $\det(\tilde{X})$ denote the determinant and $|\det(\tilde{X})|$ denote the absolute value of the determinant of \tilde{X} . As a result, $|\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})}$ is a real-valued scalar function of \tilde{X} . Let us consider the following integral, denoted by $\tilde{\Gamma}_p(\alpha)$:

$$\tilde{\Gamma}_p(\alpha) = \int_{\tilde{X} > O} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})} d\tilde{X}, \quad (5.2a.1)$$

which was evaluated in Sect. 5.1a. In fact, (5.1a.3) provides two representations of the complex matrix-variate gamma function $\tilde{\Gamma}_p(\alpha)$. With the help of (5.1a.3), we can define the complex $p \times p$ matrix-variate gamma density as follows:

$$\tilde{f}_1(\tilde{X}) = \begin{cases} \frac{1}{\tilde{\Gamma}_p(\alpha)} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})}, & \tilde{X} > O, \Re(\alpha) > p - 1 \\ 0, & \text{elsewhere.} \end{cases} \quad (5.2a.2)$$

For example, let us examine the 2×2 complex matrix-variate case. Let \tilde{X} be a matrix in the complex domain, $\tilde{\bar{X}}$ denoting its complex conjugate and \tilde{X}^* , its conjugate transpose. When $\tilde{X} = \tilde{X}^*$, the matrix is Hermitian and its diagonal elements are real. In the 2×2 Hermitian case, let

$$\tilde{X} = \begin{bmatrix} x_1 & x_2 + iy_2 \\ x_2 - iy_2 & x_3 \end{bmatrix} \Rightarrow \tilde{\bar{X}} = \begin{bmatrix} x_1 & x_2 - iy_2 \\ x_2 + iy_2 & x_3 \end{bmatrix} \Rightarrow \tilde{X}^* = \tilde{X}.$$

Then, the determinants are

$$\begin{aligned} \det(\tilde{X}) &= x_1x_3 - (x_2 - iy_2)(x_2 + iy_2) = x_1x_3 - (x_2^2 + y_2^2) \\ &= \det(\tilde{X}^*), \quad x_1 > 0, \quad x_3 > 0, \quad x_1x_3 - (x_2^2 + y_2^2) > 0, \end{aligned}$$

due to Hermitian positive definiteness of \tilde{X} . As well,

$$|\det(\tilde{X})| = +[(\det(\tilde{X})(\det(\tilde{X}^*)))]^{\frac{1}{2}} = x_1x_3 - (x_2^2 + y_2^2) > 0.$$

Note that $\text{tr}(\tilde{X}) = x_1 + x_3$ and $\tilde{\Gamma}_2(\alpha) = \pi^{\frac{2(1)}{2}} \Gamma(\alpha)\Gamma(\alpha - 1)$, $\Re(\alpha) > 1$, $p = 2$. The density is then of the following form:

$$\begin{aligned} f_1(\tilde{X}) &= \frac{1}{\tilde{\Gamma}_2(\alpha)} |\det(\tilde{X})|^{\alpha-2} e^{-\text{tr}(\tilde{X})} \\ &= \frac{1}{\pi \Gamma(\alpha)\Gamma(\alpha - 1)} [x_1x_3 - (x_2^2 + y_2^2)]^{\alpha-2} e^{-(x_1+x_3)} \end{aligned}$$

for $x_1 > 0$, $x_3 > 0$, $x_1x_3 - (x_2^2 + y_2^2) > 0$, $\Re(\alpha) > 1$, and $f_1(\tilde{X}) = 0$ elsewhere.

Now, consider a $p \times p$ parameter matrix $\tilde{B} > O$. We can obtain the following identity corresponding to the identity in the real case:

$$|\det(\tilde{B})|^{-\alpha} \equiv \frac{1}{\tilde{\Gamma}_p(\alpha)} \int_{\tilde{X} > O} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{B}\tilde{X})} d\tilde{X}, \quad \Re(\alpha) > p - 1. \quad (5.2a.3)$$

A two-parameter gamma density in the complex domain can then be derived by proceeding as in the real case; it is given by

$$\tilde{f}(\tilde{X}) = \begin{cases} \frac{|\det(\tilde{B})|^\alpha}{\tilde{\Gamma}_p(\alpha)} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{B}\tilde{X})}, & \tilde{B} > O, \tilde{X} > O, \Re(\alpha) > p - 1 \\ 0, & \text{elsewhere.} \end{cases} \quad (5.2a.4)$$

5.2a.1. The mgf of the complex matrix-variate gamma distribution

The moment generating function in the complex domain is slightly different from that in the real case. Let $\tilde{T} > O$ be a $p \times p$ parameter matrix and let \tilde{X} be $p \times p$ two-parameter gamma distributed as in (5.2a.4). Then $\tilde{T} = T_1 + iT_2$ and $\tilde{X} = X_1 + iX_2$, with T_1, T_2, X_1, X_2 real and $i = \sqrt{-1}$. When \tilde{T} and \tilde{X} are Hermitian positive definite, T_1 and X_1 are real symmetric and T_2 and X_2 are real skew symmetric. Then consider

$$\text{tr}(\tilde{T}^* \tilde{X}) = \text{tr}(T_1 X_1) + \text{tr}(T_2 X_2) + i[\text{tr}(T_1 X_2) - \text{tr}(T_2 X_1)].$$

Note that $\text{tr}(T_1 X_1) + \text{tr}(T_2 X_2)$ contains all the real variables involved multiplied by the corresponding parameters, where the diagonal elements appear once and the off-diagonal elements each appear twice. Thus, as in the real case, \tilde{T} has to be replaced by $*\tilde{T} = *T_1 + i*T_2$. A term containing i still remains; however, as a result of the following properties, this term will disappear.

Lemma 5.2a.1. *Let $\tilde{T}, \tilde{X}, T_1, T_2, X_1, X_2$ be as defined above. Then, $\text{tr}(T_1 X_2) = 0$, $\text{tr}(T_2 X_1) = 0$, $\text{tr}(*T_1 X_2) = 0$, $\text{tr}(*T_2 X_1) = 0$.*

Proof: For any real square matrix A , $\text{tr}(A) = \text{tr}(A')$ and for any two matrices A and B where AB and BA are defined, $\text{tr}(AB) = \text{tr}(BA)$. With the help of these two results, we have the following:

$$\text{tr}(T_1 X_2) = \text{tr}(T_1 X_2)' = \text{tr}(X_2' T_1') = -\text{tr}(X_2 T_1) = -\text{tr}(T_1 X_2)$$

as T_1 is symmetric and X_2 is skew symmetric. Now, $\text{tr}(T_1 X_2) = -\text{tr}(T_1 X_2) \Rightarrow \text{tr}(T_1 X_2) = 0$ since it is a real quantity. It can be similarly established that the other results stated in the lemma hold.

We may now define the mgf in the complex case, denoted by $M_{\tilde{X}}(*T)$, as follows:

$$\begin{aligned} M_{\tilde{X}}(*\tilde{T}) &= E[e^{\text{tr}(*\tilde{T}^* \tilde{X})}] = \int_{\tilde{X} > O} e^{\text{tr}(*\tilde{T}^* \tilde{X})} \tilde{f}(\tilde{X}) d\tilde{X} \\ &= \frac{|\det(\tilde{B})|^\alpha}{\tilde{\Gamma}_p(\alpha)} \int_{\tilde{X} > O} e^{-\text{tr}(\tilde{B} - *\tilde{T}^*) \tilde{X}} d\tilde{X}. \end{aligned}$$

Since $\text{tr}(\tilde{X}(\tilde{B} - *\tilde{T}^*)) = \text{tr}(C\tilde{X}C^*)$ for $C = (\tilde{B} - *\tilde{T}^*)^{\frac{1}{2}}$ and $C > O$, it follows from Theorem 1:6a.5 that $\tilde{Y} = C\tilde{X}C^* \Rightarrow d\tilde{Y} = |\det(CC^*)|^p d\tilde{X}$, that is, $d\tilde{X} = |\det(\tilde{B} - *\tilde{T}^*)|^{-p} d\tilde{Y}$ for $\tilde{B} - *\tilde{T}^* > O$. Then,

$$\begin{aligned}
 M_{\tilde{X}}(*\tilde{T}) &= \frac{|\det(\tilde{B})|^\alpha}{\tilde{I}_p(\alpha)} |\det(\tilde{B} - *\tilde{T}^*)|^{-\alpha} \int_{\tilde{Y} > O} |\det(\tilde{Y})|^{\alpha-p} e^{-\text{tr}(\tilde{Y})} d\tilde{Y} \\
 &= |\det(\tilde{B})|^\alpha |\det(\tilde{B} - *\tilde{T}^*)|^{-\alpha} \text{ for } \tilde{B} - *\tilde{T}^* > O \\
 &= |\det(I - \tilde{B}^{-1} *\tilde{T}^*)|^{-\alpha}, \quad I - \tilde{B}^{-1} *\tilde{T}^* > O.
 \end{aligned}
 \tag{5.2a.5}$$

For example, let $p = 2$ and

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{12}^* & \tilde{x}_{22} \end{bmatrix}, \quad B = \begin{bmatrix} 3 & i \\ -i & 2 \end{bmatrix} \text{ and } *\tilde{T} = \begin{bmatrix} *\tilde{t}_{11} & *\tilde{t}_{12} \\ *\tilde{t}_{12}^* & *\tilde{t}_{22} \end{bmatrix},$$

with $\tilde{x}_{21} = \tilde{x}_{12}^*$ and $*\tilde{t}_{21} = *\tilde{t}_{12}^*$. In this case, the conjugate transpose is only the conjugate since the quantities are scalar. Note that $B = B^*$ and hence B is Hermitian. The leading minors of B being $|(3)| = 3 > 0$ and $|B| = (3)(2) - (-i)(i) = 5 > 0$, B is Hermitian positive definite. Accordingly,

$$\begin{aligned}
 M_{\tilde{X}}(*\tilde{T}) &= |\det(B)|^\alpha |\det(B - *\tilde{T}^*)|^{-\alpha} \\
 &= 5^\alpha [(3 - *\tilde{t}_{11}^*)(2 - *\tilde{t}_{22}^*) + (i + *\tilde{t}_{12}^*)(i - *\tilde{t}_{12})].
 \end{aligned}$$

Now, consider the partitioning of the following $p \times p$ matrices:

$$\tilde{X} = \begin{bmatrix} \tilde{X}_{11} & \tilde{X}_{12} \\ \tilde{X}_{21} & \tilde{X}_{22} \end{bmatrix}, \quad *\tilde{T} = \begin{bmatrix} *\tilde{T}_{11} & *\tilde{T}_{12} \\ *\tilde{T}_{21} & *\tilde{T}_{22} \end{bmatrix} \text{ and } \tilde{B} = \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix} \tag{i}$$

where \tilde{X}_{11} and $*\tilde{T}_{11}$ are $r \times r$, $r \leq p$. Then, proceeding as in the real case, we have the following results:

Theorem 5.2a.1. *Let \tilde{X} have a $p \times p$ complex matrix-variate gamma density with shape parameter α and scale parameter I_p , and \tilde{X} be partitioned as in (i). Then, \tilde{X}_{11} has an $r \times r$ complex matrix-variate gamma density with shape parameter α and scale parameter I_r and \tilde{X}_{22} has a $(p - r) \times (p - r)$ complex matrix-variate gamma density with shape parameter α and scale parameter I_{p-r} .*

Theorem 5.2a.2. *Let the $p \times p$ complex matrix \tilde{X} have a $p \times p$ complex matrix-variate gamma density with the parameters $(\alpha, \tilde{B} > O)$ and let \tilde{X} and \tilde{B} be partitioned as in (i) and $\tilde{B}_{12} = O$, $\tilde{B}_{21} = O$. Then \tilde{X}_{11} and \tilde{X}_{22} have $r \times r$ and $(p - r) \times (p - r)$ complex matrix-variate gamma densities with shape parameter α and scale parameters \tilde{B}_{11} and \tilde{B}_{22} , respectively.*

Theorem 5.2a.3. *Let \tilde{X} , \tilde{X}_{11} , \tilde{X}_{22} and \tilde{B} be as specified in Theorems 5.2a.1 or 5.2a.2. Then, \tilde{X}_{11} and \tilde{X}_{22} are statistically independently distributed as complex matrix-variate gamma random variables on $r \times r$ and $(p - r) \times (p - r)$ matrices, respectively.*

For a general matrix B where the sub-matrices B_{12} and B_{21} are not assumed to be null, the marginal densities of \tilde{X}_{11} and \tilde{X}_{22} being given in the next result can be determined by proceeding as in the real case.

Theorem 5.2a.4. *Let \tilde{X} have a complex matrix-variate gamma density with shape parameter α and scale parameter matrix $B = B^* > O$. Letting \tilde{X} and B be partitioned as in (i), then the sub-matrix \tilde{X}_{11} has a complex matrix-variate gamma density with shape parameter α and scale parameter matrix $B_{11} - B_{12}B_{22}^{-1}B_{21}$, and the sub-matrix \tilde{X}_{22} has a complex matrix-variate gamma density with shape parameter α and scale parameter matrix $B_{22} - B_{21}B_{11}^{-1}B_{12}$.*

Exercises 5.2

5.2.1. Show that

$$\pi^{\frac{(p-r)(p-r-1)}{4}} \pi^{\frac{r(r-1)}{4}} \pi^{\frac{2r(p-r)}{4}} = \pi^{\frac{p(p-1)}{4}}.$$

5.2.2. Show that $\Gamma_r(\alpha)\Gamma_{p-r}(\alpha - \frac{r}{2}) = \Gamma_p(\alpha)$.

5.2.3. Evaluate (1): $\int_{X>O} e^{-\text{tr}(X)} dX$, (2): $\int_{X>O} |X| e^{-\text{tr}(X)} dX$.

5.2.4. Write down (1): $\Gamma_3(\alpha)$, (2): $\Gamma_4(\alpha)$ explicitly in the real and complex cases.

5.2.5. Evaluate the integrals in Exercise 5.2.3 for the complex case. In (2) replace $\det(X)$ by $|\det(X)|$.

5.3. Matrix-variate Type-1 Beta and Type-2 Beta Densities, Real Case

The $p \times p$ matrix-variate beta function denoted by $B_p(\alpha, \beta)$ is defined as follows in the real case:

$$B_p(\alpha, \beta) = \frac{\Gamma_p(\alpha)\Gamma_p(\beta)}{\Gamma_p(\alpha + \beta)}, \Re(\alpha) > \frac{p-1}{2}, \Re(\beta) > \frac{p-1}{2}. \quad (5.3.1)$$

This function has the following integral representations in the real case where it is assumed that $\Re(\alpha) > \frac{p-1}{2}$ and $\Re(\beta) > \frac{p-1}{2}$:

$$B_p(\alpha, \beta) = \int_{0 < X < I} |X|^{\alpha - \frac{p+1}{2}} |I - X|^{\beta - \frac{p+1}{2}} dX, \text{ a type-1 beta integral} \quad (5.3.2)$$

$$B_p(\beta, \alpha) = \int_{0 < Y < I} |Y|^{\beta - \frac{p+1}{2}} |I - Y|^{\alpha - \frac{p+1}{2}} dY, \text{ a type-1 beta integral} \quad (5.3.3)$$

$$B_p(\alpha, \beta) = \int_{Z > 0} |Z|^{\alpha - \frac{p+1}{2}} |I + Z|^{-(\alpha + \beta)} dZ, \text{ a type-2 beta integral} \quad (5.3.4)$$

$$B_p(\beta, \alpha) = \int_{T > 0} |T|^{\beta - \frac{p+1}{2}} |I + T|^{-(\alpha + \beta)} dT, \text{ a type-2 beta integral.} \quad (5.3.5)$$

For example, for $p = 2$, let

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{bmatrix} \Rightarrow |X| = x_{11}x_{22} - x_{12}^2 \text{ and } |I - X| = (1 - x_{11})(1 - x_{22}) - x_{12}^2.$$

Then for example, (5.3.2) will be of the following form:

$$\begin{aligned} B_p(\alpha, \beta) &= \int_{X > 0} |X|^{\alpha - \frac{p+1}{2}} |I - X|^{\beta - \frac{p+1}{2}} dX \\ &= \int_{x_{11} > 0} \int_{x_{22} > 0} \int_{x_{11}x_{22} - x_{12}^2 > 0} [x_{11}x_{22} - x_{12}^2]^{\alpha - \frac{3}{2}} \\ &\quad \times [(1 - x_{11})(1 - x_{22}) - x_{12}^2]^{\beta - \frac{3}{2}} dx_{11} \wedge dx_{12} \wedge dx_{22}. \end{aligned}$$

We will derive two of the integrals (5.3.2)–(5.3.5), the other ones being then directly obtained. Let us begin with the integral representations of $\Gamma_p(\alpha)$ and $\Gamma_p(\beta)$ for $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$:

$$\begin{aligned} \Gamma_p(\alpha)\Gamma_p(\beta) &= \left[\int_{X > 0} |X|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(X)} dX \right] \left[\int_{Y > 0} |Y|^{\beta - \frac{p+1}{2}} e^{-\text{tr}(Y)} dY \right] \\ &= \int_X \int_Y |X|^{\alpha - \frac{p+1}{2}} |Y|^{\beta - \frac{p+1}{2}} e^{-\text{tr}(X+Y)} dX \wedge dY. \end{aligned}$$

Making the transformation $U = X + Y$, $X = V$, whose Jacobian is 1, taking out U from $|U - V| = |U| |I - U^{-\frac{1}{2}} V U^{-\frac{1}{2}}|$, and then letting $W = U^{-\frac{1}{2}} V U^{-\frac{1}{2}} \Rightarrow dV = |U|^{\frac{p+1}{2}} dW$, we have

$$\begin{aligned}
\Gamma_p(\alpha)\Gamma_p(\beta) &= \int_U \int_V |V|^{\alpha-\frac{p+1}{2}} |U-V|^{\beta-\frac{p+1}{2}} e^{-\text{tr}(U)} dU \wedge dV \\
&= \left\{ \int_{U>O} |U|^{\alpha+\beta-\frac{p+1}{2}} e^{-\text{tr}(U)} dU \right\} \\
&\quad \times \left\{ \int_{O<W<I} |W|^{\alpha-\frac{p+1}{2}} |I-W|^{\beta-\frac{p+1}{2}} dW \right\} \\
&= \Gamma_p(\alpha+\beta) \int_{O<W<I} |W|^{\alpha-\frac{p+1}{2}} |I-W|^{\beta-\frac{p+1}{2}} dW.
\end{aligned}$$

Thus, on dividing both sides by $\Gamma_p(\alpha, \beta)$, we have

$$B_p(\alpha, \beta) = \int_{O<W<I} |W|^{\alpha-\frac{p+1}{2}} |I-W|^{\beta-\frac{p+1}{2}} dW. \quad (i)$$

This establishes (5.3.2). The initial conditions $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$ are sufficient to justify all the steps above, and hence no conditions are listed at each stage. Now, take $Y = I - W$ to obtain (5.3.3). Let us take the W of (i) above and consider the transformation

$$Z = (I - W)^{-\frac{1}{2}} W (I - W)^{-\frac{1}{2}} = (W^{-1} - I)^{-\frac{1}{2}} (W^{-1} - I)^{-\frac{1}{2}} = (W^{-1} - I)^{-1}$$

which gives

$$Z^{-1} = W^{-1} - I \Rightarrow |Z|^{-(p+1)} dZ = |W|^{-(p+1)} dW. \quad (ii)$$

Taking determinants and substituting in (ii) we have

$$dW = |I + Z|^{-(p+1)} dZ.$$

On expressing W , $I - W$ and dW in terms of Z , we have the result (5.3.4). Now, let $T = Z^{-1}$ with the Jacobian $dT = |Z|^{-(p+1)} dZ$, then (5.3.4) transforms into the integral (5.3.5). These establish all four integral representations of the real matrix-variate beta function. We may also observe that $B_p(\alpha, \beta) = B_p(\beta, \alpha)$ or α and β can be interchanged in the beta function. Consider the function

$$f_3(X) = \frac{\Gamma_p(\alpha + \beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)} |X|^{\alpha-\frac{p+1}{2}} |I-X|^{\beta-\frac{p+1}{2}} \quad (5.3.6)$$

for $O < X < I$, $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$, and $f_3(X) = 0$ elsewhere. This is a type-1 real matrix-variate beta density with the parameters (α, β) , where $O < X < I$ means $X > O$, $I - X > O$ so that all the eigenvalues of X are in the open interval $(0, 1)$. As for

$$f_4(Z) = \frac{\Gamma_p(\alpha + \beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)} |Z|^{\alpha-\frac{p+1}{2}} |I+Z|^{-(\alpha+\beta)} \quad (5.3.7)$$

whenever $Z > O$, $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$, and $f_4(Z) = 0$ elsewhere, this is a $p \times p$ real matrix-variate type-2 beta density with the parameters (α, β) .

5.3.1. Some properties of real matrix-variate type-1 and type-2 beta densities

In the course of the above derivations, it was shown that the following results hold. If X is a $p \times p$ real positive definite matrix having a real matrix-variate type-1 beta density with the parameters (α, β) , then:

- (1): $Y_1 = I - X$ is real type-1 beta distributed with the parameters (β, α) ;
- (2): $Y_2 = (I - X)^{-\frac{1}{2}} X (I - X)^{-\frac{1}{2}}$ is real type-2 beta distributed with the parameters (α, β) ;
- (3): $Y_3 = (I - X)^{\frac{1}{2}} X^{-1} (I - X)^{\frac{1}{2}}$ is real type-2 beta distributed with the parameters (β, α) .
If Y is real type-2 beta distributed with the parameters (α, β) then:
- (4): $Z_1 = Y^{-1}$ is real type-2 beta distributed with the parameters (β, α) ;
- (5): $Z_2 = (I + Y)^{-\frac{1}{2}} Y (I + Y)^{-\frac{1}{2}}$ is real type-1 beta distributed with the parameters (α, β) ;
- (6): $Z_3 = I - (I + Y)^{-\frac{1}{2}} Y (I + Y)^{-\frac{1}{2}} = (I + Y)^{-1}$ is real type-1 beta distributed with the parameters (β, α) .

5.3a. Matrix-variate Type-1 and Type-2 Beta Densities, Complex Case

A matrix-variate beta function in the complex domain is defined as

$$\tilde{B}_p(\alpha, \beta) = \frac{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(\beta)}{\tilde{\Gamma}_p(\alpha + \beta)}, \Re(\alpha) > p - 1, \Re(\beta) > p - 1 \tag{5.3a.1}$$

with a tilde over B . As $\tilde{B}_p(\alpha, \beta) = \tilde{B}_p(\beta, \alpha)$, clearly α and β can be interchanged. Then, $\tilde{B}_p(\alpha, \beta)$ has the following integral representations, where $\Re(\alpha) > p - 1, \Re(\beta) > p - 1$:

$$\tilde{B}_p(\alpha, \beta) = \int_{O < \tilde{X} < I} |\det(\tilde{X})|^{\alpha-p} |\det(I - \tilde{X})|^{\beta-p} d\tilde{X}, \text{ a type-1 beta integral} \tag{5.3a.2}$$

$$\tilde{B}_p(\beta, \alpha) = \int_{O < \tilde{Y} < I} |\det(\tilde{Y})|^{\beta-p} |\det(I - \tilde{Y})|^{\alpha-p} d\tilde{Y}, \text{ a type-1 beta integral} \tag{5.3a.3}$$

$$\tilde{B}_p(\alpha, \beta) = \int_{\tilde{Z} > O} |\det(\tilde{Z})|^{\alpha-p} |\det(I + \tilde{Z})|^{-(\alpha+\beta)} d\tilde{Z}, \text{ a type-2 beta integral} \tag{5.3a.4}$$

$$\tilde{B}_p(\beta, \alpha) = \int_{\tilde{T} > O} |\det(\tilde{T})|^{\beta-p} |\det(I + \tilde{T})|^{-(\alpha+\beta)} d\tilde{T}, \text{ a type-2 beta integral.} \tag{5.3a.5}$$

For instance, consider the integrand in (5.3a.2) for the case $p = 2$. Let

$$\tilde{X} = \begin{bmatrix} x_1 & x_2 + iy_2 \\ x_2 - iy_2 & x_3 \end{bmatrix}, \quad \tilde{X} = \tilde{X}^* > O, \quad i = \sqrt{-1},$$

the diagonal elements of \tilde{X} being real; since \tilde{X} is Hermitian positive definite, we have $x_1 > 0$, $x_3 > 0$, $x_1x_2 - (x_2^2 + y_2^2) > 0$, $\det(\tilde{X}) = x_1x_3 - (x_2^2 + y_2^2) > 0$ and $\det(I - \tilde{X}) = (1 - x_1)(1 - x_3) - (x_2^2 + y_2^2) > 0$. The integrand in (5.3a.2) is then

$$[x_1x_3 - (x_2^2 + y_2^2)]^{\alpha - \frac{3}{2}} [(1 - x_1)(1 - x_3) - (x_2^2 + y_2^2)]^{\beta - \frac{3}{2}}.$$

The derivations of (5.3a.2)–(5.3a.5) being parallel to those provided in the real case, they are omitted. We will list one case for each of a type-1 and a type-2 beta density in the complex $p \times p$ matrix-variate case:

$$\tilde{f}_3(\tilde{X}) = \frac{\tilde{\Gamma}_p(\alpha + \beta)}{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(\beta)} |\det(\tilde{X})|^{\alpha - p} |\det(I - \tilde{X})|^{\beta - p} \quad (5.3a.6)$$

for $O < \tilde{X} < I$, $\Re(\alpha) > p - 1$, $\Re(\beta) > p - 1$ and $\tilde{f}_3(\tilde{X}) = 0$ elsewhere;

$$\tilde{f}_4(\tilde{Z}) = \frac{\tilde{\Gamma}_p(\alpha + \beta)}{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(\beta)} |\det(\tilde{Z})|^{\alpha - p} |\det(I + \tilde{Z})|^{-(\alpha + \beta)} \quad (5.3a.7)$$

for $\tilde{Z} > O$, $\Re(\alpha) > p - 1$, $\Re(\beta) > p - 1$ and $\tilde{f}_4(\tilde{Z}) = 0$ elsewhere.

Properties parallel to (1) to (6) which are listed in Sect. 5.3.1 also hold in the complex case.

5.3.2. Explicit evaluation of type-1 matrix-variate beta integrals, real case

A detailed evaluation of a type-1 matrix-variate beta integral as a multiple integral is presented in this section as the steps will prove useful in connection with other computations; the reader may also refer to Mathai (2014,b). The real matrix-variate type-1 beta function which is denoted by

$$B_p(\alpha, \beta) = \frac{\Gamma_p(\alpha)\Gamma_p(\beta)}{\Gamma_p(\alpha + \beta)}, \quad \Re(\alpha) > \frac{p-1}{2}, \quad \Re(\beta) > \frac{p-1}{2},$$

has the following type-1 beta integral representation:

$$B_p(\alpha, \beta) = \int_{O < X < I} |X|^{\alpha - \frac{p+1}{2}} |I - X|^{\beta - \frac{p+1}{2}} dX,$$

for $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$ where X is a real $p \times p$ symmetric positive definite matrix. The standard derivation of this integral relies on the properties of real matrix-variate gamma integrals after making suitable transformations, as was previously done. It is also possible to evaluate the integral directly and show that it is equal to $\Gamma_p(\alpha)\Gamma_p(\beta)/\Gamma_p(\alpha + \beta)$ where, for example,

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha)\Gamma(\alpha - 1/2) \cdots \Gamma(\alpha - (p - 1)/2), \quad \Re(\alpha) > \frac{p - 1}{2}.$$

A convenient technique for evaluating a real matrix-variate gamma integral consists of making the transformation $X = TT'$ where T is a lower triangular matrix whose diagonal elements are positive. However, on applying this transformation, the type-1 beta integral does not simplify due to the presence of the factor $|I - X|^{\beta - \frac{p+1}{2}}$. Hence, we will attempt to evaluate this integral by appropriately partitioning the matrices and then, successively integrating out the variables. Letting $X = (x_{ij})$ be a $p \times p$ real matrix, x_{pp} can then be extracted from the determinants of $|X|$ and $|I - X|$ after partitioning the matrices. Thus, let

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

where X_{11} is the $(p - 1) \times (p - 1)$ leading sub-matrix, X_{21} is $1 \times (p - 1)$, $X_{22} = x_{pp}$ and $X_{12} = X'_{21}$. Then $|X| = |X_{11}||x_{pp} - X_{21}X_{11}^{-1}X_{12}|$ so that

$$|X|^{\alpha - \frac{p+1}{2}} = |X_{11}|^{\alpha - \frac{p+1}{2}} [x_{pp} - X_{21}X_{11}^{-1}X_{12}]^{\alpha - \frac{p+1}{2}}, \tag{i}$$

and

$$|I - X|^{\beta - \frac{p+1}{2}} = |I - X_{11}|^{\beta - \frac{p+1}{2}} [(1 - x_{pp}) - X_{21}(I - X_{11})^{-1}X_{12}]^{\beta - \frac{p+1}{2}}. \tag{ii}$$

It follows from (i) that $x_{pp} > X_{21}X_{11}^{-1}X_{12}$ and, from (ii) that $x_{pp} < 1 - X_{21}(I - X_{11})^{-1}X_{12}$; thus, we have $X_{21}X_{11}^{-1}X_{12} < x_{pp} < 1 - X_{21}(I - X_{11})^{-1}X_{12}$. Let $y = x_{pp} - X_{21}X_{11}^{-1}X_{12} \Rightarrow dy = dx_{pp}$ for fixed X_{21}, X_{11} , so that $0 < y < b$ where

$$\begin{aligned} b &= 1 - X_{21}X_{11}^{-1}X_{12} - X_{21}(I - X_{11})^{-1}X_{12} \\ &= 1 - X_{21}X_{11}^{-\frac{1}{2}}(I - X_{11})^{-\frac{1}{2}}(I - X_{11})^{-\frac{1}{2}}X_{11}^{-\frac{1}{2}}X_{12} \\ &= 1 - WW', \quad W = X_{21}X_{11}^{-\frac{1}{2}}(I - X_{11})^{-\frac{1}{2}}. \end{aligned}$$

The second factor on the right-hand side of (ii) then becomes

$$[b - y]^{\beta - \frac{p+1}{2}} = b^{\beta - \frac{p+1}{2}} [1 - y/b]^{\beta - \frac{p+1}{2}}.$$

Now letting $u = \frac{y}{b}$ for fixed b , the terms containing u and b become $b^{\alpha + \beta - (p+1) + 1} u^{\alpha - \frac{p+1}{2}} (1 - u)^{\beta - \frac{p+1}{2}}$. Integration over u then gives

$$\int_0^1 u^{\alpha - \frac{p+1}{2}} (1 - u)^{\beta - \frac{p+1}{2}} du = \frac{\Gamma(\alpha - \frac{p-1}{2}) \Gamma(\beta - \frac{p-1}{2})}{\Gamma(\alpha + \beta - (p - 1))}$$

for $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$. Letting $W = X_{21} X_{11}^{-\frac{1}{2}} (I - X_{11})^{-\frac{1}{2}}$ for fixed X_{11} , $dX_{21} = |X_{11}|^{\frac{1}{2}} |I - X_{11}|^{\frac{1}{2}} dW$ from Theorem 1.6.1 of Chap. 1 or Theorem 1.18 of Mathai (1997), where X_{11} is a $(p - 1) \times (p - 1)$ matrix. Now, letting $v = WW'$ and integrating out over the Stiefel manifold by applying Theorem 4.2.3 of Chap. 4 or Theorem 2.16 and Remark 2.13 of Mathai (1997), we have

$$dW = \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} v^{\frac{p-1}{2} - 1} dv.$$

Thus, the integral over b becomes

$$\begin{aligned} \int b^{\alpha + \beta - p} dX_{21} &= \int_0^1 v^{\frac{p-1}{2} - 1} (1 - v)^{\alpha + \beta - p} dv \\ &= \frac{\Gamma(\frac{p-1}{2}) \Gamma(\alpha + \beta - (p - 1))}{\Gamma(\alpha + \beta - \frac{p-1}{2})}, \quad \Re(\alpha + \beta) > p - 1. \end{aligned}$$

Then, on multiplying all the factors together, we have

$$|X_{11}^{(1)}|^{\alpha + \frac{1}{2} - \frac{p+1}{2}} |I - X_{11}^{(1)}|^{\beta + \frac{1}{2} - \frac{p+1}{2}} \pi^{\frac{p-1}{2}} \frac{\Gamma(\alpha - \frac{p-1}{2}) \Gamma(\beta - \frac{p-1}{2})}{\Gamma(\alpha + \beta - \frac{p-1}{2})}$$

whenever $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\beta) > \frac{p-1}{2}$. In this case, $X_{11}^{(1)}$ represents the $(p - 1) \times (p - 1)$ leading sub-matrix at the end of the first set of operations. At the end of the second set of operations, we will denote the $(p - 2) \times (p - 2)$ leading sub-matrix by $X_{11}^{(2)}$, and so on. The second step of the operations begins by extracting $x_{p-1, p-1}$ and writing

$$|X_{11}^{(1)}| = |X_{11}^{(2)}| [x_{p-1, p-1} - X_{21}^{(2)} [X_{11}^{(2)}]^{-1} X_{12}^{(2)}]$$

where $X_{21}^{(2)}$ is a $1 \times (p - 2)$ vector. We then proceed as in the first sequence of steps to obtain the final factors in the following form:

$$|X_{11}^{(2)}|^{\alpha+1-\frac{p+1}{2}} |I - X_{11}^{(2)}|^{\beta+1-\frac{p+1}{2}} \pi^{\frac{p-2}{2}} \frac{\Gamma(\alpha - \frac{p-2}{2})\Gamma(\beta - \frac{p-2}{2})}{\Gamma(\alpha + \beta - \frac{p-2}{2})}$$

for $\Re(\alpha) > \frac{p-2}{2}$, $\Re(\beta) > \frac{p-2}{2}$. Proceeding in such a manner, in the end, the exponent of π will be

$$\frac{p-1}{2} + \frac{p-2}{2} + \dots + \frac{1}{2} = \frac{p(p-1)}{4},$$

and the gamma product will be

$$\frac{\Gamma(\alpha - \frac{p-1}{2})\Gamma(\alpha - \frac{p-2}{2}) \dots \Gamma(\alpha)\Gamma(\beta - \frac{p-1}{2}) \dots \Gamma(\beta)}{\Gamma(\alpha + \beta - \frac{p-1}{2}) \dots \Gamma(\alpha + \beta)}.$$

These gamma products, along with $\pi^{\frac{p(p-1)}{4}}$, can be written as $\frac{\Gamma_p(\alpha)\Gamma_p(\beta)}{\Gamma_p(\alpha+\beta)} = B_p(\alpha, \beta)$; hence the result. It is thus possible to obtain the beta function in the real matrix-variate case by direct evaluation of a type-1 real matrix-variate beta integral.

A similar approach can yield the real matrix-variate beta function from a type-2 real matrix-variate beta integral of the form

$$\int_{X>O} |X|^{\alpha-\frac{p+1}{2}} |I + X|^{-(\alpha+\beta)} dX$$

where X is a $p \times p$ positive definite symmetric matrix and it is assumed that $\Re(\alpha) > \frac{p-1}{2}$ and $\Re(\beta) > \frac{p-1}{2}$, the evaluation procedure being parallel.

Example 5.3.1. By direct evaluation as a multiple integral, show that

$$\int_{X>O} |X|^{\alpha-\frac{p+1}{2}} |I - X|^{\beta-\frac{p+1}{2}} dX = \frac{\Gamma_p(\alpha)\Gamma_p(\beta)}{\Gamma_p(\alpha + \beta)}$$

for $p = 2$.

Solution 5.3.1. The integral to be evaluated will be denoted by δ . Let

$$|X| = x_{11}[x_{22} - x_{21}x_{11}^{-1}x_{12}] = x_{11}\left[x_{22} - \frac{x_{12}^2}{x_{11}}\right]$$

$$|I - X| = [1 - x_{11}][(1 - x_{22}) - x_{12}(1 - x_{11})^{-1}x_{12}] = (1 - x_{11})\left[1 - x_{22} - \frac{x_{12}^2}{1 - x_{11}}\right]. \tag{i}$$

It is seen from (i) that

$$\frac{x_{12}^2}{x_{11}} \leq x_{22} \leq 1 - \frac{x_{12}^2}{1 - x_{11}}.$$

Letting $y = x_{22} - \frac{x_{12}^2}{x_{11}}$ so that $0 \leq y \leq b$, and $b = 1 - \frac{x_{12}^2}{x_{11}} - \frac{x_{12}^2}{1 - x_{11}} = 1 - \frac{x_{12}^2}{x_{11}(1 - x_{11})}$, we have

$$\begin{aligned} |X|^{\alpha - \frac{3}{2}} |I - X|^{\beta - \frac{3}{2}} dX &= x_{11}^{\alpha - \frac{3}{2}} (1 - x_{11})^{\beta - \frac{3}{2}} y^{\alpha - \frac{3}{2}} \\ &\times (b - y)^{\beta - \frac{3}{2}} dx_{11} \wedge dx_{22} \wedge dy. \end{aligned}$$

Now, integrating out y , we have

$$\begin{aligned} \int_{y=0}^b y^{\alpha - \frac{3}{2}} (b - y)^{\beta - \frac{3}{2}} dy &= b^{\alpha + \beta - 3 + 1} \int_0^1 v^{\alpha - \frac{3}{2}} (1 - v)^{\beta - \frac{3}{2}} dv, \quad v = \frac{y}{b} \\ &= b^{\alpha + \beta - 2} \frac{\Gamma(\alpha - \frac{1}{2}) \Gamma(\beta - \frac{1}{2})}{\Gamma(\alpha + \beta - 1)} \end{aligned} \quad (ii)$$

whenever $\Re(\alpha) > \frac{1}{2}$ and $\Re(\beta) > \frac{1}{2}$, b being as previously defined. Letting $w = \frac{x_{12}}{[x(1-x)]^{\frac{1}{2}}}$,

$dx_{12} = [x_{11}(1 - x_{11})]^{\frac{1}{2}} dw$ for fixed x_{11} . The exponents of x_{11} and $(1 - x_{11})$ then become $\alpha - \frac{3}{2} + \frac{1}{2}$ and $\beta - \frac{3}{2} + \frac{1}{2}$, and the integral over w gives the following:

$$\begin{aligned} \int_{-1}^1 (1 - w^2)^{\alpha + \beta - 2} dw &= 2 \int_0^1 (1 - w^2)^{\alpha + \beta - 2} dw = \int_0^1 z^{\frac{1}{2} - 1} (1 - z)^{\alpha + \beta - 2} dz \\ &= \frac{\Gamma(\frac{1}{2}) \Gamma(\alpha + \beta - 1)}{\Gamma(\alpha + \beta - \frac{1}{2})}. \end{aligned} \quad (iii)$$

Now, integrating out x_{11} , we obtain

$$\int_0^1 x_{11}^{\alpha - 1} (1 - x_{11})^{\beta - 1} dx_{11} = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}. \quad (iv)$$

Then, on collecting the factors from (i) to (iv), we have

$$\delta = \Gamma(1/2) \frac{\Gamma(\alpha) \Gamma(\alpha - \frac{1}{2}) \Gamma(\beta) \Gamma(\beta - \frac{1}{2})}{\Gamma(\alpha + \beta) \Gamma(\alpha + \beta - \frac{1}{2})}.$$

Finally, noting that for $p = 2$, $\pi^{\frac{p(p-1)}{4}} = \pi^{\frac{1}{2}} = \frac{\pi^{\frac{1}{2}} \pi^{\frac{1}{2}}}{\pi^{\frac{1}{2}}}$, the desired result is obtained, that is,

$$\delta = \frac{\Gamma_2(\alpha) \Gamma_2(\beta)}{\Gamma_2(\alpha + \beta)} = B_2(\alpha, \beta).$$

This completes the computations.

5.3a.1. Evaluation of matrix-variate type-1 beta integrals, complex case

The integral representation for $B_p(\alpha, \beta)$ in the complex case is

$$\int_{O < \tilde{X} < I} |\det(\tilde{X})|^{\alpha-p} |\det(I - \tilde{X})|^{\beta-p} d\tilde{X} = \tilde{B}_p(\alpha, \beta)$$

whenever $\Re(\alpha) > p - 1$, $\Re(\beta) > p - 1$ where $\det(\cdot)$ denotes the determinant of (\cdot) and $|\det(\cdot)|$, the absolute value (or modulus) of the determinant of (\cdot) . In this case, $\tilde{X} = (\tilde{x}_{ij})$ is a $p \times p$ Hermitian positive definite matrix and accordingly, all of its diagonal elements are real and positive. As in the real case, let us extract x_{pp} by partitioning \tilde{X} as follows:

$$\tilde{X} = \begin{bmatrix} \tilde{X}_{11} & \tilde{X}_{12} \\ \tilde{X}_{21} & \tilde{X}_{22} \end{bmatrix} \text{ so that } I - \tilde{X} = \begin{bmatrix} I - \tilde{X}_{11} & -\tilde{X}_{12} \\ -\tilde{X}_{21} & I - \tilde{X}_{22} \end{bmatrix},$$

where $\tilde{X}_{22} \equiv x_{pp}$ is a real scalar. Then, the absolute value of the determinants have the following representations:

$$|\det(\tilde{X})|^{\alpha-p} = |\det(\tilde{X}_{11})|^{\alpha-p} |x_{pp} - \tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}^*|^{\alpha-p} \tag{i}$$

where * indicates conjugate transpose, and

$$|\det(I - \tilde{X})|^{\beta-p} = |\det(I - \tilde{X}_{11})|^{\beta-p} |(1 - x_{pp}) - \tilde{X}_{21}(I - \tilde{X}_{11})^{-1} \tilde{X}_{12}^*|^{\beta-p}. \tag{ii}$$

Note that whenever \tilde{X} and $I - \tilde{X}$ are Hermitian positive definite, \tilde{X}_{11}^{-1} and $(I - \tilde{X}_{11})^{-1}$ are too Hermitian positive definite. Further, the Hermitian forms $\tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}^*$ and $\tilde{X}_{21}(I - \tilde{X}_{11})^{-1} \tilde{X}_{12}^*$ remain real and positive. It follows from (i) and (ii) that

$$\tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}^* < x_{pp} < 1 - \tilde{X}_{21}(I - \tilde{X}_{11})^{-1} \tilde{X}_{12}^*.$$

Since the traces of Hermitian forms are real, the lower and upper bounds of x_{pp} are real as well. Let

$$\tilde{W} = \tilde{X}_{21} \tilde{X}_{11}^{-\frac{1}{2}} (I - \tilde{X}_{11})^{-\frac{1}{2}}$$

for fixed \tilde{X}_{11} . Then

$$d\tilde{X}_{21} = |\det(\tilde{X}_{11})|^{-1} |\det(I - \tilde{X}_{11})|^{-1} d\tilde{W}$$

and $|\det(\tilde{X})|^{\alpha-p}$, $|\det(I - \tilde{X}_{11})|^{\beta-p}$ will become $|\det(\tilde{X}_{11})|^{\alpha+1-p}$, $|\det(I - \tilde{X}_{11})|^{\beta+1-p}$, respectively. Then, we can write

$$\begin{aligned} & |(1 - x_{pp}) - \tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}^* - \tilde{X}_{21}(I - \tilde{X}_{11})^{-1} \tilde{X}_{12}^*|^{\beta-p} \\ & = (b - y)^{\beta-p} = b^{\beta-p} [1 - y/b]^{\beta-p}. \end{aligned}$$

Now, letting $u = y/b$, the factors containing u and b will be of the form $u^{\alpha-p} (1-u)^{\beta-p} b^{\alpha+\beta-2p+1}$; the integral over u then gives

$$\int_0^1 u^{\alpha-p} (1-u)^{\beta-p} du = \frac{\Gamma(\alpha - (p-1))\Gamma(\beta - (p-1))}{\Gamma(\alpha + \beta - 2(p-1))},$$

for $\Re(\alpha) > p-1$, $\Re(\beta) > p-1$. Letting $v = \tilde{W}\tilde{W}^*$ and integrating out over the Stiefel manifold by making use of Theorem 4.2a.3 of Chap. 4 or Corollaries 4.5.2 and 4.5.3 of Mathai (1997), we have

$$d\tilde{W} = \frac{\pi^{p-1}}{\Gamma(p-1)} v^{(p-1)-1} dv.$$

The integral over b gives

$$\begin{aligned} \int b^{\alpha+\beta-2p+1} d\tilde{X}_{21} &= \int_0^1 v^{(p-1)-1} (1-v)^{\alpha+\beta-2p+1} dv \\ &= \frac{\Gamma(p-1)\Gamma(\alpha + \beta - 2(p-1))}{\Gamma(\alpha + \beta - p + 1)}, \end{aligned}$$

for $\Re(\alpha) > p-1$, $\Re(\beta) > p-1$. Now, taking the product of all the factors yields

$$|\det(\tilde{X}_{11})|^{\alpha+1-p} |\det(I - \tilde{X}_{11})|^{\beta+1-p} \pi^{p-1} \frac{\Gamma(\alpha - p + 1)\Gamma(\beta - p + 1)}{\Gamma(\alpha + \beta - p + 1)}$$

for $\Re(\alpha) > p-1$, $\Re(\beta) > p-1$. On extracting $x_{p-1,p-1}$ from $|\tilde{X}_{11}|$ and $|I - \tilde{X}_{11}|$ and continuing this process, in the end, the exponent of π will be $(p-1) + (p-2) + \cdots + 1 = \frac{p(p-1)}{2}$ and the gamma product will be

$$\frac{\Gamma(\alpha - (p-1))\Gamma(\alpha - (p-2)) \cdots \Gamma(\alpha)\Gamma(\beta - (p-1)) \cdots \Gamma(\beta)}{\Gamma(\alpha + \beta - (p-1)) \cdots \Gamma(\alpha + \beta)}.$$

These factors, along with $\pi^{\frac{p(p-1)}{2}}$ give

$$\frac{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(\beta)}{\tilde{\Gamma}_p(\alpha + \beta)} = \tilde{B}_p(\alpha, \beta), \quad \Re(\alpha) > p-1, \quad \Re(\beta) > p-1.$$

The procedure for evaluating a type-2 matrix-variate beta integral by partitioning matrices is parallel and hence will not be detailed here.

Example 5.3a.1. For $p = 2$, evaluate the integral

$$\int_{O < \tilde{X} < I} |\det(\tilde{X})|^{\alpha-p} |\det(I - \tilde{X})|^{\beta-p} d\tilde{X}$$

as a multiple integral and show that it evaluates out to $\tilde{B}_2(\alpha, \beta)$, the beta function in the complex domain.

Solution 5.3a.1. For $p = 2$, $\pi^{\frac{p(p-1)}{2}} = \pi^{\frac{2(1)}{2}} = \pi$, and

$$\tilde{B}_2(\alpha, \beta) = \pi \frac{\Gamma(\alpha)\Gamma(\alpha-1)\Gamma(\beta)\Gamma(\beta-1)}{\Gamma(\alpha+\beta)\Gamma(\alpha+\beta-1)}$$

whenever $\Re(\alpha) > 1$ and $\Re(\beta) > 1$. For $p = 2$, our matrix and the relevant determinants are

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} \\ \tilde{x}_{12}^* & \tilde{x}_{22} \end{bmatrix}, \quad |\det(\tilde{X})| \text{ and } |\det(I - \tilde{X})|$$

where \tilde{x}_{12}^* is only the conjugate of \tilde{x}_{12} as it is a scalar quantity. By expanding the determinants of the partitioned matrices as explained in Sect. 1.3, we have the following:

$$\det(\tilde{X}) = \tilde{x}_{11}[\tilde{x}_{22} - \tilde{x}_{12}^*\tilde{x}_{11}^{-1}\tilde{x}_{12}] = \tilde{x}_{11}\left[\tilde{x}_{22} - \frac{\tilde{x}_{12}\tilde{x}_{12}^*}{\tilde{x}_{11}}\right] \quad (i)$$

$$\det(I - \tilde{X}) = (1 - \tilde{x}_{11})\left[1 - \tilde{x}_{22} - \frac{\tilde{x}_{12}\tilde{x}_{12}^*}{1 - \tilde{x}_{11}}\right]. \quad (ii)$$

From (i) and (ii), it is seen that

$$\frac{\tilde{x}_{12}\tilde{x}_{12}^*}{\tilde{x}_{11}} \leq \tilde{x}_{22} \leq 1 - \frac{\tilde{x}_{12}\tilde{x}_{12}^*}{1 - \tilde{x}_{11}}.$$

Note that when \tilde{X} is Hermitian, \tilde{x}_{11} and \tilde{x}_{22} are real and hence we may not place a tilde on these variables. Let $\tilde{y} = x_{22} - \tilde{x}_{12}\tilde{x}_{12}^*/x_{11}$. Note that \tilde{y} is also real since $\tilde{x}_{12}\tilde{x}_{12}^*$ is real. As well, $0 \leq y \leq b$, where

$$b = 1 - \left[\frac{\tilde{x}_{12}\tilde{x}_{12}^*}{x_{11}}\right] - \frac{\tilde{x}_{12}\tilde{x}_{12}^*}{1 - x_{11}} = 1 - \frac{\tilde{x}_{12}\tilde{x}_{12}^*}{x_{11}(1 - x_{11})}.$$

Further, b is a real scalar of the form $b = 1 - \tilde{w}\tilde{w}^*$ where $\tilde{w} = \frac{\tilde{x}_{12}}{[x_{11}(1-x_{11})]^{\frac{1}{2}}} \Rightarrow d\tilde{x}_{12} = x_{11}(1-x_{11})d\tilde{w}$. This will make the exponents of x_{11} and $(1-x_{11})$ as $\alpha - p + 1 = \alpha - 1$ and $\beta - 1$, respectively. Now, on integrating out y , we have

$$\int_{y=0}^b y^{\alpha-2}(b-y)^{\beta-2}dy = b^{\alpha+\beta-3} \frac{\Gamma(\alpha-1)\Gamma(\beta-1)}{\Gamma(\alpha+\beta-2)}, \quad \Re(\alpha) > 1, \quad \Re(\beta) > 1. \quad (iii)$$

Integrating out \tilde{w} , we have the following:

$$\int_{\tilde{w}} (1 - \tilde{w}\tilde{w}^*)^{\alpha+\beta-3} d\tilde{w} = \pi \frac{\Gamma(\alpha+\beta-2)}{\Gamma(\alpha+\beta-1)}. \quad (iv)$$

This integral is evaluated by writing $z = \tilde{w}\tilde{w}^*$. Then, it follows from Theorem 4.2a.3 that

$$d\tilde{w} = \frac{\pi^{p-1}}{\Gamma(p-1)} z^{(p-1)-1} dz = \pi dz \text{ for } p = 2.$$

Now, collecting all relevant factors from (i) to (iv), the required representation of the initial integral, denoted by δ , is obtained:

$$\delta = \pi \frac{\Gamma(\alpha)\Gamma(\alpha-1)\Gamma(\beta)\Gamma(\beta-1)}{\Gamma(\alpha+\beta)\Gamma(\alpha+\beta-1)} = \frac{\tilde{\Gamma}_2(\alpha)\tilde{\Gamma}_2(\beta)}{\tilde{\Gamma}_2(\alpha+\beta)} = \tilde{B}_2(\alpha+\beta)$$

whenever $\Re(\alpha) > 1$ and $\Re(\beta) > 1$. This completes the computations.

5.3.3. General partitions, real case

In Sect. 5.3.2, we have considered integrating one variable at a time by suitably partitioning the matrices. Would it also be possible to have a general partitioning and integrate a block of variables at a time, rather than integrating out individual variables? We will consider the real matrix-variate gamma integral first. Let the $p \times p$ positive definite matrix X be partitioned as follows:

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \quad X_{11} \text{ being } p_1 \times p_1 \text{ and } X_{22}, \quad p_2 \times p_2,$$

so that X_{12} is $p_1 \times p_2$ with $X_{21} = X'_{12}$ and $p_1 + p_2 = p$. Without any loss of generality, let us assume that $p_1 \geq p_2$. The determinant can be partitioned as follows:

$$\begin{aligned} |X|^{\alpha-\frac{p+1}{2}} &= |X_{11}|^{\alpha-\frac{p+1}{2}} |X_{22} - X_{21}X_{11}^{-1}X_{12}|^{\alpha-\frac{p+1}{2}} \\ &= |X_{11}|^{\alpha-\frac{p+1}{2}} |X_{22}|^{\alpha-\frac{p+1}{2}} |I - X_{22}^{-\frac{1}{2}}X_{21}X_{11}^{-1}X_{12}X_{22}^{-\frac{1}{2}}|^{\alpha-\frac{p+1}{2}}. \end{aligned}$$

Letting

$$Y = X_{22}^{-\frac{1}{2}}X_{21}X_{11}^{-\frac{1}{2}} \Rightarrow dY = |X_{22}|^{-\frac{p_1}{2}} |X_{11}|^{-\frac{p_2}{2}} dX_{21}$$

for fixed X_{11} and X_{22} by making use of Theorem 1.6.4 of Chap. 1 or Theorem 1.18 of Mathai (1997),

$$|X|^{\alpha-\frac{p+1}{2}} dX_{21} = |X_{11}|^{\alpha+\frac{p_2}{2}-\frac{p+1}{2}} |X_{22}|^{\alpha+\frac{p_1}{2}-\frac{p+1}{2}} |I - YY'|^{\alpha-\frac{p+1}{2}} dY.$$

Letting $S = YY'$ and integrating out over the Stiefel manifold, we have

$$dY = \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_2}(\frac{p_1}{2})} |S|^{\frac{p_1}{2}-\frac{p_2+1}{2}} dS;$$

refer to Theorem 2.16 and Remark 2.13 of Mathai (1997) or Theorem 4.2.3 of Chap. 4. Now, the integral over S gives

$$\int_{O < S < I} |S|^{\frac{p_1 - p_2 + 1}{2}} |I - S|^{\alpha - \frac{p_1 - p_2 + 1}{2}} dS = \frac{\Gamma_{p_2}(\frac{p_1}{2}) \Gamma_{p_2}(\alpha - \frac{p_1}{2})}{\Gamma_{p_2}(\alpha)},$$

for $\Re(\alpha) > \frac{p_1 - 1}{2}$. Collecting all the factors, we have

$$|X_{11}|^{\alpha - \frac{p_1 + 1}{2}} |X_{22}|^{\alpha - \frac{p_2 + 1}{2}} \pi^{\frac{p_1 p_2}{2}} \frac{\Gamma_{p_2}(\alpha - \frac{p_1}{2})}{\Gamma_{p_2}(\alpha)}.$$

One can observe from this result that the original determinant splits into functions of X_{11} and X_{22} . This also shows that if we are considering a real matrix-variate gamma density, then the diagonal blocks X_{11} and X_{22} are statistically independently distributed, where X_{11} will have a p_1 -variate gamma distribution and X_{22} , a p_2 -variate gamma distribution. Note that $\text{tr}(X) = \text{tr}(X_{11}) + \text{tr}(X_{22})$ and hence, the integral over X_{22} gives $\Gamma_{p_2}(\alpha)$ and the integral over X_{11} , $\Gamma_{p_1}(\alpha)$. Thus, the total integral is available as

$$\Gamma_{p_1}(\alpha) \Gamma_{p_2}(\alpha) \pi^{\frac{p_1 p_2}{2}} \frac{\Gamma_{p_2}(\alpha - \frac{p_1}{2})}{\Gamma_{p_2}(\alpha)} = \Gamma_p(\alpha)$$

since $\pi^{\frac{p_1 p_2}{2}} \Gamma_{p_1}(\alpha) \Gamma_{p_2}(\alpha - \frac{p_1}{2}) = \Gamma_p(\alpha)$.

Hence, it is seen that instead of integrating out variables one at a time, we could have also integrated out blocks of variables at a time and verified the result. A similar procedure works for real matrix-variate type-1 and type-2 beta distributions, as well as the matrix-variate gamma and type-1 and type-2 beta distributions in the complex domain.

5.3.4. Methods avoiding integration over the Stiefel manifold

The general method of partitioning matrices previously described involves the integration over the Stiefel manifold as an intermediate step and relies on Theorem 4.2.3. We will consider another procedure whereby integration over the Stiefel manifold is not required. Let us consider the real gamma case first. Again, we begin with the decomposition

$$|X|^{\alpha - \frac{p+1}{2}} = |X_{11}|^{\alpha - \frac{p+1}{2}} |X_{22} - X_{21} X_{11}^{-1} X_{12}|^{\alpha - \frac{p+1}{2}}. \tag{5.3.8}$$

Instead of integrating out X_{21} or X_{12} , let us integrate out X_{22} . Let X_{11} be a $p_1 \times p_1$ matrix and X_{22} be a $p_2 \times p_2$ matrix, with $p_1 + p_2 = p$. In the above partitioning, we require that X_{11} be nonsingular. However, when X is positive definite, both X_{11} and X_{22} will be positive definite, and thereby nonsingular. From the second factor in (5.3.8), $X_{22} >$

$X_{21}X_{11}^{-1}X_{12}$ as $X_{22} - X_{21}X_{11}^{-1}X_{12}$ is positive definite. We will attempt to integrate out X_{22} first. Let $U = X_{22} - X_{21}X_{11}^{-1}X_{12}$ so that $dU = dX_{22}$ for fixed X_{11} and X_{12} . Since $\text{tr}(X) = \text{tr}(X_{11}) + \text{tr}(X_{22})$, we have

$$e^{-\text{tr}(X_{22})} = e^{-\text{tr}(U) - \text{tr}(X_{21}X_{11}^{-1}X_{12})}.$$

On integrating out U , we obtain

$$\int_{U>O} |U|^{\alpha - \frac{p_2+1}{2}} e^{-\text{tr}(U)} dU = \Gamma_{p_2}\left(\alpha - \frac{p_2}{2}\right), \Re(\alpha) > \frac{p_2 - 1}{2}$$

since $\alpha - \frac{p_2+1}{2} = \alpha - \frac{p_2}{2} - \frac{p_2+1}{2}$. Letting

$$Y = X_{21}X_{11}^{-\frac{1}{2}} \Rightarrow dY = |X_{11}|^{-\frac{p_2}{2}} dX_{21}$$

for fixed X_{11} (Theorem 1.6.1), we have

$$\int_{X_{21}} e^{-\text{tr}(X_{21}X_{11}^{-1}X_{12})} dX_{21} = |X_{11}|^{\frac{p_2}{2}} \int_Y e^{-\text{tr}(YY')} dY.$$

But $\text{tr}(YY')$ is the sum of the squares of the $p_1 p_2$ elements of Y and each integral is of the form $\int_{-\infty}^{\infty} e^{-z^2} dz = \sqrt{\pi}$. Hence,

$$\int_Y e^{-\text{tr}(YY')} dY = \pi^{\frac{p_1 p_2}{2}}.$$

We may now integrate out X_{11} :

$$\begin{aligned} \int_{X_{11}>O} |X_{11}|^{\alpha + \frac{p_2}{2} - \frac{p_2+1}{2}} e^{-\text{tr}(X_{11})} dX_{11} \\ = \int_{X_{11}>O} |X_{11}|^{\alpha - \frac{p_1+1}{2}} e^{-\text{tr}(X_{11})} dX_{11} \\ = \Gamma_{p_1}(\alpha). \end{aligned}$$

Thus, we have the following factors:

$$\pi^{\frac{p_1 p_2}{2}} \Gamma_{p_2}\left(\alpha - \frac{p_2}{2}\right) \Gamma_{p_1}(\alpha) = \Gamma_p(\alpha)$$

since

$$\frac{p_1(p_1 - 1)}{4} + \frac{p_2(p_2 - 1)}{4} + \frac{p_1 p_2}{2} = \frac{p(p - 1)}{4}, \quad p = p_1 + p_2,$$

and

$$\begin{aligned}\Gamma_{p_1}(\alpha)\Gamma_{p_2}(\alpha - p_1/2) &= \Gamma(\alpha)\Gamma(\alpha - 1/2)\cdots\Gamma(\alpha - (p_1 - 1)/2)\Gamma_{p_2}(\alpha - (p_1)/2) \\ &= \Gamma(\alpha)\cdots\Gamma(\alpha - (p_1 + p_2 - 1)/2).\end{aligned}$$

Hence the result. This procedure avoids integration over the Stiefel manifold and does not require that $p_1 \geq p_2$. We could have integrated out X_{11} first, if needed. In that case, we would have used the following expansion:

$$|X|^{\alpha - \frac{p+1}{2}} = |X_{22}|^{\alpha - \frac{p+1}{2}} |X_{11} - X_{12}X_{22}^{-1}X_{21}|^{\alpha - \frac{p+1}{2}}.$$

We would have then proceeded as before by integrating out X_{11} first and would have ended up with

$$\pi^{\frac{p_1 p_2}{2}} \Gamma_{p_1}(\alpha - p_2/2)\Gamma_{p_2}(\alpha) = \Gamma_p(\alpha), \quad p = p_1 + p_2.$$

Note 5.3.1: If we are considering a real matrix-variate gamma density, such as the Wishart density, then from the above procedure, observe that after integrating out X_{22} , the only factor containing X_{21} is the exponential function, which has the structure of a matrix-variate Gaussian density. Hence, for a given X_{11} , X_{21} is matrix-variate Gaussian distributed. Similarly, for a given X_{22} , X_{12} is matrix-variate Gaussian distributed. Further, the diagonal blocks X_{11} and X_{22} are independently distributed.

The same procedure also applies for the evaluation of the gamma integrals in the complex domain. Since the steps are parallel, they will not be detailed here.

5.3.5. Arbitrary moments of the determinants, real gamma and beta matrices

Let the $p \times p$ real positive definite matrix X have a real matrix-variate gamma density with the parameters $(\alpha, B > O)$. Then for an arbitrary h , we can evaluate the h -th moment of the determinant of X with the help of the matrix-variate gamma integral, namely,

$$\int_{X>O} |X|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(BX)} dX = |B|^{-\alpha} \Gamma_p(\alpha). \quad (i)$$

By making use of (i), we can evaluate the h -th moment in a real matrix-variate gamma density with the parameters $(\alpha, B > O)$ by considering the associated normalizing constant. Let $u_1 = |X|$. Then, the moments of u_1 can be obtained by integrating out over the density of X :

$$\begin{aligned}
E[u_1]^h &= \frac{|B|^\alpha}{\Gamma_p(\alpha)} \int_{X>O} u_1^h |X|^{\alpha-\frac{p+1}{2}} e^{-\text{tr}(BX)} dX \\
&= \frac{|B|^\alpha}{\Gamma_p(\alpha)} \int_{X>O} |X|^{\alpha+h-\frac{p+1}{2}} e^{-\text{tr}(BX)} dX \\
&= \frac{|B|^\alpha}{\Gamma_p(\alpha)} \Gamma_p(\alpha+h) |B|^{-(\alpha+h)}, \Re(\alpha+h) > \frac{p-1}{2}.
\end{aligned}$$

Thus,

$$E[u_1]^h = |B|^{-h} \frac{\Gamma_p(\alpha+h)}{\Gamma_p(\alpha)}, \Re(\alpha+h) > \frac{p-1}{2}.$$

This is evaluated by observing that when $E[u_1]^h$ is taken, α is replaced by $\alpha+h$ in the integrand and hence, the answer is obtained from equation (i). The same procedure enables one to evaluate the h -th moment of the determinants of type-1 beta and type-2 beta matrices. Let Y be a $p \times p$ real positive definite matrix having a real matrix-variate type-1 beta density with the parameters (α, β) and $u_2 = |Y|$. Then, the h -th moment of Y is obtained as follows:

$$\begin{aligned}
E[u_2]^h &= \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)} \int_{O<Y<I} u_2^h |Y|^{\alpha-\frac{p+1}{2}} |I-Y|^{\beta-\frac{p+1}{2}} dY \\
&= \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)} \int_{O<Y<I} |Y|^{\alpha+h-\frac{p+1}{2}} |I-Y|^{\beta-\frac{p+1}{2}} dY \\
&= \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha)\Gamma_p(\beta)} \frac{\Gamma_p(\alpha+h)\Gamma_p(\beta)}{\Gamma_p(\alpha+\beta+h)}, \Re(\alpha+h) > \frac{p-1}{2}, \\
&= \frac{\Gamma_p(\alpha+h)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\alpha+\beta)}{\Gamma_p(\alpha+\beta+h)}, \Re(\alpha+h) > \frac{p-1}{2}.
\end{aligned}$$

In a similar manner, let $u_3 = |Z|$ where Z has a $p \times p$ real matrix-variate type-2 beta density with the parameters (α, β) . In this case, take $\alpha+\beta = (\alpha+h) + (\beta-h)$, replacing α by $\alpha+h$ and β by $\beta-h$. Then, considering the normalizing constant of a real matrix-variate type-2 beta density, we obtain the h -th moment of u_3 as follows:

$$E[u_3]^h = \frac{\Gamma_p(\alpha+h)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\beta-h)}{\Gamma_p(\beta)}, \Re(\alpha+h) > \frac{p-1}{2}, \Re(\beta-h) > \frac{p-1}{2}.$$

Relatively few moments will exist in this case, as $\Re(\alpha+h) > \frac{p-1}{2}$ implies that $\Re(h) > -\Re(\alpha) + \frac{p-1}{2}$ and $\Re(\beta-h) > \frac{p-1}{2}$ means that $\Re(h) < \Re(\beta) - \frac{p-1}{2}$. Accordingly, only moments in the range $-\Re(\alpha) + \frac{p-1}{2} < \Re(h) < \Re(\beta) - \frac{p-1}{2}$ will exist. We can summarize

the above results as follows: When X is distributed as a real $p \times p$ matrix-variate gamma with the parameters $(\alpha, B > O)$,

$$E|X|^h = \frac{|B|^{-h} \Gamma_p(\alpha + h)}{\Gamma_p(\alpha)}, \Re(\alpha) > \frac{p-1}{2}. \tag{5.3.9}$$

When Y has a $p \times p$ real matrix-variate type-1 beta density with the parameters (α, β) and if $u_2 = |Y|$ then

$$E[u_2]^h = \frac{\Gamma_p(\alpha + h)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\alpha + \beta)}{\Gamma_p(\alpha + \beta + h)}, \Re(\alpha + h) > \frac{p-1}{2}. \tag{5.3.10}$$

When the $p \times p$ real positive definite matrix Z has a real matrix-variate type-2 beta density with the parameters (α, β) , then letting $u_3 = |Z|$,

$$E[u_3]^h = \frac{\Gamma_p(\alpha + h)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\beta - h)}{\Gamma_p(\beta)}, -\Re(\alpha) + \frac{p-1}{2} < \Re(h) < \Re(\beta) - \frac{p-1}{2}. \tag{5.3.11}$$

Let us examine (5.3.9):

$$\begin{aligned} E|X|^h &= |B|^{-h} \frac{\Gamma_p(\alpha + h)}{\Gamma_p(\alpha)} \\ &= |B|^{-h} \frac{\Gamma(\alpha + h)}{\Gamma(\alpha)} \frac{\Gamma(\alpha - \frac{1}{2} + h)}{\Gamma(\alpha - \frac{1}{2})} \cdots \frac{\Gamma(\alpha - \frac{p-1}{2} + h)}{\Gamma(\alpha - \frac{p-1}{2})} \\ &= E[x_1^h] E[x_2^h] \cdots E[x_p^h] \end{aligned}$$

where x_j is a real scalar gamma random variable with shape parameter $\alpha - \frac{j-1}{2}$ and scale parameter λ_j where $\lambda_j > 0, j = 1, \dots, p$ are the eigenvalues of $B > O$ by observing that the determinant is the product and trace is the sum of the eigenvalues $\lambda_1, \dots, \lambda_p$. Further, x_1, \dots, x_p , are independently distributed. Hence, structurally, we have the following representation:

$$|X| = \prod_{j=1}^p x_j \tag{5.3.12}$$

where x_j has the density

$$f_{1j}(x_j) = \frac{\lambda_j^{\alpha - \frac{j-1}{2}}}{\Gamma(\alpha - \frac{j-1}{2})} x_j^{\alpha - \frac{j-1}{2} - 1} e^{-\lambda_j x_j}, \quad 0 \leq x_j < \infty,$$

for $\Re(\alpha) > \frac{j-1}{2}$, $\lambda_j > 0$ and zero otherwise. Similarly, when the $p \times p$ real positive definite matrix Y has a real matrix-variate type-1 beta density with the parameters (α, β) , the determinant, $|Y|$, has the structural representation

$$|Y| = \prod_{j=1}^p y_j \quad (5.3.13)$$

where y_j is a real scalar type-1 beta random variable with the parameter $(\alpha - \frac{j-1}{2}, \beta)$ for $j = 1, \dots, p$. When the $p \times p$ real positive definite matrix Z has a real matrix-variate type-2 beta density, then $|Z|$, the determinant of Z , has the following structural representation:

$$|Z| = \prod_{j=1}^p z_j \quad (5.3.14)$$

where z_j has a real scalar type-2 beta density with the parameters $(\alpha - \frac{j-1}{2}, \beta - \frac{j-1}{2})$ for $j = 1, \dots, p$.

Example 5.3.2. Consider a real 2×2 matrix X having a real matrix-variate distribution. Derive the density of the determinant $|X|$ if X has (a) a gamma distribution with the parameters $(\alpha, B = I)$; (b) a real type-1 beta distribution with the parameters $(\alpha = \frac{3}{2}, \beta = \frac{3}{2})$; (c) a real type-2 beta distribution with the parameters $(\alpha = \frac{3}{2}, \beta = \frac{3}{2})$.

Solution 5.3.2. We will derive the density in these three cases by using three different methods to illustrate the possibility of making use of various approaches for solving such problems. (a) Let $u_1 = |X|$ in the gamma case. Then for an arbitrary h ,

$$E[u_1^h] = \frac{\Gamma_p(\alpha + h)}{\Gamma_p(\alpha)} = \frac{\Gamma(\alpha + h)\Gamma(\alpha + h - \frac{1}{2})}{\Gamma(\alpha)\Gamma(\alpha - \frac{1}{2})}, \quad \Re(\alpha) > \frac{1}{2}.$$

Since the gammas differ by $\frac{1}{2}$, they can be combined by utilizing the following identity:

$$\Gamma(mz) = (2\pi)^{\frac{1-m}{2}} m^{mz - \frac{1}{2}} \Gamma(z)\Gamma(z + \frac{1}{m}) \cdots \Gamma(z + \frac{m-1}{m}), \quad m = 1, 2, \dots, \quad (5.3.15)$$

which is the multiplication formula for gamma functions. For $m = 2$, we have the duplication formula:

$$\Gamma(2z) = (2\pi)^{-\frac{1}{2}} 2^{2z - \frac{1}{2}} \Gamma(z)\Gamma(z + 1/2).$$

Thus,

$$\Gamma(z)\Gamma(z + 1/2) = \pi^{\frac{1}{2}} 2^{1-2z} \Gamma(2z).$$

Now, by taking $z = \alpha - \frac{1}{2} + h$ in the numerator and $z = \alpha - \frac{1}{2}$ in the denominator, we can write

$$E[u_1^h] = \frac{\Gamma(\alpha + h)\Gamma(\alpha + h - \frac{1}{2})}{\Gamma(\alpha)\Gamma(\alpha - \frac{1}{2})} = \frac{\Gamma(2\alpha - 1 + 2h)}{\Gamma(2\alpha - 1)} 2^{-2h}.$$

Accordingly,

$$E[(4u_1)^h] = \frac{\Gamma(2\alpha - 1 + 2h)}{\Gamma(2\alpha - 1)} \Rightarrow E[2u_1^{\frac{1}{2}}]^{2h} = \frac{\Gamma(2\alpha - 1 + 2h)}{\Gamma(2\alpha - 1)}.$$

This shows that $v = 2u_1^{\frac{1}{2}}$ has a real scalar gamma distribution with the parameters $(2\alpha - 1, 1)$ whose density is

$$\begin{aligned} f(v) dv &= \frac{v^{(2\alpha-1)-1}}{\Gamma(2\alpha - 1)} e^{-v} dv = \frac{(2u_1^{\frac{1}{2}})^{2\alpha-2}}{\Gamma(2\alpha - 1)} e^{-2u_1^{\frac{1}{2}}} d(2u_1^{\frac{1}{2}}) \\ &= \frac{2^{2\alpha-2} u_1^{\alpha-\frac{1}{2}-1}}{\Gamma(2\alpha - 1)} e^{-2u_1^{\frac{1}{2}}} du_1. \end{aligned}$$

Hence the density of u_1 , denoted by $f_1(u_1)$, is the following:

$$f_1(u_1) = \frac{2^{2\alpha-2} u_1^{\alpha-\frac{1}{2}-1}}{\Gamma(2\alpha - 1)} e^{-2u_1^{\frac{1}{2}}}, \quad 0 \leq u_1 < \infty$$

and zero elsewhere. It can easily be verified that $f_1(u_1)$ is a density.

(b) Let $u_2 = |X|$. Then for an arbitrary h , $\alpha = \frac{3}{2}$ and $\beta = \frac{3}{2}$,

$$\begin{aligned} E[u_2^h] &= \frac{\Gamma_p(\alpha + h)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\alpha + \beta)}{\Gamma_p(\alpha + \beta + h)} \\ &= \frac{\Gamma(3)\Gamma(\frac{5}{2})}{\Gamma(\frac{3}{2})\Gamma(\frac{7}{2})} \frac{\Gamma(\frac{3}{2} + h)\Gamma(1 + h)}{\Gamma(3 + h)\Gamma(\frac{5}{2} + h)} \\ &= 3 \left\{ \frac{1}{(2 + h)(1 + h)(\frac{3}{2} + h)} \right\} = 3 \left\{ \frac{2}{2 + h} + \frac{2}{1 + h} - \frac{4}{\frac{3}{2} + h} \right\}, \end{aligned}$$

the last expression resulting from an application of the partial fraction technique. This results from h -th moment of the distribution of u_2 , whose density which is

$$f_2(u_2) = 6\{1 + u_2 - 2u_2^{\frac{1}{2}}\}, \quad 0 \leq u_2 \leq 1,$$

and zero elsewhere, is readily seen to be *bona fide*.

(c) Let the density $u_3 = |X|$ be denoted by $f_3(u_3)$. The Mellin transform of $f_3(u_3)$, with Mellin parameter s , is

$$\begin{aligned} E[u_3^{s-1}] &= \frac{\Gamma_p(\alpha + s - 1)}{\Gamma_p(\alpha)} \frac{\Gamma_p(\beta - s + 1)}{\Gamma_p(\beta)} = \frac{\Gamma_2(\frac{3}{2} + s - 1)}{\Gamma_2(\frac{3}{2})} \frac{\Gamma_2(\frac{3}{2} - s + 1)}{\Gamma_2(\frac{3}{2})} \\ &= \frac{1}{[\Gamma(3/2)\Gamma(1)]^2} \Gamma(1/2 + s) \Gamma(s) \Gamma(5/2 - s) \Gamma(2 - s), \end{aligned}$$

the corresponding density being available by taking the inverse Mellin transform, namely,

$$f_3(u_3) = \frac{4}{\pi} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Gamma(s) \Gamma(s + 1/2) \Gamma(5/2 - s) \Gamma(2 - s) u_3^{-s} ds \quad (i)$$

where $i = \sqrt{-1}$ and c in the integration contour is such that $0 < c < 2$. The integral in (i) is available as the sum of residues at the poles of $\Gamma(s)\Gamma(s + \frac{1}{2})$ for $0 \leq u_3 \leq 1$ and the sum of residues at the poles of $\Gamma(2 - s)\Gamma(\frac{5}{2} - s)$ for $1 < u_3 < \infty$. We can also combine $\Gamma(s)$ and $\Gamma(s + \frac{1}{2})$ as well as $\Gamma(2 - s)$ and $\Gamma(\frac{5}{2} - s)$ by making use of the duplication formula for gamma functions. We will then be able to identify the functions in each of the sectors, $0 \leq u_3 \leq 1$ and $1 < u_3 < \infty$. These will be functions of $u_3^{\frac{1}{2}}$ as done in the case (a). In order to illustrate the method relying on the inverse Mellin transform, we will evaluate the density $f_3(u_3)$ as a sum of residues. The poles of $\Gamma(s)\Gamma(s + \frac{1}{2})$ are simple and hence two sums of residues are obtained for $0 \leq u_3 \leq 1$. The poles of $\Gamma(s)$ occur at $s = -\nu$, $\nu = 0, 1, \dots$, and those of $\Gamma(s + \frac{1}{2})$ occur at $s = -\frac{1}{2} - \nu$, $\nu = 0, 1, \dots$. The residues and the sum thereof will be evaluated with the help of the following two lemmas.

Lemma 5.3.1. Consider a function $\Gamma(\gamma + s)\phi(s)u^{-s}$ whose poles are simple. The residue at the pole $s = -\gamma - \nu$, $\nu = 0, 1, \dots$, denoted by R_ν , is given by

$$R_\nu = \frac{(-1)^\nu}{\nu!} \phi(-\gamma - \nu) u^{\gamma+\nu}.$$

Lemma 5.3.2. When $\Gamma(\delta)$ and $\Gamma(\delta - \nu)$ are defined

$$\Gamma(\delta - \nu) = \frac{(-1)^\nu \Gamma(\delta)}{(-\delta + 1)_\nu}$$

where, for example, $(a)_\nu = a(a+1)\cdots(a+\nu-1)$, $(a)_0 = 1$, $a \neq 0$, is the Pochhammer symbol.

Observe that $\Gamma(\alpha)$ is defined for all $\alpha \neq 0, -1, -2, \dots$, and that an integral representation requires $\Re(\alpha) > 0$. As well, $\Gamma(\alpha + k) = \Gamma(\alpha)(\alpha)_k$, $k = 1, 2, \dots$. With the help of Lemmas 5.3.1 and 5.3.2, the sum of the residues at the poles of $\Gamma(s)$ in the integral in (i), excluding the constant $\frac{4}{\pi}$, is the following:

$$\begin{aligned} & \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{\nu!} \Gamma\left(\frac{1}{2} - \nu\right) \Gamma\left(\frac{5}{2} + \nu\right) \Gamma(2 + \nu) \\ &= \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{\nu!} \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{5}{2}\right) \Gamma(2) \frac{(-1)^\nu}{\left(\frac{1}{2}\right)_\nu} \left(\frac{5}{2}\right)_\nu (2)_\nu u_3^\nu \\ &= \frac{3}{4} \pi {}_2F_1\left(\frac{5}{2}, 2; \frac{1}{2}; u_3\right), \quad 0 \leq u_3 \leq 1, \end{aligned}$$

where the ${}_2F_1(\cdot)$ is Gauss' hypergeometric function. The same procedure consisting of taking the sum of the residues at the poles $s = -\frac{1}{2} - \nu$, $\nu = 0, 1, \dots$, gives

$$-3\pi u_3^{\frac{1}{2}} {}_2F_1\left(3, \frac{5}{2}; \frac{3}{2}; u_3\right), \quad 0 \leq u_3 \leq 1.$$

The inverse Mellin transform for the sector $1 < u_3 < \infty$ is available as the sum of residues at the poles of $\Gamma(\frac{5}{2} - s)$ and $\Gamma(2 - s)$ which occur at $s = \frac{5}{2} + \nu$ and $s = 2 + \nu$ for $\nu = 0, 1, \dots$. The sum of residues at the poles of $\Gamma(\frac{5}{2} - s)$ is the following:

$$\begin{aligned} & \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{\nu!} \Gamma\left(\frac{5}{2} + \nu\right) \Gamma(3 + \nu) \Gamma\left(-\frac{1}{2} - \nu\right) u_3^{-\frac{5}{2}-\nu} \\ &= -3\pi u_3^{-\frac{5}{2}} {}_2F_1\left(\frac{5}{2}, 3; \frac{3}{2}; \frac{1}{u_3}\right), \quad 1 < u_3 < \infty, \end{aligned}$$

and the sum of the residues at the poles of $\Gamma(2 - s)$ is given by

$$\frac{3}{4} \pi u_3^{-2} {}_2F_1\left(2, \frac{5}{2}; \frac{1}{2}; \frac{1}{u_3}\right), \quad 1 < u_3 < \infty.$$

Now, on combining all the hypergeometric series and multiplying the result by the constant $\frac{4}{\pi}$, the final representation of the required density is obtained as

$$f_3(u_3) = \begin{cases} 3 {}_2F_1\left(\frac{5}{2}, 2; \frac{1}{2}; u_3\right) - 12u_3^{\frac{1}{2}} {}_2F_1\left(3, \frac{5}{2}; \frac{3}{2}; u_3\right), & 0 \leq u_3 \leq 1, \\ 3u_3^{-2} {}_2F_1\left(2, \frac{5}{2}; \frac{1}{2}; \frac{1}{u_3}\right) - 12u_3^{-\frac{5}{2}} {}_2F_1\left(\frac{5}{2}, 3; \frac{3}{2}; \frac{1}{u_3}\right), & 1 < u_3 < \infty. \end{cases}$$

This completes the computations.

5.3a.2. Arbitrary moments of the determinants in the complex case

In the complex matrix-variate case, one can consider the absolute value of the determinant, which will be real; however, the parameters will be different from those in the real case. For example, consider the complex matrix-variate gamma density. If \tilde{X} has a $p \times p$ complex matrix-variate gamma density with the parameters $(\alpha, \tilde{B} > O)$, then the h -th moment of the absolute value of the determinant of \tilde{X} is the following:

$$\begin{aligned} E[|\det(\tilde{X})|^h] &= \frac{|\det(\tilde{B})|^{-h} \tilde{\Gamma}_p(\alpha + h)}{\tilde{\Gamma}_p(\alpha)} \\ &= (\lambda_1 \cdots \lambda_p)^{-h} \prod_{j=1}^p \frac{\Gamma(\alpha - (j-1) + h)}{\Gamma(\alpha - (j-1))} = \prod_{j=1}^p E[\tilde{x}_j]^h, \end{aligned}$$

that is, $|\det(\tilde{X})|$ has the structural representation

$$|\det(\tilde{X})| = \tilde{x}_1 \tilde{x}_2 \cdots \tilde{x}_p, \quad (5.3a.8)$$

where the \tilde{x}_j is a *real scalar gamma random variable* with the parameters $(\alpha - (j-1), \lambda_j)$, $j = 1, \dots, p$, and the \tilde{x}_j 's are independently distributed. Similarly, when \tilde{Y} is a $p \times p$ complex Hermitian positive definite matrix having a complex matrix-variate type-1 beta density with the parameters (α, β) , the absolute value of the determinant of \tilde{Y} , $|\det(\tilde{Y})|$, has the structural representation

$$|\det(\tilde{Y})| = \prod_{j=1}^p \tilde{y}_j \quad (5.3a.9)$$

where the \tilde{y}_j 's are independently distributed, \tilde{y}_j being a *real scalar type-1 beta random variable* with the parameters $(\alpha - (j-1), \beta)$, $j = 1, \dots, p$. When \tilde{Z} is a $p \times p$ Hermitian positive definite matrix having a complex matrix-variate type-2 beta density with the parameters (α, β) , then for arbitrary h , the h -th moment of the absolute value of the determinant is given by

$$\begin{aligned} E[|\det(\tilde{Z})|^h] &= \frac{\tilde{\Gamma}_p(\alpha + h) \tilde{\Gamma}_p(\beta - h)}{\tilde{\Gamma}_p(\alpha) \tilde{\Gamma}_p(\beta)} \\ &= \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha - (j-1) + h)}{\Gamma(\alpha - (j-1))} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(\beta - (j-1) - h)}{\Gamma(\beta - (j-1))} \right\} \\ &= \prod_{j=1}^p E[\tilde{z}_j]^h, \end{aligned}$$

so that the absolute value of the determinant of \tilde{Z} has the following structural representation:

$$|\det(\tilde{Z})| = \prod_{j=1}^p \tilde{z}_j \tag{5.3a.10}$$

where the \tilde{z}_j 's are independently distributed *real scalar type-2 beta random variables* with the parameters $(\alpha - (j - 1), \beta - (j - 1))$ for $j = 1, \dots, p$. Thus, in the real case, the determinant and, in the complex case, the absolute value of the determinant have structural representations in terms of products of independently distributed real scalar random variables. The following is the summary of what has been discussed so far:

Distribution	Parameters, real case	Parameters, complex case
gamma	$(\alpha - \frac{j-1}{2}, \lambda_j)$	$(\alpha - (j - 1), \lambda_j)$
type-1 beta	$(\alpha - \frac{j-1}{2}, \beta)$	$(\alpha - (j - 1), \beta)$
type-2 beta	$(\alpha - \frac{j-1}{2}, \beta - \frac{j-1}{2})$	$(\alpha - (j - 1), \beta - (j - 1))$

for $j = 1, \dots, p$. When we consider the determinant in the real case, the parameters differ by $\frac{1}{2}$ whereas the parameters differ by 1 in the complex domain. Whether in the real or complex cases, the individual variables appearing in the structural representations are real scalar variables that are independently distributed.

Example 5.3a.2. Even when $p = 2$, some of the poles will be of order 2 since the gammas differ by integers in the complex case, and hence a numerical example will not be provided for such an instance. Actually, when poles of order 2 or more are present, the series representation will contain logarithms as well as psi and zeta functions. A simple illustrative example is now considered. Let \tilde{X} be 2×2 matrix having a complex matrix-variate type-1 beta distribution with the parameters $(\alpha = 2, \beta = 2)$. Evaluate the density of $\tilde{u} = |\det(\tilde{X})|$.

Solution 5.3a.2. Let us take the $(s - 1)$ th moment of \tilde{u} which corresponds to the Mellin transform of the density of \tilde{u} , with Mellin parameter s :

$$\begin{aligned} E[\tilde{u}^{s-1}] &= \frac{\tilde{\Gamma}_p(\alpha + s - 1)}{\tilde{\Gamma}_p(\alpha)} \frac{\tilde{\Gamma}_p(\alpha + \beta)}{\tilde{\Gamma}_p(\alpha + \beta + s - 1)} \\ &= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + \beta - 1)}{\Gamma(\alpha)\Gamma(\alpha - 1)} \frac{\Gamma(\alpha + s - 1)\Gamma(\alpha + s - 2)}{\Gamma(\alpha + \beta + s - 1)\Gamma(\alpha + \beta + s - 2)} \\ &= \frac{\Gamma(4)\Gamma(3)}{\Gamma(2)\Gamma(1)} \frac{\Gamma(1 + s)\Gamma(s)}{\Gamma(3 + s)\Gamma(2 + s)} = \frac{12}{(2 + s)(1 + s)^2 s}. \end{aligned}$$

The inverse Mellin transform then yields the density of \tilde{u} , denoted by $\tilde{g}(\tilde{u})$, which is

$$\tilde{g}(\tilde{u}) = 12 \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{(2+s)(1+s)^2 s} u^{-s} ds \quad (i)$$

where the c in the contour is any real number $c > 0$. There is a pole of order 1 at $s = 0$ and another pole of order 1 at $s = -2$, the residues at these poles being obtained as follows:

$$\lim_{s \rightarrow 0} \frac{u^{-s}}{(2+s)(1+s)^2} = \frac{1}{2}, \quad \lim_{s \rightarrow -2} \frac{u^{-s}}{(1+s)^2 s} = -\frac{u^2}{2}.$$

The pole at $s = -1$ is of order 2 and hence the residue is given by

$$\begin{aligned} \lim_{s \rightarrow -1} \left\{ \frac{d}{ds} \frac{u^{-s}}{s(2+s)} \right\} &= \lim_{s \rightarrow -1} \left\{ \frac{(-\ln u)u^{-s}}{s(2+s)} - \frac{u^{-s}}{s^2(2+s)} - \frac{u^{-s}}{s(2+s)^2} \right\} \\ &= u \ln u - u + u = u \ln u. \end{aligned}$$

Hence the density is the following:

$$\tilde{g}(\tilde{u}) = 6 - 6u^2 + 12u \ln u, \quad 0 \leq u \leq 1,$$

and zero elsewhere, where u is real. It can readily be shown that $\tilde{g}(\tilde{u}) \geq 0$ and $\int_0^1 \tilde{g}(\tilde{u}) du = 1$. This completes the computations.

Exercises 5.3

5.3.1. Evaluate the real $p \times p$ matrix-variate type-2 beta integral from first principles or by direct evaluation by partitioning the matrix as in Sect. 5.3.3 (general partitioning).

5.3.2. Repeat Exercise 5.3.1 for the complex case.

5.3.3. In the 2×2 partitioning of a $p \times p$ real matrix-variate gamma density with shape parameter α and scale parameter I , where the first diagonal block X_{11} is $r \times r$, $r < p$, compute the density of the rectangular block X_{12} .

5.3.4. Repeat Exercise 5.3.3 for the complex case.

5.3.5. Let the $p \times p$ real matrices X_1 and X_2 have real matrix-variate gamma densities with the parameters $(\alpha_1, B > O)$ and $(\alpha_2, B > O)$, respectively, B being the same for both distributions. Compute the density of (1): $U_1 = X_2^{-\frac{1}{2}} X_1 X_2^{-\frac{1}{2}}$, (2): $U_2 = X_1^{\frac{1}{2}} X_2^{-1} X_1^{\frac{1}{2}}$, (3): $U_3 = (X_1 + X_2)^{-\frac{1}{2}} X_2 (X_1 + X_2)^{-\frac{1}{2}}$, when X_1 and X_2 are independently distributed.

5.3.6. Repeat Exercise 5.3.5 for the complex case.

5.3.7. In the transformation $Y = I - X$ that was used in Sect. 5.3.1, the Jacobian is $dY = (-1)^{\frac{p(p+1)}{2}} dX$. What happened to the factor $(-1)^{\frac{p(p+1)}{2}}$?

5.3.8. Consider X in the (a) 2×2 , (b) 3×3 real matrix-variate case. If X is real matrix-variate gamma distributed, then derive the densities of the determinant of X in (a) and (b) if the parameters are $\alpha = \frac{5}{2}$, $B = I$. Consider \tilde{X} in the (a) 2×2 , (b) 3×3 complex matrix-variate case. Derive the distributions of $|\det(\tilde{X})|$ in (a) and (b) if \tilde{X} is complex matrix-variate gamma distributed with parameters ($\alpha = 2 + i$, $B = I$).

5.3.9. Consider the real cases (a) and (b) in Exercise 5.3.8 except that the distribution is type-1 beta with the parameters ($\alpha = \frac{5}{2}$, $\beta = \frac{5}{2}$). Derive the density of the determinant of X .

5.3.10. Consider \tilde{X} , (a) 2×2 , (b) 3×3 complex matrix-variate type-1 beta distributed with parameters $\alpha = \frac{5}{2} + i$, $\beta = \frac{5}{2} - i$. Then derive the density of $|\det(\tilde{X})|$ in the cases (a) and (b).

5.3.11. Consider X , (a) 2×2 , (b) 3×3 real matrix-variate type-2 beta distributed with the parameters ($\alpha = \frac{3}{2}$, $\beta = \frac{3}{2}$). Derive the density of $|X|$ in the cases (a) and (b).

5.3.12. Consider \tilde{X} , (a) 2×2 , (b) 3×3 complex matrix-variate type-2 beta distributed with the parameters ($\alpha = \frac{3}{2}$, $\beta = \frac{3}{2}$). Derive the density of $|\det(\tilde{X})|$ in the cases (a) and (b).

5.4. The Densities of Some General Structures

Three cases were examined in Section 5.3: the product of real scalar gamma variables, the product of real scalar type-1 beta variables and the product of real scalar type-2 beta variables, where in all these instances, the individual variables were mutually independently distributed. Let us now consider the corresponding general structures. Let x_j be a real scalar gamma variable with shape parameter α_j and scale parameter 1 for convenience and let the x_j 's be independently distributed for $j = 1, \dots, p$. Then, letting $v_1 = x_1 \cdots x_p$,

$$E[v_1^h] = \prod_{j=1}^p \frac{\Gamma(\alpha_j + h)}{\Gamma(\alpha_j)}, \quad \Re(\alpha_j + h) > 0, \quad \Re(\alpha_j) > 0. \quad (5.4.1)$$

Now, let y_1, \dots, y_p be independently distributed real scalar type-1 beta random variables with the parameters (α_j, β_j) , $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $j = 1, \dots, p$, and $v_2 = y_1 \cdots y_p$,

$$E[v_2^h] = \prod_{j=1}^p \frac{\Gamma(\alpha_j + h)}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j + \beta_j + h)} \quad (5.4.2)$$

for $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $\Re(\alpha_j + h) > 0$, $j = 1, \dots, p$. Similarly, let z_1, \dots, z_p , be independently distributed real scalar type-2 beta random variables with the parameters (α_j, β_j) , $j = 1, \dots, p$, and let $v_3 = z_1 \cdots z_p$. Then, we have

$$E[v_3^h] = \prod_{j=1}^p \frac{\Gamma(\alpha_j + h)}{\Gamma(\alpha_j)} \frac{\Gamma(\beta_j - h)}{\Gamma(\beta_j)} \quad (5.4.3)$$

for $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $\Re(\alpha_j + h) > 0$, $\Re(\beta_j - h) > 0$, $j = 1, \dots, p$. The corresponding densities of v_1, v_2, v_3 , respectively denoted by $g_1(v_1)$, $g_2(v_2)$, $g_3(v_3)$, are available from the inverse Mellin transforms by taking (5.4.1) to (5.4.3) as the Mellin transforms of g_1, g_2, g_3 with $h = s - 1$ for a complex variable s where s is the Mellin parameter. Then, for suitable contours L , the densities can be determined as follows:

$$\begin{aligned} g_1(v_1) &= \frac{1}{2\pi i} \int_L E[v_1^{s-1}] v_1^{-s} ds, \quad i = \sqrt{-1}, \\ &= \left\{ \prod_{j=1}^p \frac{1}{\Gamma(\alpha_j)} \right\} \frac{1}{2\pi i} \int_L \left\{ \prod_{j=1}^p \Gamma(\alpha_j + s - 1) \right\} v_1^{-s} ds \\ &= \left\{ \prod_{j=1}^p \frac{1}{\Gamma(\alpha_j)} \right\} G_{0,p}^{p,0}[v_1 |_{\alpha_j-1, j=1, \dots, p}], \quad 0 \leq v_1 < \infty, \end{aligned} \quad (5.4.4)$$

where $\Re(\alpha_j + s - 1) > 0$, $j = 1, \dots, p$, and $g_1(v_1) = 0$ elsewhere. This last representation is expressed in terms of a G-function, which will be defined in Sect. 5.4.1.

$$\begin{aligned} g_2(v_2) &= \frac{1}{2\pi i} \int_L E[v_2^{s-1}] v_2^{-s} ds, \quad i = \sqrt{-1}, \\ &= \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)} \right\} \frac{1}{2\pi i} \int_L \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha_j + s - 1)}{\Gamma(\alpha_j + \beta_j + s - 1)} \right\} v_2^{-s} ds \\ &= \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)} \right\} G_{p,p}^{p,0} \left[v_2 \middle|_{\alpha_j-1, j=1, \dots, p}^{\alpha_j+\beta_j-1, j=1, \dots, p} \right], \quad 0 \leq v_2 \leq 1, \end{aligned} \quad (5.4.5)$$

where $G_{p,p}^{p,0}$ is a G-function, $\Re(\alpha_j + s - 1) > 0$, $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $j = 1, \dots, p$, and $g_2(v_2) = 0$ elsewhere.

$$\begin{aligned}
 g_3(v_3) &= \frac{1}{2\pi i} \int_L E[v_3^{s-1}] v_3^{-s} ds, \quad i = \sqrt{-1}, \\
 &= \left\{ \prod_{j=1}^p \frac{1}{\Gamma(\alpha_j)\Gamma(\beta_j)} \right\} \int_L \left\{ \prod_{j=1}^p \Gamma(\alpha_j + s - 1)\Gamma(\beta_j - s + 1) \right\} v_3^{-s} ds \\
 &= \left\{ \prod_{j=1}^p \frac{1}{\Gamma(\alpha_j)\Gamma(\beta_j)} \right\} G_{p,p}^{p,p} \left[v_3 \middle|_{\alpha_j-1, j=1,\dots,p}^{-\beta_j, j=1,\dots,p} \right], \quad 0 \leq v_3 < \infty, \quad (5.4.6)
 \end{aligned}$$

where $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $\Re(\alpha_j + s - 1) > 0$, $\Re(\beta_j - s + 1) > 0$, $j = 1, \dots, p$, and $g_3(v_3) = 0$ elsewhere.

5.4.1. The G-function

The G-function is defined in terms of the following Mellin-Barnes integral:

$$\begin{aligned}
 G(z) &= G_{p,q}^{m,n}(z) = G_{p,q}^{m,n} \left[z \middle|_{b_1, \dots, b_q}^{a_1, \dots, a_p} \right] \\
 &= \frac{1}{2\pi i} \int_L \phi(s) z^{-s} ds, \quad i = \sqrt{-1} \\
 \phi(s) &= \frac{\{\prod_{j=1}^m \Gamma(b_j + s)\} \{\prod_{j=1}^n \Gamma(1 - a_j - s)\}}{\{\prod_{j=m+1}^q \Gamma(1 - b_j - s)\} \{\prod_{j=n+1}^p \Gamma(a_j + s)\}}
 \end{aligned}$$

where the parameters a_j , $j = 1, \dots, p$, b_j , $j = 1, \dots, q$, can be complex numbers. There are three general contours L , say L_1, L_2, L_3 where L_1 is a loop starting and ending at $-\infty$ that contains all the poles of $\Gamma(b_j + s)$, $j = 1, \dots, m$, and none of those of $\Gamma(1 - a_j - s)$, $j = 1, \dots, n$. In general L will separate the poles of $\Gamma(b_j + s)$, $j = 1, \dots, m$, from those of $\Gamma(1 - a_j - s)$, $j = 1, \dots, n$, which lie on either side of the contour. L_2 is a loop starting and ending at $+\infty$, which encloses all the poles of $\Gamma(1 - a_j - s)$, $j = 1, \dots, n$. L_3 is the straight line contour $c - i\infty$ to $c + i\infty$. The existence of the contours, convergence conditions, explicit series forms for general parameters as well as applications are available in Mathai (1993). G-functions can readily be evaluated with symbolic computing packages such as MAPLE and *Mathematica*.

Example 5.4.1. Let x_1, x_2, x_3 be independently distributed real scalar random variables, x_1 being real gamma distributed with the parameters $(\alpha_1 = 3, \beta_1 = 2)$, x_2 , real type-1 beta distributed with the parameters $(\alpha_2 = \frac{3}{2} + 2i, \beta_2 = \frac{1}{2})$ and x_3 , real type-2 beta distributed

with the parameters $(\alpha_3 = \frac{5}{2} + i, \beta_3 = 2 - i)$. Let $u_1 = x_1 x_2 x_3$, $u_2 = \frac{x_1}{x_2 x_3}$ and $u_3 = \frac{x_2}{x_1 x_3}$ with densities $g_j(u_j)$, $j = 1, 2, 3$, respectively. Derive the densities $g_j(u_j)$, $j = 1, 2, 3$, and represent them in terms of G-functions.

Solution 5.4.1. Observe that $E[\frac{1}{x_j}]^{s-1} = E[x_j^{-s+1}]$, $j = 1, 2, 3$, and that $g_1(u_1)$, $g_2(u_2)$ and $g_3(u_3)$ will share the same ‘normalizing constant’, say c , which is the product of the parts of the normalizing constants in the densities of x_1 , x_2 and x_3 that do not cancel out when determining the moments, respectively denoted by c_1 , c_2 and c_3 , that is, $c = c_1 c_2 c_3$. Thus,

$$\begin{aligned} c &= \frac{1}{\Gamma(\alpha_1)} \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)} \frac{1}{\Gamma(\alpha_3)\Gamma(\beta_3)} \\ &= \frac{1}{\Gamma(3)} \frac{\Gamma(2 + 2i)}{\Gamma(\frac{3}{2} + 2i)} \frac{1}{\Gamma(\frac{5}{2} + i)\Gamma(2 - i)}. \end{aligned} \quad (i)$$

The following are $E[x_j^{s-1}]$ and $E[x_j^{-s+1}]$ for $j = 1, 2, 3$:

$$E[x_1^{s-1}] = c_1 2^{s-1} \Gamma(2 + s), \quad E[x_1^{-s+1}] = c_1 2^{-s+1} \Gamma(4 - s) \quad (ii)$$

$$E[x_2^{s-1}] = c_2 \frac{\Gamma(\frac{1}{2} + 2i + s)}{\Gamma(1 + 2i + s)}, \quad E[x_2^{-s+1}] = c_2 \frac{\Gamma(\frac{5}{2} + 2i - s)}{\Gamma(3 + 2i - s)} \quad (iii)$$

$$E[x_3^{s-1}] = c_3 \Gamma(3/2 + i + s)\Gamma(3 - i - s), \quad E[x_3^{-s+1}] = c_3 \Gamma(7/2 + i - s)\Gamma(1 - i + s). \quad (iv)$$

Then from (i)-(iv),

$$E[u_1^{s-1}] = c 2^{s-1} \Gamma(2 + s) \frac{\Gamma(\frac{1}{2} + 2i + s)}{\Gamma(1 + 2i + s)} \Gamma(3/2 + i + s)\Gamma(3 - i - s).$$

Taking the inverse Mellin transform and writing the density $g_1(u_1)$ in terms of a G-function, we have

$$g_1(u_1) = \frac{c}{2} G_{2,3}^{3,1} \left[\frac{u_1}{2} \middle| \begin{matrix} -2+i, 1+2i \\ 2, \frac{1}{2}+2i, \frac{3}{2}+i \end{matrix} \right].$$

Using (i)-(iv) and rearranging the gamma functions so that those involving $+s$ appear together in the numerator, we have the following:

$$E[u_2^{s-1}] = \frac{c}{2} 2^s \Gamma(2 + s) \frac{\Gamma(1 - i + s)}{\Gamma(3 + 2i - s)} \Gamma(5/2 + 2i - s)\Gamma(7/2 + i - s).$$

Taking the inverse Mellin transform and expressing the result in terms of a G-function, we obtain the density $g_2(u_2)$ as

$$g_2(u_2) = \frac{c}{2} G_{2,3}^{2,2} \left[\frac{u_2}{2} \middle| \begin{matrix} -\frac{3}{2}-2i, -\frac{5}{2}-i \\ 2, 1-i, \frac{5}{2}+2i, -2-2i \end{matrix} \right].$$

Using (i)-(iv) and conveniently rearranging the gamma functions involving $+s$, we have

$$E[u_3^{s-1}] = 2c 2^{-s} \Gamma(1/2 + 2i + s) \Gamma(1 - i + s) \Gamma(4 - s) \frac{\Gamma(\frac{7}{2} + i - s)}{\Gamma(1 + 2i + s)}.$$

On taking the inverse Mellin transform, the following density is obtained:

$$g_3(u_3) = 2c G_{3,2}^{2,2} \left[2u_3 \left| \begin{matrix} -3, -\frac{5}{2}-i, 1+2i \\ \frac{1}{2}+2i, 1-i \end{matrix} \right. \right].$$

This completes the computations.

5.4.2. Some special cases of the G-function

Certain special cases of the G-function can be written in terms of elementary functions. Here are some of them:

$$\begin{aligned} G_{0,1}^{1,0}(z|a) &= z^a e^{-z}, \quad z \neq 0 \\ G_{1,1}^{1,1}[-z|_0^{1-a}] &= \Gamma(a)(1-z)^{-a}, \quad |z| < 1 \\ G_{1,1}^{1,0}[z|_\alpha^{\alpha+\beta+1}] &= \frac{1}{\Gamma(\beta+1)} z^\alpha (1-z)^\beta, \quad |z| < 1 \\ G_{1,1}^{1,1}[az^\alpha|_{\beta/\alpha}^{\beta/\alpha}] &= a^{\beta/\alpha} \left[\frac{z^\beta}{1+az^\alpha} \right], \quad |az^\alpha| < 1 \\ G_{1,1}^{1,1}[az^\alpha|_{\beta/\alpha}^{1-\gamma+\beta/\alpha}] &= \Gamma(\gamma) a^{\beta/\alpha} \left[\frac{z^\beta}{(1+az^\alpha)^\gamma} \right], \quad |az^\alpha| < 1 \\ G_{2,2}^{1,2}[-z^2|_{0, \frac{1}{2}}^{1-a, \frac{1}{2}-a}] &= \frac{\Gamma(2a)}{2^{2a}} [(1+z)^{-2a} + (1-z)^{-2a}], \quad |z| < 1 \\ G_{2,2}^{1,2}[z|_{0, -2a}^{\frac{1}{2}-a, 1-a}] &= \frac{\pi^{\frac{1}{2}}}{a} [1 + (1+z)^{\frac{1}{2}}]^{-2a}, \quad |z| < 1 \\ G_{0,2}^{1,0}\left[\frac{z^2}{4} \left| \begin{matrix} \frac{1}{4}, -\frac{1}{4} \\ -\frac{1}{4}, \frac{1}{4} \end{matrix} \right. \right] &= \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \sin z \\ G_{0,2}^{1,0}\left[\frac{z^2}{4} \left| \begin{matrix} -\frac{1}{4}, \frac{1}{4} \\ \frac{1}{4}, -\frac{1}{4} \end{matrix} \right. \right] &= \left(\frac{2}{\pi z}\right)^{\frac{1}{2}} \cos z \\ G_{0,2}^{1,0}\left[-\frac{z^2}{4} \left| \begin{matrix} 0, -\frac{1}{2} \\ \frac{1}{2}, 0 \end{matrix} \right. \right] &= \frac{2}{z\pi^{\frac{1}{2}}} \sinh z \\ G_{0,2}^{1,0}\left[-\frac{z^2}{4} \left| \begin{matrix} 0, \frac{1}{2} \\ \frac{1}{2}, 0 \end{matrix} \right. \right] &= \pi^{-\frac{1}{2}} \cosh z \\ G_{2,2}^{1,2}[\pm z|_{1,0}^{1,1}] &= \ln(1 \pm z), \quad |z| < 1 \end{aligned}$$

$$G_{p,q+1}^{1,p} \left[z \middle| \begin{matrix} 1-a_1, \dots, 1-a_p \\ 0, 1-b_1, \dots, 1-b_q \end{matrix} \right] \\ = \left[\frac{\Gamma(a_1) \cdots \Gamma(a_p)}{\Gamma(b_1) \cdots \Gamma(b_q)} \right] {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; -z)$$

for $p \leq q$ or $p = q + 1$ and $|z| < 1$.

5.4.3. The H-function

If we have a general structure corresponding to v_1, v_2 and v_3 of Sect. 5.4, say w_1, w_2 and w_3 of the form

$$w_1 = x_1^{\delta_1} x_2^{\delta_2} \cdots x_p^{\delta_p} \quad (5.4.7)$$

$$w_2 = y_1^{\delta_1} y_2^{\delta_2} \cdots y_p^{\delta_p} \quad (5.4.8)$$

$$w_3 = z_1^{\delta_1} z_2^{\delta_2} \cdots z_p^{\delta_p} \quad (5.4.9)$$

for some $\delta_j > 0$, $j = 1, \dots, p$ the densities of w_1, w_2 and w_3 are then available in terms of a more general function known as the H-function. It is again a Mellin-Barnes type integral defined and denoted as follows:

$$H(z) = H_{p,q}^{m,n}(z) = H_{p,q}^{m,n} \left[z \middle| \begin{matrix} (a_1, \alpha_1), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right] \\ = \frac{1}{2\pi i} \int_L \psi(s) z^{-s} ds, \quad i = \sqrt{-1}, \\ \psi(s) = \frac{\{\prod_{j=1}^m \Gamma(b_j + \beta_j s)\} \{\prod_{j=1}^n \Gamma(1 - a_j - \alpha_j s)\}}{\{\prod_{j=m+1}^q \Gamma(1 - b_j - \beta_j s)\} \{\prod_{j=n+1}^p \Gamma(a_j + \alpha_j s)\}} \quad (5.4.10)$$

where $\alpha_j > 0$, $j = 1, \dots, p$, $\beta_j > 0$, $j = 1, \dots, q$, are real and positive, a_j , $j = 1, \dots, p$, and b_j , $j = 1, \dots, q$, are complex numbers. Three main contours L_1, L_2, L_3 are utilized, similarly to those described in connection with the G-function. Existence conditions, properties and applications of this generalized hypergeometric function are available from Mathai et al. (2010) among other monographs. Numerous special cases can be expressed in terms of known elementary functions.

Example 5.4.2. Let x_1 and x_2 be independently distributed real type-1 beta random variables with the parameters $(\alpha_j > 0, \beta_j > 0)$, $j = 1, 2$, respectively. Let $y_1 = x_1^{\delta_1}$, $\delta_1 > 0$, and $y_2 = x_2^{\delta_2}$, $\delta_2 > 0$. Compute the density of $u = y_1 y_2$.

Solution 5.4.2. Arbitrary moments of y_1 and y_2 are available from those of x_1 and x_2 .

$$\begin{aligned}
 E[x_j^h] &= \frac{\Gamma(\alpha_j + h)}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j + \beta_j + h)}, \Re(\alpha_j + h) > 0, j = 1, 2, \\
 E[y_j^h] &= E[x_j^{\delta_j h}] = \frac{\Gamma(\alpha_j + \delta_j h)}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j + \beta_j + \delta_j h)}, \Re(\alpha_j + \delta_j h) > 0, \\
 E[u^{s-1}] &= E[y_1^{s-1}]E[y_2^{s-1}] \\
 &= \prod_{j=1}^2 \frac{\Gamma(\alpha_j + \delta_j(s-1))}{\Gamma(\alpha_j)} \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j + \beta_j + \delta_j(s-1))}. \tag{5.4.11}
 \end{aligned}$$

Accordingly, the density of u , denoted by $g(u)$, is the following:

$$\begin{aligned}
 g(u) &= C \frac{1}{2\pi i} \int_L \left\{ \prod_{j=1}^2 \frac{\Gamma(\alpha_j - \delta_j + \delta_j s)}{\Gamma(\alpha_j + \beta_j - \delta_j + \delta_j s)} \right\} u^{-s} ds \\
 &= C H_{2,2}^{2,0} \left[u \left| \begin{matrix} (\alpha_1 + \beta_1 - \delta_1, \delta_1), (\alpha_2 + \beta_2 - \delta_2, \delta_2) \\ (\alpha_1 - \delta_1, \delta_1), (\alpha_2 - \delta_2, \delta_2) \end{matrix} \right. \right], \\
 C &= \prod_{j=1}^2 \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)} \tag{5.4.12}
 \end{aligned}$$

where $0 \leq u \leq 1$, $\Re(\alpha_j - \delta_j + \delta_j s) > 0$, $\Re(\alpha_j) > 0$, $\Re(\beta_j) > 0$, $j = 1, 2$ and $g(u) = 0$ elsewhere.

When $\alpha_1 = 1 = \dots = \alpha_p$, $\beta_1 = 1 = \dots = \beta_q$, the H-function reduces to a G-function. This G-function is frequently referred to as Meijer's G-function and the H-function, as Fox's H-function.

5.4.4. Some special cases of the H-function

Certain special cases of the H-function are listed next.

$$\begin{aligned}
 H_{0,1}^{1,0}[x|(b,\beta)] &= \beta^{-1} x^{\frac{b}{\beta}} e^{-x^{\frac{1}{\beta}}}; \\
 H_{1,1}^{1,1}[z|_{(0,1)}^{(1-\nu,1)}] &= \Gamma(\nu)(1+z)^{-\nu} = \Gamma(\nu)_1 F_0(\nu; \ ; -z), |z| < 1; \\
 H_{0,2}^{1,0}\left[\frac{z^2}{4} \left| \begin{matrix} \nu \\ (\frac{a+\nu}{2}, \frac{a-\nu}{2}, 1) \end{matrix} \right. \right] &= \left(\frac{\nu}{2}\right)^a J_\nu(z)
 \end{aligned}$$

where the Bessel function

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{k! \Gamma(\nu+k+1)} = \frac{(z/2)^\nu}{\Gamma(\nu+1)} {}_0F_1\left(; 1+\nu; -\frac{z^2}{4}\right);$$

$$H_{1,2}^{1,1} \left[z \middle| \begin{matrix} (1-a, 1) \\ (0, 1), (1-c, 1) \end{matrix} \right] = \frac{\Gamma(a)}{\Gamma(c)} {}_1F_1(a; c; -z);$$

$$H_{2,2}^{1,2} \left[x \middle| \begin{matrix} (1-a, 1), (1-b, 1) \\ (0, 1), (1-c, 1) \end{matrix} \right] = \frac{\Gamma(a)\Gamma(b)}{\Gamma(c)} {}_2F_1(a, b; c; -z);$$

$$H_{1,2}^{1,1} \left[-z \middle| \begin{matrix} (1-\gamma, 1) \\ (0, 1), (1-\beta, \alpha) \end{matrix} \right] = \Gamma(\gamma) E_{\alpha, \beta}^\gamma(z), \quad \Re(\gamma) > 0,$$

where the generalized Mittag-Leffler function

$$E_{\alpha, \beta}^\gamma(z) = \sum_{k=0}^{\infty} \frac{(\gamma)_k}{k!} \frac{z^k}{\Gamma(\beta + \alpha k)}, \quad \Re(\alpha) > 0, \quad \Re(\beta) > 0,$$

where $\Gamma(\gamma)$ is defined. For $\gamma = 1$, we have $E_{\alpha, \beta}^1(z) = E_{\alpha, \beta}(z)$; when $\gamma = 1$, $\beta = 1$, $E_{\alpha, 1}^1(z) = E_\alpha(z)$ and when $\gamma = 1 = \beta = \alpha$, we have $E_1(z) = e^z$.

$$H_{0,2}^{2,0} \left[z \middle| \begin{matrix} \\ (0, 1), (\frac{\nu}{\rho}, \frac{1}{\rho}) \end{matrix} \right] = \rho K_\rho^\nu(z)$$

where $K_\rho^\nu(z)$ is Krätzel function

$$K_\rho^\nu(z) = \int_0^\infty t^{\nu-1} e^{-t^\rho - \frac{z}{t}} dt, \quad \Re(z) > 0.$$

$$H_{1,1}^{1,0} \left[x \middle| \begin{matrix} (\alpha + \frac{1}{2}, 1) \\ (\alpha, 1) \end{matrix} \right] = \pi^{-\frac{1}{2}} z^\alpha (1-z)^{-\frac{1}{2}}, \quad |z| < 1;$$

$$\begin{aligned}
 H_{2,2}^{2,0} \left[z \middle| \begin{matrix} (\alpha + \frac{1}{3}, 1), (\alpha + \frac{2}{3}, 1) \\ (\alpha, 1), (\alpha, 1) \end{matrix} \right] \\
 = z^\alpha {}_2F_1 \left(\frac{2}{3}, \frac{1}{3}; 1; 1 - z \right), \quad |1 - z| < 1.
 \end{aligned}$$

Exercises 5.4

5.4.1. Show that

$$z^\gamma G_{0,1}^{1,0} [pz^\alpha |_{\beta/\alpha}] = p^{\beta/\alpha} z^{\beta+\gamma} e^{-pz^\alpha}.$$

5.4.2. Show that

$$e^{-z} = G_{1,2}^{1,1} \left[z \middle| \begin{matrix} 1/3 \\ 0, 1/3 \end{matrix} \right] = G_{2,3}^{2,1} \left[z \middle| \begin{matrix} -\frac{1}{2}, \frac{1}{2} \\ 0, \frac{1}{2}, -\frac{1}{2} \end{matrix} \right].$$

5.4.3. Show that

$$z^{\frac{1}{3}} (1 - z)^{-\frac{5}{6}} = \Gamma(1/6) G_{1,1}^{1,0} \left[z \middle| \begin{matrix} \frac{1}{2} \\ \frac{1}{3} \end{matrix} \right].$$

5.4.4. Show that

$$\begin{aligned}
 \int_0^\infty x^{a-1} (1-x)^{b-c} (1+x-zx)^{-b} dx \\
 = \frac{\Gamma(a)\Gamma(c-a)}{\Gamma(c)} {}_2F_1(a, b; c; z), \quad |z| < 1, \Re(c-a) > 0.
 \end{aligned}$$

5.4.5. Show that

$$\begin{aligned}
 (a_1 - a_2) H_{p,q}^{m,n} \left[z \middle| \begin{matrix} (a_1, \alpha_1), (a_2, \alpha_1), (a_3, \alpha_3), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right] \\
 = H_{p,q}^{m,n} \left[z \middle| \begin{matrix} (a_1, \alpha_1), (a_2 - 1, \alpha_1), (a_3, \alpha_3), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right] \\
 - H_{p,q}^{m,n} \left[z \middle| \begin{matrix} (a_1 - 1, \alpha_1), (a_2, \alpha_1), (a_3, \alpha_3), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_q, \beta_q) \end{matrix} \right], \quad n \geq 2.
 \end{aligned}$$

5.5, 5.5a. The Wishart Density

A particular case of the real $p \times p$ matrix-variate gamma distribution, known as the Wishart distribution, is the preeminent distribution in multivariate statistical analysis. In the general $p \times p$ real matrix-variate gamma density with parameters $(\alpha, B > O)$, let $\alpha = \frac{m}{2}$, $B = \frac{1}{2}\Sigma^{-1}$ and $\Sigma > O$; the resulting density is called a Wishart density with

degrees of freedom m and parameter matrix $\Sigma > O$. This density, denoted by $f_w(W)$, is given by

$$f_w(W) = \frac{|W|^{\frac{m-p+1}{2}}}{2^{\frac{mp}{2}} |\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1}W)}, \quad W > O, \quad \Sigma > O, \quad (5.5.1)$$

for $m \geq p$, and $f_w(W) = 0$ elsewhere. This will be denoted as $W \sim W_p(m, \Sigma)$. Clearly, all the properties discussed in connection with the real matrix-variate gamma density still hold in this case. Algebraic evaluations of the marginal densities and explicit evaluations of the densities of sub-matrices will be considered, some aspects having already been discussed in Sects. 5.2 and 5.2.1.

In the complex case, the density is the following, denoted by $\tilde{f}_w(\tilde{W})$:

$$\tilde{f}_w(\tilde{W}) = \frac{|\det(\tilde{W})|^{m-p} e^{-\text{tr}(\Sigma^{-1}\tilde{W})}}{|\det(\Sigma)|^m \tilde{\Gamma}_p(m)}, \quad \tilde{W} > O, \quad \Sigma > O, \quad m \geq p, \quad (5.5a.1)$$

and $\tilde{f}_w(\tilde{W}) = 0$ elsewhere. This will be denoted as $\tilde{W} \sim \tilde{W}_p(m, \Sigma)$.

5.5.1. Explicit evaluations of the matrix-variate gamma integral, real case

Is it possible to evaluate the matrix-variate gamma integral explicitly by using conventional integration? We will now investigate some aspects of this question.

When the Wishart density is derived from samples coming from a Gaussian population, the basic technique relies on the triangularization process. When $\Sigma = I$, that is, $W \sim W_p(m, I)$, can the integral of the right-hand side of (5.5.1) be evaluated by resorting to conventional methods or by direct evaluation? We will address this problem by making use of the technique of partitioning matrices. Let us partition

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

where let $X_{22} = x_{pp}$ so that $X_{21} = (x_{p1}, \dots, x_{pp-1})$, $X_{12} = X'_{21}$. Then, on applying a result from Sect. 1.3, we have

$$|X|^{\alpha - \frac{p+1}{2}} = |X_{11}|^{\alpha - \frac{p+1}{2}} [x_{pp} - X_{21} X_{11}^{-1} X_{12}]^{\alpha - \frac{p+1}{2}}. \quad (5.5.2)$$

Note that when X is positive definite, $X_{11} > O$ and $x_{pp} > 0$, and the quadratic form $X_{21} X_{11}^{-1} X_{12} > 0$. As well,

$$[x_{pp} - X_{21} X_{11}^{-1} X_{12}]^{\alpha - \frac{p+1}{2}} = x_{pp}^{\alpha - \frac{p+1}{2}} [1 - x_{pp}^{-\frac{1}{2}} X_{21} X_{11}^{-\frac{1}{2}} X_{11}^{-\frac{1}{2}} X_{12} x_{pp}^{-\frac{1}{2}}]^{\alpha - \frac{p+1}{2}}. \quad (5.5.3)$$

Letting $Y = x_{pp}^{-\frac{1}{2}} X_{21} X_{11}^{-\frac{1}{2}}$, then referring to Mathai (1997, Theorem 1.18) or Theorem 1.6.4 of Chap. 1, $dY = x_{pp}^{-\frac{p-1}{2}} |X_{11}|^{-\frac{1}{2}} dX_{21}$ for fixed X_{11} and x_{pp} . The integral over x_{pp} gives

$$\int_0^\infty x_{pp}^{\alpha + \frac{p-1}{2} - \frac{p+1}{2}} e^{-x_{pp}} dx_{pp} = \Gamma(\alpha), \Re(\alpha) > 0.$$

If we let $u = YY'$, then from Theorem 2.16 and Remark 2.13 of Mathai (1997) or using Theorem 4.2.3, after integrating out over the Stiefel manifold, we have

$$dY = \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} u^{\frac{p-1}{2}-1} du.$$

(Note that n in Theorem 2.16 of Mathai (1997) corresponds to $p - 1$ and p is 1). Then, the integral over u gives

$$\int_0^1 u^{\frac{p-1}{2}-1} (1-u)^{\alpha - \frac{p+1}{2}} du = \frac{\Gamma(\frac{p-1}{2})\Gamma(\alpha - \frac{p-1}{2})}{\Gamma(\alpha)}, \Re(\alpha) > \frac{p-1}{2}.$$

Now, collecting all the factors, we have

$$\begin{aligned} |X_{11}|^{\alpha + \frac{1}{2} - \frac{p+1}{2}} \Gamma(\alpha) \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \frac{\Gamma(\frac{p-1}{2})\Gamma(\alpha - \frac{p-1}{2})}{\Gamma(\alpha)} \\ = |X_{11}^{(1)}|^{\alpha + \frac{1}{2} - \frac{p+1}{2}} \pi^{\frac{p-1}{2}} \Gamma(\alpha - (p-1)/2) \end{aligned}$$

for $\Re(\alpha) > \frac{p-1}{2}$. Note that $|X_{11}^{(1)}|$ is $(p-1) \times (p-1)$ and $|X_{11}|$, after the completion of the first part of the operations, is denoted by $|X_{11}^{(1)}|$, the exponent being changed to $\alpha + \frac{1}{2} - \frac{p+1}{2}$. Now repeat the process by separating $x_{p-1,p-1}$, that is, by writing

$$X_{11}^{(1)} = \begin{bmatrix} X_{11}^{(2)} & X_{12}^{(2)} \\ X_{21}^{(2)} & x_{p-1,p-1} \end{bmatrix}.$$

Here, $X_{11}^{(2)}$ is of order $(p-2) \times (p-2)$ and $X_{21}^{(2)}$ is of order $1 \times (p-2)$. As before, letting $u = YY'$ with $Y = x_{p-1,p-1}^{-\frac{1}{2}} X_{21}^{(2)} [X_{11}^{(2)}]^{-\frac{1}{2}}$, $dY = x_{p-1,p-1}^{-\frac{p-2}{2}} |X_{11}^{(2)}|^{-\frac{1}{2}} dX_{21}^{(2)}$. The integral over the Stiefel manifold gives $\frac{\pi^{\frac{p-2}{2}}}{\Gamma(\frac{p-2}{2})} u^{\frac{p-2}{2}-1} du$ and the factor containing $(1-u)$ is $(1-u)^{\alpha + \frac{1}{2} - \frac{p+1}{2}}$, the integral over u yielding

$$\int_0^1 u^{\frac{p-2}{2}-1} (1-u)^{\alpha + \frac{1}{2} - \frac{p+1}{2}} du = \frac{\Gamma(\frac{p-2}{2})\Gamma(\alpha - \frac{p-2}{2})}{\Gamma(\alpha)}$$

and that over $v = x_{p-1,p-1}$ giving

$$\int_0^1 v^{\alpha + \frac{1}{2} + \frac{p-2}{2} - \frac{p+1}{2}} e^{-v} dv = \Gamma(\alpha), \Re(\alpha) > 0.$$

The product of these factors is then

$$|X_{11}^{(2)}|^{\alpha+1-\frac{p+1}{2}} \pi^{\frac{p-2}{2}} \Gamma(\alpha - (p-2)/2), \Re(\alpha) > \frac{p-2}{2}.$$

Successive evaluations carried out by employing the same procedure yield the exponent of π as $\frac{p-1}{2} + \frac{p-2}{2} + \dots + \frac{1}{2} = \frac{p(p-1)}{4}$ and the gamma product, $\Gamma(\alpha - \frac{p-1}{2})\Gamma(\alpha - \frac{p-2}{2}) \dots \Gamma(\alpha)$, the final result being $\Gamma_p(\alpha)$. The result is thus verified.

5.5a.1. Evaluation of matrix-variate gamma integrals in the complex case

The matrices and gamma functions belonging to the complex domain will be denoted with a tilde. As well, in the complex case, all matrices appearing in the integrals will be $p \times p$ Hermitian positive definite unless otherwise stated; as an example, for such a matrix X , this will be denoted by $\tilde{X} > O$. The integral of interest is

$$\tilde{I}_p(\alpha) = \int_{\tilde{X} > O} |\det(\tilde{X})|^{\alpha-p} e^{-\text{tr}(\tilde{X})} d\tilde{X}. \quad (5.5a.2)$$

A standard procedure for evaluating the integral in (5.5a.2) consists of expressing the positive definite Hermitian matrix as $\tilde{X} = \tilde{T}\tilde{T}^*$ where \tilde{T} is a lower triangular matrix with real and positive diagonal elements $t_{jj} > 0$, $j = 1, \dots, p$, where an asterisk indicates the conjugate transpose. Then, referring to (Mathai (1997, Theorem 3.7) or Theorem 1.6.7 of Chap. 1, the Jacobian is seen to be as follows:

$$d\tilde{X} = 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T} \quad (5.5a.3)$$

and then

$$\begin{aligned} \text{tr}(\tilde{X}) &= \text{tr}(\tilde{T}\tilde{T}^*) \\ &= t_{11}^2 + \dots + t_{pp}^2 + |\tilde{t}_{21}|^2 + \dots + |\tilde{t}_{p1}|^2 + \dots + |\tilde{t}_{pp-1}|^2 \end{aligned}$$

and

$$|\det(\tilde{X})|^{\alpha-p} d\tilde{X} = 2^p \left\{ \prod_{j=1}^p t_{jj}^{2\alpha-2j+1} \right\} d\tilde{T}.$$

Now, integrating out over \tilde{t}_{jk} for $j > k$,

$$\int_{\tilde{t}_{jk}} e^{-|\tilde{t}_{jk}|^2} d\tilde{t}_{jk} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(t_{jk1}^2+t_{jk2}^2)} dt_{jk1} \wedge dt_{jk2} = \pi$$

and

$$\prod_{j>k} \pi = \pi^{\frac{p(p-1)}{2}}.$$

As well,

$$2 \int_0^{\infty} t_{jj}^{2\alpha-2j+1} e^{-t_{jj}^2} dt_{jj} = \Gamma(\alpha - j + 1), \Re(\alpha) > j - 1,$$

for $j = 1, \dots, p$. Taking the product of all these factors then gives

$$\pi^{\frac{p(p-1)}{2}} \Gamma(\alpha)\Gamma(\alpha - 1) \cdots \Gamma(\alpha - p + 1) = \tilde{\Gamma}_p(\alpha), \Re(\alpha) > p - 1,$$

and hence the result is verified.

An alternative method based on partitioned matrix, complex case

The approach discussed in this section relies on the successive extraction of the diagonal elements of \tilde{X} , a $p \times p$ positive definite Hermitian matrix, all of these elements being necessarily real and positive, that is, $x_{jj} > 0, j = 1, \dots, p$. Let

$$\tilde{X} = \begin{bmatrix} \tilde{X}_{11} & \tilde{X}_{12} \\ \tilde{X}_{21} & x_{pp} \end{bmatrix}$$

where \tilde{X}_{11} is $(p - 1) \times (p - 1)$ and

$$|\det(\tilde{X})|^{\alpha-p} = |\det(\tilde{X}_{11})|^{\alpha-p} |x_{pp} - \tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}|^{\alpha-p}$$

and

$$\text{tr}(\tilde{X}) = \text{tr}(\tilde{X}_{11}) + x_{pp}.$$

Then,

$$|x_{pp} - \tilde{X}_{21} \tilde{X}_{11}^{-1} \tilde{X}_{12}|^{\alpha-p} = x_{pp}^{\alpha-p} |1 - x_{pp}^{-\frac{1}{2}} \tilde{X}_{21} \tilde{X}_{11}^{-\frac{1}{2}} \tilde{X}_{11}^{-\frac{1}{2}} \tilde{X}_{12} x_{pp}^{-\frac{1}{2}}|^{\alpha-p}.$$

Let

$$\tilde{Y} = x_{pp}^{-\frac{1}{2}} \tilde{X}_{21} \tilde{X}_{11}^{-\frac{1}{2}} \Rightarrow d\tilde{Y} = x_{pp}^{-(p-1)} |\det(\tilde{X}_{11})|^{-1} d\tilde{X}_{21},$$

referring to Theorem 1.6a.4 or Mathai (1997, Theorem 3.2(c)) for fixed x_{pp} and X_{11} . Now, the integral over x_{pp} gives

$$\int_0^\infty x_{pp}^{\alpha-p+(p-1)} e^{-x_{pp}} dx_{pp} = \Gamma(\alpha), \Re(\alpha) > 0.$$

Letting $u = \tilde{Y}\tilde{Y}^*$, $d\tilde{Y} = u^{p-2} \frac{\pi^{p-1}}{\Gamma(p-1)} du$ by applying Theorem 4.2a.3 or Corollaries 4.5.2 and 4.5.3 of Mathai (1997), and noting that u is real and positive, the integral over u gives

$$\int_0^\infty u^{(p-1)-1} (1-u)^{\alpha-(p-1)-1} du = \frac{\Gamma(p-1)\Gamma(\alpha-(p-1))}{\Gamma(\alpha)}, \Re(\alpha) > p-1.$$

Taking the product, we obtain

$$\begin{aligned} & |\det(\tilde{X}_{11}^{(1)})|^{\alpha+1-p} \Gamma(\alpha) \frac{\pi^{p-1}}{\Gamma(p-1)} \frac{\Gamma(p-1)\Gamma(\alpha-(p-1))}{\Gamma(\alpha)} \\ &= \pi^{p-1} \Gamma(\alpha-(p-1)) |\det(\tilde{X}_{11}^{(1)})|^{\alpha+1-p} \end{aligned}$$

where $\tilde{X}_{11}^{(1)}$ stands for \tilde{X}_{11} after having completed the first set of integrations. In the second stage, we extract $x_{p-1,p-1}$, the first $(p-2) \times (p-2)$ submatrix being denoted by $\tilde{X}_{11}^{(2)}$ and we continue as previously explained to obtain $|\det(\tilde{X}_{11}^{(2)})|^{\alpha+2-p} \pi^{p-2} \Gamma(\alpha-(p-2))$. Proceeding successively in this manner, we have the exponent of π as $(p-1) + (p-2) + \dots + 1 = p(p-1)/2$ and the gamma product as $\Gamma(\alpha-(p-1))\Gamma(\alpha-(p-2)) \dots \Gamma(\alpha)$ for $\Re(\alpha) > p-1$. That is,

$$\pi^{\frac{p(p-1)}{2}} \Gamma(\alpha)\Gamma(\alpha-1) \dots \Gamma(\alpha-(p-1)) = \tilde{\Gamma}_p(\alpha).$$

5.5.2. Triangularization of the Wishart matrix in the real case

Let $W \sim W_p(m, \Sigma)$, $\Sigma > O$ be a $p \times p$ matrix having a Wishart distribution with m degrees of freedom and parameter matrix $\Sigma > O$, that is, let W have a density of the following form for $\Sigma = I$:

$$f_w(W) = \frac{|W|^{\frac{m-p+1}{2}} e^{-\frac{1}{2}\text{tr}(W)}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})}, \quad W > O, \quad m \geq p, \quad (5.5.4)$$

and $f_w(W) = 0$ elsewhere. Let us consider the transformation $W = TT'$ where T is a lower triangular matrix with positive diagonal elements. Since $W > O$, the transformation

$W = TT'$ with the diagonal elements of T being positive is one-to-one. We have already evaluated the associated Jacobian in Theorem 1.6.7, namely,

$$dW = 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT. \tag{5.5.5}$$

Under this transformation,

$$\begin{aligned} f(W)dW &= \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \left\{ \prod_{j=1}^p (t_{jj}^2)^{\frac{m}{2}-\frac{p+1}{2}} \right\} e^{-\frac{1}{2} \sum_{i \geq j} t_{ij}^2} 2^p \left\{ \prod_{j=1}^p t_{jj}^{p+1-j} \right\} dT \\ &= \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} 2^p \left\{ \prod_{j=1}^p (t_{jj}^2)^{\frac{m}{2}-\frac{j}{2}} \right\} e^{-\frac{1}{2} \sum_{j=1}^p t_{jj}^2 - \frac{1}{2} \sum_{i>j} t_{ij}^2} dT. \end{aligned} \tag{5.5.6}$$

In view of (5.5.6), it is evident that t_{jj} , $j = 1, \dots, p$ and the t_{ij} 's, $i > j$ are mutually independently distributed. The form of the function containing t_{ij} , $i > j$, is $e^{-\frac{1}{2}t_{ij}^2}$, and hence the t_{ij} 's for $i > j$ are mutually independently distributed real standard normal variables. It is also seen from (5.5.6) that the density of t_{jj}^2 is of the form

$$c_j y_j^{\frac{m}{2}-\frac{j-1}{2}-1} e^{-\frac{1}{2}y_j}, \quad y_j = t_{jj}^2,$$

which is the density of a real chisquare variable having $m - (j - 1)$ degrees of freedom for $j = 1, \dots, p$, where c_j is the normalizing constant. Hence, the following result:

Theorem 5.5.1. *Let the real $p \times p$ positive definite matrix W have a real Wishart density as specified in (5.5.4) and let $W = TT'$ where $T = (t_{ij})$ is a lower triangular matrix whose diagonal elements are positive. Then, the non-diagonal elements t_{ij} such that $i > j$ are mutually independently distributed as real standard normal variables, the diagonal elements t_{jj}^2 , $j = 1, \dots, p$, are independently distributed as a real chisquare variables having $m - (j - 1)$ degrees of freedom for $j = 1, \dots, p$, and the t_{jj}^2 's and t_{ij} 's are mutually independently distributed.*

Corollary 5.5.1. *Let $W \sim W_p(n, \sigma^2 I)$, where $\sigma^2 > 0$ is a real scalar quantity. Let $W = TT'$ where $T = (t_{ij})$ is a lower triangular matrix whose diagonal elements are positive. Then, the t_{jj} 's are independently distributed for $j = 1, \dots, p$, the t_{ij} 's, $i > j$, are independently distributed, and all t_{jj} 's and t_{ij} 's are mutually independently distributed, where t_{jj}^2/σ^2 has a real chisquare distribution with $m - (j - 1)$ degrees of freedom for $j = 1, \dots, p$, and t_{ij} , $i > j$, has a real scalar Gaussian distribution with mean value zero and variance σ^2 , that is, $t_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ for all $i > j$.*

5.5a.2. Triangularization of the Wishart matrix in the complex domain

Let \tilde{W} have the following Wishart density in the complex domain:

$$\tilde{f}_w(\tilde{W}) = \frac{1}{\tilde{\Gamma}_p(m)} |\det(\tilde{W})|^{m-p} e^{-\text{tr}(\tilde{W})}, \quad \tilde{W} > O, \quad m \geq p, \quad (5.5a.4)$$

and $\tilde{f}_w(\tilde{W}) = 0$ elsewhere, which is denoted $\tilde{W} \sim \tilde{W}_p(m, I)$. Consider the transformation $\tilde{W} = \tilde{T}\tilde{T}^*$ where \tilde{T} is lower triangular whose diagonal elements are real and positive. The transformation $\tilde{W} = \tilde{T}\tilde{T}^*$ is then one-to-one and its associated Jacobian, as given in Theorem 1.6a.7, is the following:

$$d\tilde{W} = 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T}. \quad (5.5a.5)$$

Then we have

$$\begin{aligned} \tilde{f}_w(\tilde{W}) d\tilde{W} &= \frac{1}{\tilde{\Gamma}_p(m)} \left\{ \prod_{j=1}^p (t_{jj}^2)^{m-p} \right\} e^{-\sum_{i \geq j} |\tilde{t}_{ij}|^2} 2^p \left\{ \prod_{j=1}^p t_{jj}^{2(p-j)+1} \right\} d\tilde{T} \\ &= \frac{1}{\tilde{\Gamma}_p(m)} 2^p \left\{ \prod_{j=1}^p (t_{jj}^2)^{m-j+\frac{1}{2}} \right\} e^{-\sum_{j=1}^p t_{jj}^2 - \sum_{i>j} |\tilde{t}_{ij}|^2} d\tilde{T}. \end{aligned} \quad (5.5a.6)$$

In light of (5.5a.6), it is clear that all the t_{jj} 's and \tilde{t}_{ij} 's are mutually independently distributed where \tilde{t}_{ij} , $i > j$, has a complex standard Gaussian density and t_{jj}^2 has a complex chisquare density with degrees of freedom $m - (j - 1)$ or a real gamma density with the parameters $(\alpha = m - (j - 1), \beta = 1)$, for $j = 1, \dots, p$. Hence, we have the following result:

Theorem 5.5a.1. *Let the complex Wishart density be as specified in (5.5a.4), that is, $\tilde{W} \sim \tilde{W}_p(m, I)$. Consider the transformation $\tilde{W} = \tilde{T}\tilde{T}^*$ where $\tilde{T} = (\tilde{t}_{ij})$ is a lower triangular matrix in the complex domain whose diagonal elements are real and positive. Then, for $i > j$, the \tilde{t}_{ij} 's are standard Gaussian distributed in the complex domain, that is, $\tilde{t}_{ij} \sim \tilde{N}_1(0, 1)$, $i > j$, t_{jj}^2 is real gamma distributed with the parameters $(\alpha = m - (j - 1), \beta = 1)$ for $j = 1, \dots, p$, and all the t_{jj} 's and \tilde{t}_{ij} 's, $i > j$, are mutually independently distributed.*

Corollary 5.5a.1. *Let $\tilde{W} \sim \tilde{W}_p(m, \sigma^2 I)$ where $\sigma^2 > 0$ is a real positive scalar. Let \tilde{T} , t_{jj} , \tilde{t}_{ij} , $i > j$, be as defined in Theorem 5.5a.1. Then, t_{jj}^2/σ^2 is a real gamma variable with the parameters $(\alpha = m - (j - 1), \beta = 1)$ for $j = 1, \dots, p$, $\tilde{t}_{ij} \sim \tilde{N}_1(0, \sigma^2)$ for all $i > j$, and the t_{jj} 's and \tilde{t}_{ij} 's are mutually independently distributed.*

5.5.3. Samples from a p -variate Gaussian population and the Wishart density

Let the $p \times 1$ real vector X_j be normally distributed, $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$. Let X_1, \dots, X_n be a simple random sample of size n from this normal population and the $p \times n$ sample matrix be denoted in bold face lettering as $\mathbf{X} = (X_1, X_2, \dots, X_n)$ where $X_j' = (x_{1j}, x_{2j}, \dots, x_{pj})$. Let the sample mean be $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ and the matrix of sample means be denoted by the bold face $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$. Then, the $p \times p$ sample sum of products matrix S is given by

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = (s_{ij}), \quad s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

where $\bar{x}_r = \sum_{k=1}^n x_{rk}/n$, $r = 1, \dots, p$, are the averages on the components. It has already been shown in Sect. 3.5 for instance that the joint density of the sample values X_1, \dots, X_n , denoted by L , can be written as

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu)}. \quad (5.5.7)$$

But $(\mathbf{X} - \bar{\mathbf{X}})J = O$, $J' = (1, \dots, 1)$, which implies that the columns of $(\mathbf{X} - \bar{\mathbf{X}})$ are linearly related, and hence the elements in $(\mathbf{X} - \bar{\mathbf{X}})$ are not distinct. In light of equation (4.5.17), one can write the sample sum of products matrix S in terms of a $p \times (n-1)$ matrix Z_{n-1} of distinct elements so that $S = Z_{n-1}Z_{n-1}'$. As well, according to Theorem 3.5.3 of Chap. 3, S and \bar{X} are independently distributed. The $p \times n$ matrix Z is obtained through the orthonormal transformation $\mathbf{X}P = Z$, $PP' = I$, $P'P = I$ where P is $n \times n$. Then $d\mathbf{X} = dZ$, ignoring the sign. Let the last column of P be p_n . We can specify p_n to be $\frac{1}{\sqrt{n}}J$ so that $\mathbf{X}p_n = \sqrt{n}\bar{\mathbf{X}}$. Note that in light of (4.5.17), the deleted column in Z corresponds to $\sqrt{n}\bar{\mathbf{X}}$. The following considerations will be helpful to those who might need further confirmation of the validity of the above statement. Observe that $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}(I - B)$, with $B = \frac{1}{n}JJ'$ where J is a $n \times 1$ vector of unities. Since $I - B$ is idempotent and of rank $n - 1$, the eigenvalues are 1 repeated $n - 1$ times and a zero. An eigenvector, corresponding to the eigenvalue zero, is J normalized or $\frac{1}{\sqrt{n}}J$. Taking this as the last column p_n of P , we have $\mathbf{X}p_n = \sqrt{n}\bar{\mathbf{X}}$. Note that the other columns of P , namely p_1, \dots, p_{n-1} , correspond to the $n - 1$ orthonormal solutions coming from the equation $BY = Y$ where Y is a $n \times 1$ non-null vector. Hence we can write $dZ = dZ_{n-1} \wedge d\bar{X}$. Now, integrating out \bar{X} from (5.5.7), we have

$$L dZ_{n-1} = c e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}Z_{n-1}Z_{n-1}')} dZ_{n-1}, \quad S = Z_{n-1}Z_{n-1}', \quad (5.5.8)$$

where c is a constant. Since Z_{n-1} contains $p(n-1)$ distinct real variables, we may apply Theorems 4.2.1, 4.2.2 and 4.2.3, and write dZ_{n-1} in terms of dS as

$$dZ_{n-1} = \frac{\pi^{\frac{p(n-1)}{2}}}{\Gamma_p(\frac{n-1}{2})} |S|^{\frac{n-1}{2} - \frac{p+1}{2}} dS, \quad n-1 \geq p. \quad (5.5.9)$$

Then, if the density of S is denoted by $f(S)$,

$$f(S)dS = c_1 \frac{|S|^{\frac{n-1}{2} - \frac{p+1}{2}}}{\Gamma_p(\frac{n-1}{2})} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} dS$$

where c_1 is a constant. From a real matrix-variate gamma density, we have the normalizing constant, thereby the value of c_1 . Hence

$$f(S)dS = \frac{|S|^{\frac{n-1}{2} - \frac{p+1}{2}}}{2^{\frac{(n-1)p}{2}} |\Sigma|^{\frac{n-1}{2}} \Gamma_p(\frac{n-1}{2})} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} dS \quad (5.5.10)$$

for $S > O$, $\Sigma > O$, $n-1 \geq p$ and $f(X) = 0$ elsewhere, $\Gamma_p(\cdot)$ being the real matrix-variate gamma given by

$$\Gamma_p(\alpha) = \pi^{\frac{p(p-1)}{4}} \Gamma(\alpha) \Gamma(\alpha - 1/2) \cdots \Gamma(\alpha - (p-1)/2), \quad \Re(\alpha) > (p-1)/2.$$

Usually the sample size is taken as N so that $N-1 = n$ the number of degrees of freedom associated with the Wishart density in (5.5.10). Since we have taken the sample size as n , the number of degrees of freedom is $n-1$ and the parameter matrix is $\Sigma > O$. Then S in (5.5.10) is written as $S \sim W_p(m, \Sigma)$, with $m = n-1 \geq p$. Thus, the following result:

Theorem 5.5.2. *Let X_1, \dots, X_n be a simple random sample of size n from a $N_p(\mu, \Sigma)$, $\Sigma > O$. Let X_j , \mathbf{X} , \bar{X} , $\bar{\mathbf{X}}$, S be as defined in Sect. 5.5.3. Then, the density of S is a real Wishart density with $m = n-1$ degrees of freedom and parameter matrix $\Sigma > O$, as given in (5.5.10).*

5.5a.3. Sample from a complex Gaussian population and the Wishart density

Let $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$ be independently distributed. Let $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)$, $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$, $\bar{\tilde{\mathbf{X}}} = (\bar{\tilde{X}}, \dots, \bar{\tilde{X}})$ and let $\tilde{S} = (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})(\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})^*$ where a * indicates the conjugate transpose. We have already shown in Sect. 3.5a that the joint density of $\tilde{X}_1, \dots, \tilde{X}_n$, denoted by \tilde{L} , can be written as

$$\tilde{L} = \frac{1}{\pi^{np} |\det(\Sigma)|^n} e^{-\text{tr}(\Sigma^{-1}\tilde{S}) - n(\bar{\tilde{X}} - \tilde{\mu})^* \Sigma^{-1}(\bar{\tilde{X}} - \tilde{\mu})}. \quad (5.5a.7)$$

Then, following steps parallel to (5.5.7) to (5.5.10), we obtain the density of \tilde{S} , denoted by $\tilde{f}(\tilde{S})$, as the following:

$$\tilde{f}(\tilde{S})d\tilde{S} = \frac{|\det(S)|^{m-p}}{|\det(\Sigma)|^m \tilde{\Gamma}_p(m)} e^{-\text{tr}(\Sigma^{-1}S)} d\tilde{S}, \quad m = n - 1 \geq p, \quad (5.5a.8)$$

for $\tilde{S} > O$, $\Sigma > O$, $n - 1 \geq p$, and $\tilde{f}(\tilde{S}) = 0$ elsewhere, where the complex matrix-variate gamma function being given by

$$\tilde{\Gamma}_p(\alpha) = \pi^{\frac{p(p-1)}{2}} \Gamma(\alpha) \Gamma(\alpha - 1) \cdots \Gamma(\alpha - p + 1), \quad \Re(\alpha) > p - 1.$$

Hence, we have the following result:

Theorem 5.5a.2. *Let $\tilde{X}_j \sim \tilde{N}_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, be independently and identically distributed. Let $\tilde{\mathbf{X}}$, \tilde{X} , $\tilde{\mathbf{X}}$, \tilde{S} be as previously defined. Then, \tilde{S} has a complex matrix-variate Wishart density with $m = n - 1$ degrees of freedom and parameter matrix $\Sigma > O$, as given in (5.5a.8).*

5.5.4. Some properties of the Wishart distribution, real case

If we have statistically independently distributed Wishart matrices with the same parameter matrix Σ , then it is easy to see that the sum is again a Wishart matrix. This can be noted by considering the Laplace transform of matrix-variate random variables discussed in Sect. 5.2. If $S_j \sim W_p(m_j, \Sigma)$, $j = 1, \dots, k$, with the same parameter matrix $\Sigma > O$ and the S_j 's are statistically independently distributed, then from equation (5.2.6), the Laplace transform of the density of S_j is

$$L_{S_j}(*T) = |I + 2\Sigma_*T|^{-\frac{m_j}{2}}, \quad I + 2\Sigma_*T > O, \quad j = 1, \dots, k, \quad (5.5.11)$$

where $*T$ is a symmetric parameter matrix $T = (t_{ij}) = T' > O$ with off-diagonal elements weighted by $\frac{1}{2}$. When S_j 's are independently distributed, then the Laplace transform of the sum $S = S_1 + \cdots + S_k$ is the product of the Laplace transforms:

$$\prod_{j=1}^k |I + 2\Sigma_*T|^{-\frac{m_j}{2}} = |I + 2\Sigma_*T|^{-\frac{1}{2}(m_1 + \cdots + m_k)} \Rightarrow S \sim W_p(m_1 + \cdots + m_k, \Sigma). \quad (5.5.12)$$

Hence, the following result:

Theorems 5.5.3, 5.5a.3. Let $S_j \sim W_p(m_j, \Sigma)$, $\Sigma > O$, $j = 1, \dots, k$, be statistically independently distributed real Wishart matrices with m_1, \dots, m_k degrees of freedoms and the same parameter matrix $\Sigma > O$. Then the sum $S = S_1 + \dots + S_k$ is real Wishart distributed with degrees of freedom $m_1 + \dots + m_k$ and the same parameter matrix $\Sigma > O$, that is, $S \sim W_p(m_1 + \dots + m_k, \Sigma)$, $\Sigma > O$. In the complex case, let $\tilde{S}_j \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$, $j = 1, \dots, k$, be independently distributed with the same $\tilde{\Sigma}$. Then, the sum $\tilde{S} = \tilde{S}_1 + \dots + \tilde{S}_k \sim \tilde{W}_p(m_1 + \dots + m_k, \tilde{\Sigma})$.

We now consider linear functions of independent Wishart matrices. Let $S_j \sim W_p(m_j, \Sigma)$, $\Sigma > O$, $j = 1, \dots, k$, be independently distributed and $S_a = a_1 S_1 + \dots + a_k S_k$ where a_1, \dots, a_k are real scalar constants, then the Laplace transform of the density of S_a is

$$L_{S_a}(*T) = \prod_{j=1}^k |I + 2a_j \Sigma_* T|^{-\frac{m_j}{2}}, \quad I + 2a_j \Sigma_* T > O, \quad j = 1, \dots, k. \quad (i)$$

The inverse is quite complicated and the corresponding density cannot be easily determined; moreover, the density is not a Wishart density unless $a_1 = \dots = a_k$. The types of complications occurring can be apprehended from the real scalar case $p = 1$ which is discussed in Mathai and Provost (1992). Instead of real scalars, we can also consider $p \times p$ constant matrices as coefficients, in which case the inversion of the Laplace transform will be more complicated. We can also consider Wishart matrices with different parameter matrices. Let $U_j \sim W_p(m_j, \Sigma_j)$, $\Sigma_j > O$, $j = 1, \dots, k$, be independently distributed and $U = U_1 + \dots + U_k$. Then, the Laplace transform of the density of U , denoted by $L_U(*T)$, is the following:

$$L_U(*T) = \prod_{j=1}^k |I + 2\Sigma_{j*} T|^{-\frac{m_j}{2}}, \quad I + 2\Sigma_{j*} T > O, \quad j = 1, \dots, k. \quad (ii)$$

This case does not yield a Wishart density as an inverse Laplace transform either, unless $\Sigma_1 = \dots = \Sigma_k$. In both (i) and (ii), we have linear functions of independent Wishart matrices; however, these linear functions do not have Wishart distributions.

Let us consider a symmetric transformation on a Wishart matrix S . Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$ and $U = ASA'$ where A is a $p \times p$ nonsingular constant matrix. Let us take the Laplace transform of the density of U :

$$\begin{aligned}
L_U(*T) &= E[e^{-\text{tr}(*T'U)}] = E[e^{-\text{tr}(*T'ASA')}] = E[e^{-\text{tr}(A'*TAS)}] \\
&= E[e^{-\text{tr}[(A'*TA)'S]}] = L_S(A'*TA) = |I + 2\Sigma(A'*TA)|^{-\frac{m}{2}} \\
&= |I + 2(A\Sigma A')_*T|^{-\frac{m}{2}} \\
&\Rightarrow U \sim W_p(m, A\Sigma A'), \Sigma > O, |A| \neq 0.
\end{aligned}$$

Hence we have the following result:

Theorems 5.5.4, 5.5a.4. *Let $S \sim W_p(m, \Sigma > O)$ and $U = ASA'$, $|A| \neq 0$. Then, $U \sim W_p(m, A\Sigma A')$, $\Sigma > O$, $|A| \neq 0$, that is, when $U = ASA'$ where A is a nonsingular $p \times p$ constant matrix, then U is Wishart distributed with degrees of freedom m and parameter matrix $A\Sigma A'$. In the complex case, the constant $p \times p$ nonsingular matrix A can be real or in the complex domain. Let $\tilde{S} \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$. Then $\tilde{U} = A\tilde{S}A^* \sim \tilde{W}_p(m, A\tilde{\Sigma}A^*)$.*

If A is not a nonsingular matrix, is there a corresponding result? Let B be a constant $q \times p$ matrix, $q \leq p$, which is of full rank q . Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, be iid so that we have a simple random sample of size n from a real p -variate Gaussian population. Let the $q \times 1$ vectors $Y_j = BX_j$, $j = 1, \dots, n$, be iid. Then $E[Y_j] = B\mu$, $\text{Cov}(Y_j) = E[(Y_j - E(Y_j))(Y_j - E(Y_j))'] = BE[(X_j - E(X_j))(X_j - E(X_j))'B'] = B\Sigma B'$ which is $q \times q$. As well, $Y_j \sim N_q(B\mu, B\Sigma B')$, $B\Sigma B' > O$. Consider the sample matrix formed from the Y_j 's, namely the $q \times n$ matrix $\mathbf{Y} = (Y_1, \dots, Y_n) = (BX_1, \dots, BX_n) = B(X_1, \dots, X_n) = B\mathbf{X}$ where \mathbf{X} is the $p \times n$ sample matrix from X_j . Then, the sample sum of products matrix in \mathbf{Y} is $(\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})' = S_y$, say, where the usual notation is utilized, namely, $\bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n)$ and $\bar{\mathbf{Y}} = (\bar{Y}, \dots, \bar{Y})$. Now, the problem is equivalent to taking a simple random sample of size n from a q -variate real Gaussian population with mean value vector $B\mu$ and positive definite covariance matrix $B\Sigma B' > O$. Hence, the following result:

Theorems 5.5.5, 5.5a.5. *Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, be iid, and S be the sample sum of products matrix in this p -variate real Gaussian population. Let B be a $q \times p$ constant matrix, $q \leq p$, which has full rank q . Then BSB' is real Wishart distributed with degrees of freedom $m = n - 1$, n being equal to the sample size, and parameter matrix $B\Sigma B' > O$, that is, $BSB \sim W_q(m, B\Sigma B')$. Similarly, in the complex case, let B be a $q \times p$, $q \leq p$, constant matrix of full rank q , where B may be in the real or complex domain. Then, $B\tilde{S}B^*$ is Wishart distributed with degrees of freedom m and parameter matrix $B\tilde{\Sigma}B^*$, that is, $B\tilde{S}B^* \sim \tilde{W}_q(m, B\tilde{\Sigma}B^*)$.*

5.5.5. The generalized variance

Let X_j , $X_j' = (x_{1j}, \dots, x_{pj})$, be a real $p \times 1$ vector random variable for $j = 1, \dots, n$, and the X_j 's be iid (independently and identically distributed) as X_j . Let the covariance

matrix associated with X_j be $\text{Cov}(X_j) = E[(X_j - E(X_j))(X_j - E(X_j))'] = \Sigma$, $\Sigma \geq O$, for $j = 1, \dots, n$ in the real case and $\tilde{\Sigma} = E[(\tilde{X}_j - E(\tilde{X}_j))(\tilde{X}_j - E(\tilde{X}_j))^*]$ in the complex case, where an asterisk indicates the conjugate transpose. Then, the diagonal elements in Σ represent the squares of a measure of scatter or variances associated with the elements x_{1j}, \dots, x_{pj} and the off-diagonal elements in Σ provide the corresponding measure of joint dispersion or joint scatter in the pair (x_{rj}, x_{sj}) for all $r \neq s$. Thus, Σ gives a configuration of individual and joint squared scatter in all the elements x_{1j}, \dots, x_{pj} . If we wish to have a single number or single scalar quantity representing this configuration of individual and joint scatter in the elements x_{1j}, \dots, x_{pj} what should be that measure? Wilks had taken the determinant of Σ , $|\Sigma|$, as that measure and called it the *generalized variance* or square of the scatter representing the whole configuration of scatter in all the elements x_{1j}, \dots, x_{pj} . If there is no scatter in one or in a few elements but there is scatter or dispersion in all other elements, then the determinant is zero. If the matrix is singular then the determinant is zero, but this does not mean that there is no scatter in these elements. Thus, determinant as a measure of scatter or dispersion, violates a very basic condition that if the proposed measure is zero then there should not be any scatter in any of the elements or Σ should be a null matrix. Hence, the first author suggested to take a norm of Σ , $\|\Sigma\|$, as a single measure of scatter in the whole configuration, such as $\|\Sigma\|_1 = \max_i \sum_j |\sigma_{ij}|$ or $\|\Sigma\|_2 = \text{largest eigenvalue of } \Sigma$ since Σ is at least positive semi-definite. Note that normality is not assumed in the above discussion.

If $S \sim W_p(m, \Sigma)$, $\Sigma > O$, what is then the distribution of Wilks' generalized variance in S , namely $|S|$, which can be referred to as the sample generalized variance? Let us determine the h -th moment of the sample generalized variance $|S|$ for an arbitrary h . This has already been discussed for real and complex matrix-variate gamma distributions in Sect. 5.4.1 and can be obtained from the normalizing constant in the Wishart density:

$$\begin{aligned} E[|S|^h] &= \frac{\int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} dS}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma|^{\frac{m}{2}}} \\ &= 2^{ph} |\Sigma|^h \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})}, \Re(\frac{m}{2} + h) > \frac{p-1}{2}. \end{aligned} \quad (5.5.13)$$

Then

$$\begin{aligned} E[(2\Sigma)^{-1}S]^h &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} = \prod_{j=1}^p \frac{\Gamma(\frac{m}{2} + h - \frac{j-1}{2})}{\Gamma(\frac{m}{2} - \frac{j-1}{2})} \\ &= E[y_1^h] E[y_2^h] \cdots E[y_p^h] \end{aligned} \quad (5.5.14)$$

where y_1, \dots, y_p are independently distributed real scalar gamma random variables with the parameters $(\frac{m}{2} - \frac{j-1}{2}, 1)$, $j = 1, \dots, p$. In the complex case

$$\begin{aligned}
 E[|\det((\tilde{\Sigma})^{-1}\tilde{S})|^h] &= \frac{\tilde{\Gamma}_p(m+h)}{\tilde{\Gamma}_p(m)} = \prod_{j=1}^p \frac{\tilde{\Gamma}(m-(j-1)+h)}{\tilde{\Gamma}(m-(j-1))} \\
 &= E[\tilde{y}_1^h] \cdots E[\tilde{y}_p^h] \tag{5.5a.9}
 \end{aligned}$$

where $\tilde{y}_1, \dots, \tilde{y}_p$ and independently distributed real scalar gamma random variables with the parameters $(m - (j - 1), 1)$, $j = 1, \dots, p$. Note that if we consider $E[|\Sigma^{-1}S|^h]$ instead of $E[|(2\Sigma)^{-1}S|^h]$ in (5.5.14), then the y_j 's are independently distributed as real chisquare random variables having $m - (j - 1)$ degrees of freedom for $j = 1, \dots, p$. This can be stated as a result.

Theorems 5.5.6, 5.5a.6. *Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$, and $|S|$ be the generalized variance associated with this Wishart matrix or the sample generalized variance in the corresponding p -variate real Gaussian population. Then, $E[|(2\Sigma)^{-1}S|^h] = E[y_1^h] \cdots E[y_p^h]$ so that $|(2\Sigma)^{-1}S|$ has the structural representation $|(2\Sigma)^{-1}S| = y_1 \cdots y_p$ where the y_j 's are independently distributed real gamma random variables with the parameters $(\frac{m}{2} - \frac{j-1}{2}, 1)$, $j = 1, \dots, p$. Equivalently, $E[|\Sigma^{-1}S|^h] = E[z_1^h] \cdots E[z_p^h]$ where the z_j 's are independently distributed real chisquare random variables having $m - (j - 1)$, $j = 1, \dots, p$, degrees of freedom. In the complex case, if we let $\tilde{S} \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$, and $|\det(\tilde{S})|$ be the generalized variance, then $|\det((\tilde{\Sigma})^{-1}\tilde{S})|$ has the structural representation $|\det((\tilde{\Sigma})^{-1}\tilde{S})| = \tilde{y}_1 \cdots \tilde{y}_p$ where the \tilde{y}_j 's are independently distributed real scalar gamma random variables with the parameters $(m - (j - 1), 1)$, $j = 1, \dots, p$ or chisquare random variables in the complex domain having $m - (j - 1)$, $j = 1, \dots, p$, degrees of freedom.*

5.5.6. Inverse Wishart distribution

When $S \sim W_p(m, \Sigma)$, $\Sigma > O$, what is then the distribution of S^{-1} ? Since S has a real matrix-variate gamma distribution, that of its inverse is directly available from the transformation $U = S^{-1}$. In light of Theorem 1.6.6, we have $dS = |U|^{-(p+1)}dU$ for the real case and $d\tilde{X} = |\det(\tilde{U}\tilde{U}^*)|^{-p}d\tilde{U}$ in the complex domain. Thus, denoting the density of U by $g(U)$, we have the following result:

Theorems 5.5.7, 5.5a.7. *Let the real Wishart matrix $S \sim W_p(m, \Sigma)$, $\Sigma > O$, and the Wishart matrix in the complex domain $\tilde{S} \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$. Let $U = S^{-1}$ and $\tilde{U} = \tilde{S}^{-1}$. Letting the density of S be denoted by $g(U)$ and that of \tilde{U} be denoted by $\tilde{g}(\tilde{U})$,*

$$g(U) = \frac{|U|^{-\frac{m}{2}-\frac{p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})|\Sigma|^{\frac{m}{2}}} e^{-\frac{1}{2}tr(\Sigma^{-1}U^{-1})}, \quad U > O, \quad \Sigma > O, \tag{5.5.15}$$

and zero elsewhere, and

$$\tilde{g}(\tilde{U}) = \frac{|\det(\tilde{U})|^{-m-p}}{\tilde{\Gamma}_p(m)|\det(\tilde{\Sigma})|^m} e^{-\text{tr}(\tilde{\Sigma}^{-1}\tilde{U}^{-1})}, \quad \tilde{U} = \tilde{U}^* > O, \quad \tilde{\Sigma} = \tilde{\Sigma}^* > O, \quad (5.5a.10)$$

and zero elsewhere.

5.5.7. Marginal distributions of a Wishart matrix

At the beginning of this chapter, we had explicitly evaluated real and complex matrix-variate gamma integrals and determined that the diagonal blocks are again real and complex matrix-variate gamma integrals. Hence, the following results are already available from the discussion on the matrix-variate gamma distribution. We will now establish the results via Laplace transforms. Let S be Wishart distributed with degrees of freedom m and parameter matrix $\Sigma > O$, that is, $S \sim W_p(m, \Sigma)$, $\Sigma > O$, $m \geq p$. Let us partition S and Σ as follows:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad (i)$$

(referred to as a 2×2 partitioning) S_{11} , Σ_{11} being $r \times r$ and S_{22} , Σ_{22} being $(p-r) \times (p-r)$ – refer to Sect. 1.3 for results on partitioned matrices. Let *T be a similarly partitioned $p \times p$ parameter matrix with ${}^*T_{11}$ being $r \times r$ where

$${}^*T = \begin{bmatrix} {}^*T_{11} & O \\ O & O \end{bmatrix}, \quad {}^*T_{11} = {}^*T'_{11} > O. \quad (ii)$$

Observe that *T is a slightly modified parameter matrix $T = (t_{ij}) = T'$ where the t_{ij} 's are weighted with $\frac{1}{2}$ for $i \neq j$ to obtain *T . Noting that $\text{tr}({}^*T'S) = \text{tr}({}^*T'_{11}S_{11})$, the Laplace transform of the Wishart density $W_p(m, \Sigma)$, $\Sigma > O$, with *T as defined above, is given by

$$|I + 2\Sigma{}^*T|^{-\frac{m}{2}} = \begin{vmatrix} I_r + 2\Sigma_{11}{}^*T_{11} & O \\ 2\Sigma_{21}{}^*T_{11} & I_{p-r} \end{vmatrix}^{-\frac{m}{2}} = |I_r + 2\Sigma_{11}{}^*T_{11}|^{-\frac{m}{2}}. \quad (5.5.16)$$

Thus, S_{11} has a Wishart distribution with m degrees of freedom and parameter matrix Σ_{11} . It can be similarly established that S_{22} is Wishart distributed with degrees of freedom m and parameter matrix Σ_{22} . Hence, the following result:

Theorems 5.5.8, 5.5a.8. *Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$. Let S and Σ be partitioned into a 2×2 partitioning as above. Then, the sub-matrices $S_{11} \sim W_r(m, \Sigma_{11})$, $\Sigma_{11} > O$, and $S_{22} \sim W_{p-r}(m, \Sigma_{22})$, $\Sigma_{22} > O$. In the complex case, let $\tilde{S} \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$. Letting \tilde{S} be partitioned as in the real case, $\tilde{S}_{11} \sim \tilde{W}_r(m, \tilde{\Sigma}_{11})$ and $\tilde{S}_{22} \sim \tilde{W}_{p-r}(m, \tilde{\Sigma}_{22})$.*

Corollaries 5.5.2, 5.5a.2. Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$. Suppose that $\Sigma_{12} = O$ in the 2×2 partitioning of Σ . Then S_{11} and S_{22} are independently distributed with $S_{11} \sim W_r(m, \Sigma_{11})$ and $S_{22} \sim W_{p-r}(m, \Sigma_{22})$. Consider a $k \times k$ partitioning of S and Σ , the order of the diagonal blocks S_{jj} and Σ_{jj} being $p_j \times p_j$, $p_1 + \dots + p_k = p$. If $\Sigma_{ij} = O$ for all $i \neq j$, then the S_{jj} 's are independently distributed as Wishart matrices on p_j components, with degrees of freedom m and parameter matrices $\Sigma_{jj} > O$, $j = 1, \dots, k$. In the complex case, consider the same type of partitioning as in the real case. Then, if $\tilde{\Sigma}_{ij} = O$ for all $i \neq j$, \tilde{S}_{jj} , $j = 1, \dots, k$, are independently distributed as $\tilde{S}_{jj} \sim \tilde{W}_{p_j}(m, \tilde{\Sigma}_{jj})$, $j = 1, \dots, k$, $p_1 + \dots + p_k = p$.

Let S be a $p \times p$ real Wishart matrix with m degrees of freedom and parameter matrix $\Sigma > O$. Consider the following 2×2 partitioning of S and Σ^{-1} :

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{11} \text{ being } r \times r, \quad \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}.$$

Then, the density, denoted by $f(S)$, can be written as

$$\begin{aligned} f(S) &= \frac{|S|^{\frac{m}{2} - \frac{p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma|^{\frac{m}{2}}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1}S)} \\ &= \frac{|S_{11}|^{\frac{m}{2} - \frac{p+1}{2}} |S_{22} - S_{21}S_{11}^{-1}S_{12}|^{\frac{m}{2} - \frac{p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma|^{\frac{m}{2}}} \\ &\quad \times e^{-\frac{1}{2} [\text{tr}(\Sigma^{11}S_{11}) + \text{tr}(\Sigma^{22}S_{22}) + \text{tr}(\Sigma^{12}S_{21}) + \text{tr}(\Sigma^{21}S_{12})]}. \end{aligned}$$

In this case, $dS = dS_{11} \wedge dS_{22} \wedge dS_{12}$. Let $U_2 = S_{22} - S_{21}S_{11}^{-1}S_{12}$. Referring to Sect. 1.3, the coefficient of S_{11} in the exponent is $\Sigma^{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$. Let $U_2 = S_{22} - S_{21}S_{11}^{-1}S_{12}$ so that $S_{22} = U_2 + S_{21}S_{11}^{-1}S_{12}$ and $dS_{22} = dU_2$ for fixed S_{11} and S_{12} . Then, the function of U_2 is of the form

$$|U_2|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{22}U_2)}.$$

However, U_2 is $(p-r) \times (p-r)$ and we can write $\frac{m}{2} - \frac{p+1}{2} = \frac{m-r}{2} - \frac{p-r+1}{2}$. Therefore $U_2 \sim W_{p-r}(m-r, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$ as $\Sigma^{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$. From symmetry, $U_1 = S_{11} - S_{12}\Sigma_{22}^{-1}S_{21} \sim W_r(m - (p-r), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. After replacing S_{22} in the exponent by $U_2 + S_{21}S_{11}^{-1}S_{12}$, the exponent, excluding $-\frac{1}{2}$, can be written as $\text{tr}[\Sigma^{11}S_{11}] + \text{tr}[\Sigma^{22}S_{12}\Sigma_{22}^{-1}S_{21}] + \text{tr}[\Sigma^{12}S_{21}] + \text{tr}[\Sigma^{21}S_{12}]$. Let us try to integrate out S_{12} . To this end, let $V = S_{11}^{-\frac{1}{2}}S_{12} \Rightarrow dS_{12} = |S_{11}|^{\frac{p-r}{2}} dV$ for fixed S_{11} . Then the determinant of S_{11} in $f(X)$

becomes $|S_{11}|^{\frac{m}{2} - \frac{p+1}{2}} \times |S_{11}|^{\frac{p-r}{2}} = |S_{11}|^{\frac{m}{2} - \frac{r+1}{2}}$. The exponent, excluding $-\frac{1}{2}$ becomes the following, denoting it by ρ :

$$\rho = \text{tr}(\Sigma^{12} V' S_{11}^{\frac{1}{2}}) + \text{tr}(\Sigma^{21} S_{11}^{\frac{1}{2}} V) + \text{tr}(\Sigma^{22} V' V). \quad (i)$$

Note that $\text{tr}(\Sigma^{22} V' V) = \text{tr}(V \Sigma^{22} V')$ and

$$(V + C) \Sigma^{22} (V + C)' = V \Sigma^{22} V' + V \Sigma^{22} C' + C \Sigma^{22} V + C \Sigma^{22} C'. \quad (ii)$$

On comparing (i) and (ii), we have $C' = (\Sigma^{22})^{-1} \Sigma^{21} S_{11}^{\frac{1}{2}}$. Substituting for C and C' in ρ , the term containing S_{11} in the exponent becomes $-\frac{1}{2} \text{tr}(S_{11} (\Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21})) = -\frac{1}{2} \text{tr}(S_{11} \Sigma_{11}^{-1})$. Collecting the factors containing S_{11} , we have $S_{11} \sim W_r(m, \Sigma_{11})$ and from symmetry, $S_{22} \sim W_{p-r}(m, \Sigma_{22})$. Since the density $f(S)$ splits into a function of U_2 , S_{11} and $S_{11}^{-\frac{1}{2}} S_{12}$, these quantities are independently distributed. Similarly, U_1 , S_{22} and $S_{22}^{-\frac{1}{2}} S_{21}$ are independently distributed. The exponent of $|U_1|$ is $\frac{m}{2} - \frac{p+1}{2} = (\frac{m}{2} - \frac{p-r}{2}) - \frac{r+1}{2}$. Observing that U_1 is $r \times r$, we have the density of $U_1 = S_{11} - S_{12} S_{22}^{-1} S_{21}$ as a real Wishart density on r components, with degrees of freedom $m - (p - r)$ and parameter matrix $\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ whose the density, denoted by $f_1(U)$, is the following:

$$f_1(U_1) = \frac{|U_1|^{\frac{m-(p-r)}{2} - \frac{r+1}{2}}}{2^{\frac{r(m-(p-r))}{2}} \Gamma_r(\frac{m-(p-r)}{2}) |\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}|^{\frac{m-(p-r)}{2}}} e^{-\frac{1}{2} \text{tr}[U_1 (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1}]}. \quad (iii)$$

A similar expression can be obtained for the density of U_2 . Thus, the following result:

Theorems 5.5.9, 5.5a.9. *Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$, $m \geq p$. Consider the 2×2 partitioning of S as specified above, S_{11} being $r \times r$. Let $U_1 = S_{11} - S_{12} S_{22}^{-1} S_{21}$. Then,*

$$U_1 \sim W_r(m - (p - r), \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}). \quad (5.5.17)$$

In the complex case, let $\tilde{S} \sim \tilde{W}_p(m, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^ > O$. Consider the same partitioning as in the real case and let \tilde{S}_{11} be $r \times r$. Then, letting $\tilde{U}_1 = \tilde{S}_{11} - \tilde{S}_{12} \tilde{S}_{22}^{-1} \tilde{S}_{21}$, \tilde{U}_1 is Wishart distributed as*

$$\tilde{U}_1 \sim \tilde{W}_r(m - (p - r), \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12} \tilde{\Sigma}_{22}^{-1} \tilde{\Sigma}_{21}). \quad (5.5a.11)$$

A similar density is obtained for $\tilde{U}_2 = \tilde{S}_{22} - \tilde{S}_{21} \tilde{S}_{11}^{-1} \tilde{S}_{12}$.

Example 5.5.1. Let the 3×3 matrix $S \sim W_3(5, \Sigma)$, $\Sigma > O$. Determine the distributions of $Y_1 = S_{22} - S_{21}S_{11}^{-1}S_{12}$, $Y_2 = S_{22}$ and $Y_3 = S_{11}$ where

$$\Sigma = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

with S_{11} being 2×2 .

Solution 5.5.1. Let the densities of Y_j be denoted by $f_j(Y_j)$, $j = 1, 2, 3$. We need the following matrix, denoted by B :

$$\begin{aligned} B &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 3 - [0, 1] \begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 3 - \frac{2}{5} = \frac{13}{5}. \end{aligned}$$

From our usual notations, $Y_1 \sim W_{p-r}(m-r, B)$. Observing that Y_1 is a real scalar, we denote it by y_1 , its density being given by

$$\begin{aligned} f_1(y_1) &= \frac{y_1^{\frac{m-r}{2} - \frac{(p-r)+1}{2}}}{2^{\frac{(m-r)(p-r)}{2}} |B|^{\frac{m-r}{2}} \Gamma(\frac{m-r}{2})} e^{-\frac{1}{2} \text{tr}(B^{-1}y_1)} \\ &= \frac{y_1^{\frac{3}{2}-1} e^{-\frac{5}{26}y_1}}{2^{\frac{3}{2}} (13/5)^{\frac{3}{2}} \Gamma(\frac{3}{2})}, \quad 0 \leq y_1 < \infty, \end{aligned}$$

and zero elsewhere. Now, consider Y_2 which is also a real scalar that will be denoted by y_2 . As per our notation, $Y_2 = S_{22} \sim W_{p-r}(m, \Sigma_{22})$. Its density is then as follows, observing that $\Sigma_{22} = (3)$, $|\Sigma_{22}| = 3$ and $\Sigma_{22}^{-1} = (\frac{1}{3})$:

$$\begin{aligned} f_2(y_2) &= \frac{y_2^{\frac{m}{2} - \frac{(p-r)+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma_{22}^{-1}y_2)}}{2^{\frac{m(p-r)}{2}} \Gamma_{p-r}(\frac{m}{2}) |\Sigma_{22}|^{\frac{m}{2}}} \\ &= \frac{y_2^{\frac{5}{2}-1} e^{-\frac{1}{6}y_2}}{2^{\frac{5}{2}} \Gamma(\frac{5}{2}) 3^{\frac{5}{2}}}, \quad 0 \leq y_2 < \infty, \end{aligned}$$

and zero elsewhere. Note that $Y_3 = S_{11}$ is 2×2 . With our usual notations, $p = 3$, $r = 2$, $m = 5$ and $|\Sigma_{11}| = 5$; as well,

$$S_{11} = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}, \quad \Sigma_{11}^{-1} = \frac{1}{5} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \quad \text{tr}(\Sigma_{11}^{-1}S_{11}) = \frac{1}{5}[3s_{11} + 2s_{12} + 2s_{22}].$$

Thus, the density of Y_3 is

$$\begin{aligned} f_3(Y_3) &= \frac{|S_{11}|^{\frac{m}{2} - \frac{r+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma_{11}^{-1} S_{11})}}{2^{\frac{mr}{2}} \Gamma_r(\frac{m}{2}) |S_{11}|^{\frac{m}{2}}} \\ &= \frac{[s_{11}s_{22} - s_{12}^2] e^{-\frac{1}{10}(3s_{11} + 2s_{12} + 2s_{22})}}{(3)(2^3)(5)^{\frac{5}{2}} \pi}, \quad Y_3 > O, \end{aligned}$$

and zero elsewhere. This completes the calculations.

Example 5.5a.1. Let the 3×3 Hermitian positive definite matrix \tilde{S} have a complex Wishart density with degrees of freedom $m = 5$ and parameter matrix $\Sigma > O$. Determine the densities of $\tilde{Y}_1 = \tilde{S}_{22} - \tilde{S}_{21}\tilde{S}_{11}^{-1}\tilde{S}_{12}$, $\tilde{Y}_2 = \tilde{S}_{22}$ and $\tilde{Y}_3 = \tilde{S}_{11}$ where

$$\tilde{S} = \begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} \\ \tilde{S}_{21} & \tilde{S}_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 3 & -i & 0 \\ i & 2 & i \\ 0 & -i & 2 \end{bmatrix}$$

with S_{11} and Σ_{11} being 2×2 .

Solution 5.5a.1. Observe that Σ is Hermitian positive definite. We need the following numerical results:

$$B \equiv \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 2 - [0, -i] \begin{bmatrix} 1 \\ 5 \end{bmatrix} \begin{bmatrix} 2 & i \\ -i & 3 \end{bmatrix} \begin{bmatrix} 0 \\ i \end{bmatrix} = 2 - \frac{3}{5} = \frac{7}{5};$$

$$B^{-1} = \frac{5}{7}, \quad |B| = \frac{7}{5}, \quad \Sigma_{11}^{-1} = \frac{1}{5} \begin{bmatrix} 2 & i \\ -i & 3 \end{bmatrix}.$$

Note that \tilde{Y}_1 and \tilde{Y}_2 are real scalar quantities which will be denoted as y_1 and y_2 , respectively. Let the densities of y_1 and y_2 be $f_j(y_j)$, $j = 1, 2$. Then, with our usual notations, $f_1(y_1)$ is

$$\begin{aligned} f_1(y_1) &= \frac{|\det(\tilde{y}_1)|^{(m-r)-(p-r)} e^{-\text{tr}(B^{-1}\tilde{y}_1)}}{|\det(B)|^{m-r} \tilde{\Gamma}_{p-r}(m-r)} \\ &= \frac{y_1^2 e^{-\frac{5}{7}y_1}}{(\frac{7}{5})^3 \Gamma(3)}, \quad 0 \leq y_1 < \infty, \end{aligned}$$

and zero elsewhere, and the density of \tilde{y}_2 , is

$$\begin{aligned} f_2(y_2) &= \frac{|\det(\tilde{y}_2)|^{m-(p-r)} e^{-\text{tr}(\Sigma_{22}^{-1} \tilde{y}_2)}}{|\det(\Sigma_{22})|^m \tilde{\Gamma}_{p-r}(m)} \\ &= \frac{y_2^4 e^{-\frac{1}{2}y_2}}{2^5 \Gamma(5)}, \quad 0 \leq y_2 < \infty, \end{aligned}$$

and zero elsewhere. Note that $\tilde{Y}_3 = \tilde{S}_{11}$ is 2×2 . Letting

$$\tilde{S}_{11} = \begin{bmatrix} s_{11} & \tilde{s}_{12} \\ \tilde{s}_{12}^* & s_{22} \end{bmatrix}, \quad \text{tr}(\Sigma_{11}^{-1} \tilde{S}_{11}) = \frac{1}{5}[2s_{11} + 3s_{22} + i\tilde{s}_{12}^* - i\tilde{s}_{12}]$$

and $|\det(\tilde{Y}_3)| = [s_{11}s_{22} - \tilde{s}_{12}^*\tilde{s}_{12}]$. With our usual notations, the density of \tilde{Y}_3 , denoted by $\tilde{f}_3(\tilde{Y}_3)$, is the following:

$$\begin{aligned} \tilde{f}_3(\tilde{Y}_3) &= \frac{|\det(\tilde{Y}_3)|^{m-r} e^{-\text{tr}(\Sigma_{11}^{-1} \tilde{Y}_3)}}{|\det(\Sigma_{11})|^m \tilde{\Gamma}_r(m)} \\ &= \frac{[s_{11}s_{22} - \tilde{s}_{12}\tilde{s}_{12}^*]^3 e^{-\frac{1}{5}[2s_{11}+3s_{22}+i\tilde{s}_{12}^*-i\tilde{s}_{12}]}}{5^5 \tilde{\Gamma}_2(5)}, \quad \tilde{Y}_3 > O, \end{aligned}$$

and zero elsewhere, where $5^5 \tilde{\Gamma}_2(5) = 3125(144)\pi$. This completes the computations.

5.5.8. Connections to geometrical probability problems

Consider the representation of the Wishart matrix $S = Z_{n-1}Z'_{n-1}$ given in (5.5.8) where the p rows are linearly independent $1 \times (n - 1)$ vectors. Then, these p linearly independent rows, taken in order, form a convex hull and determine a p -parallelotope in that hull, which is determined by the p points in the $(n - 1)$ -dimensional Euclidean space, $n - 1 \geq p$. Then, as explained in Mathai (1999), the volume content of this parallelotope is $v = |Z_{n-1}Z'_{n-1}|^{\frac{1}{2}} = |S|^{\frac{1}{2}}$, where $S \sim W_p(n - 1, \Sigma)$, $\Sigma > O$. Thus, the volume content of this parallelotope is the positive square root of the generalized variance $|S|$. The distributions of this random volume when the p random points are uniformly, type-1 beta, type-2 beta and gamma distributed are provided in Chap. 4 of Mathai (1999).

5.6. The Distribution of the Sample Correlation Coefficient

Consider the real Wishart density or matrix-variate gamma in (5.5.10) for $p = 2$. For convenience, let us take the degrees of freedom parameter $n - 1 = m$. Then for $p = 2$, $f(S)$ in (5.5.10), denoted by $f_2(S)$, is the following, observing that $|S| = s_{11}s_{22}(1 - r^2)$ where r is the sample correlation coefficient:

$$f_2(S) = f_2(s_{11}, s_{22}, r) = \frac{[s_{11}s_{22}(1-r^2)]^{\frac{m}{2}-\frac{3}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1}S)}}{2^m [\sigma_{11}\sigma_{22}(1-\rho^2)]^{\frac{m}{2}} \Gamma_2(\frac{m}{2})} \quad (5.6.1)$$

where ρ = the population correlation coefficient, $|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho^2)$, $\Gamma_2(\frac{m}{2}) = \pi^{\frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m-1}{2})$, $-1 < \rho < 1$,

$$\begin{aligned} \Sigma^{-1} &= \frac{1}{|\Sigma|} \text{Cof}(\Sigma) = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho^2)} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \\ &= \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_{11}} & -\frac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}} \\ -\frac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}} & \frac{1}{\sigma_{22}} \end{bmatrix}, \quad \sigma_{12} = \rho \sqrt{\sigma_{11}\sigma_{22}}, \end{aligned} \quad (i)$$

$$\begin{aligned} \text{tr}(\Sigma^{-1}S) &= \frac{1}{1-\rho^2} \left\{ \frac{s_{11}}{\sigma_{11}} - 2\rho \frac{s_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{s_{22}}{\sigma_{22}} \right\} \\ &= \frac{1}{1-\rho^2} \left\{ \frac{s_{11}}{\sigma_{11}} - 2\rho r \frac{\sqrt{s_{11}s_{22}}}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{s_{22}}{\sigma_{22}} \right\}. \end{aligned} \quad (ii)$$

Let us make the substitution $x_1 = \frac{s_{11}}{\sigma_{11}}$, $x_2 = \frac{s_{22}}{\sigma_{22}}$. Note that $dS = ds_{11} \wedge ds_{22} \wedge ds_{12}$. But $ds_{12} = \sqrt{s_{11}s_{22}} dr$ for fixed s_{11} and s_{22} . In order to obtain the density of r , we must integrate out x_1 and x_2 , observing that $\sqrt{s_{11}s_{22}}$ is coming from ds_{12} :

$$\begin{aligned} \int_{s_{11}, s_{22}} f_2(S) ds_{11} \wedge ds_{22} &= \int_{x_1 > 0} \int_{x_2 > 0} \frac{1}{2^m (\sigma_{11}\sigma_{22})^{\frac{m}{2}} (1-\rho^2)^{\frac{m}{2}} \pi^{\frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m-1}{2})} \\ &\quad \times (1-r^2)^{\frac{m-3}{2}} (\sigma_{11}\sigma_{22}x_1x_2)^{\frac{m}{2}-1} e^{-\frac{1}{2(1-\rho^2)}\{x_1-2r\rho\sqrt{x_1x_2}+x_2\}} \sigma_{11}\sigma_{22} dx_1 \wedge dx_2. \end{aligned} \quad (iii)$$

For convenience, let us expand

$$e^{-\frac{1}{2(1-\rho^2)}(-2r\rho\sqrt{x_1x_2})} = \sum_{k=0}^{\infty} \left(\frac{r\rho}{1-\rho^2} \right)^k \frac{x_1^{\frac{k}{2}} x_2^{\frac{k}{2}}}{k!}. \quad (iv)$$

Then the part containing x_1 gives the integral

$$\int_{x_1=0}^{\infty} x_1^{\frac{m}{2}-1+\frac{k}{2}} e^{-\frac{x_1}{2(1-\rho^2)}} dx_1 = [2(1-\rho^2)]^{\frac{m}{2}+\frac{k}{2}} \Gamma\left(\frac{m}{2} + \frac{k}{2}\right), \quad m \geq 2. \quad (v)$$

By symmetry, the integral over x_2 gives $[2(1-\rho^2)]^{\frac{m}{2}+\frac{k}{2}} \Gamma(\frac{m+k}{2})$, $m \geq 2$. Collecting all the constants we have

$$\frac{(\sigma_{11}\sigma_{22})^{\frac{m}{2}} 2^{m+k} (1-\rho^2)^{m+k} \Gamma^2(\frac{m+k}{2})}{2^m (\sigma_{11}\sigma_{22})^{\frac{m}{2}} (1-\rho^2)^{\frac{m}{2}} \pi^{\frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m-1}{2})} = \frac{(1-\rho^2)^{\frac{m}{2}+k} 2^k \Gamma^2(\frac{m+k}{2})}{\pi^{\frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m-1}{2})}. \quad (vi)$$

We can simplify $\Gamma(\frac{m}{2})\Gamma(\frac{m}{2}-\frac{1}{2})$ by using the duplication formula for gamma functions, namely

$$\Gamma(2z) = \pi^{-\frac{1}{2}} 2^{2z-1} \Gamma(z)\Gamma\left(z + \frac{1}{2}\right), \quad z = \frac{m-1}{2}, \quad (5.6.2)$$

Then,

$$\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{m-1}{2}\right) = \frac{\Gamma(m-1)\pi^{\frac{1}{2}}}{2^{m-2}}. \quad (vii)$$

Hence the density of r , denoted by $f_r(r)$, is the following:

$$f_r(r) = \frac{2^{m-2}(1-\rho^2)^{\frac{m}{2}}}{\Gamma(m-1)\pi} (1-r^2)^{\frac{1}{2}(m-3)} \sum_{k=0}^{\infty} \frac{(2r\rho)^k}{k!} \Gamma^2\left(\frac{m+k}{2}\right), \quad -1 \leq r \leq 1, \quad (5.6.3)$$

and zero elsewhere, $m = n - 1$, n being the sample size.

5.6.1. The special case $\rho = 0$

In this case, (5.6.3) becomes

$$f_r(r) = \frac{2^{m-2}\Gamma^2(\frac{m}{2})}{\Gamma(m-1)\pi} (1-r^2)^{\frac{m-1}{2}-1}, \quad -1 \leq r \leq 1, \quad m = n - 1 \quad (5.6.4)$$

$$= \frac{\Gamma(\frac{m}{2})}{\sqrt{\pi}\Gamma(\frac{m-1}{2})} (1-r^2)^{\frac{m-1}{2}-1}, \quad -1 \leq r \leq 1, \quad (5.6.5)$$

zero elsewhere, $m = n - 1 \geq 2$, n being the sample size. The simplification is made by using the duplication formula and writing $\Gamma(m-1) = \pi^{-\frac{1}{2}} 2^{m-2} \Gamma(\frac{m-1}{2})\Gamma(\frac{m}{2})$. For testing the hypothesis $H_o : \rho = 0$, the test statistic is r and the null distribution, that is, the distribution under the null hypothesis H_o is given in (5.6.5). Numerical tables of percentage points obtained from (5.6.5) are available. If $\rho \neq 0$, the non-null distribution is available from (5.6.3); so, if we wish to test the hypothesis $H_o : \rho = \rho_o$ where ρ_o is a given quantity, we can compute the percentage points from (5.6.3). It can be shown from (5.6.5) that for $\rho = 0$, $t_m = \sqrt{m} \frac{r}{\sqrt{1-r^2}}$ is distributed as a Student- t with m degrees of freedom, and hence for testing $H_o : \rho = 0$ against $H_1 : \rho \neq 0$, the null hypothesis can be rejected if $|t_m| = \sqrt{m} \left| \frac{r}{\sqrt{1-r^2}} \right| \geq t_{m, \frac{\alpha}{2}}$ where $Pr\{|t_m| \geq t_{m, \frac{\alpha}{2}}\} = \alpha$. For tests that make use of the Student- t statistic, refer to Mathai and Haubold (2017b). Since the density given in (5.6.5) is an even function, when $\rho = 0$, all odd order moments are equal to zero and the even order moments can easily be evaluated from type-1 beta integrals.

5.6.2. The multiple and partial correlation coefficients

Let the $p \times 1$ real vector X_j with $X_j' = (x_{1j}, \dots, x_{pj})$ have a p -variate distribution whose mean value $E(X_j) = \mu$ and covariance matrix $\text{Cov}(X_j) = \Sigma$, $\Sigma > O$, where μ is $p \times 1$ and Σ is $p \times p$, for $j = 1, \dots, n$, the X_j 's being iid (independently and identically distributed). Consider the following partitioning of Σ :

$$\Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \sigma_{11} > 0 \text{ is } 1 \times 1, \quad \Sigma_{22} > O \text{ is } (p-1) \times (p-1), \quad \Sigma'_{12} = \Sigma_{21}.$$

Let

$$\rho_{1.(2\dots p)}^2 = \frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11}}. \quad (5.6.6)$$

Then, $\rho_{1.(2\dots p)}$ is called the *multiple correlation coefficient of x_{1j} on x_{2j}, \dots, x_{pj}* . The sample value corresponding to $\rho_{1.(2\dots p)}^2$ which is denoted by $r_{1.(2\dots p)}^2$ and referred to as the square of the sample multiple correlation coefficient, is given by

$$r_{1.(2\dots p)}^2 = \frac{S_{12} S_{22}^{-1} S_{21}}{s_{11}} \quad \text{with } S = \begin{bmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \quad (5.6.7)$$

where s_{11} is 1×1 , S_{22} is $(p-1) \times (p-1)$, $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$, $\mathbf{X} = (X_1, \dots, X_n)$ is the $p \times n$ sample matrix, n being the sample size, $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ is $p \times n$, the X_j 's, $j = 1, \dots, n$, being iid according to a given p -variate population having mean value vector μ and covariance matrix $\Sigma > O$, which need not be Gaussian.

5.6.3. Different derivations of $\rho_{1.(2\dots p)}$

Consider a prediction problem involving real scalar variables where x_1 is predicted by making use of x_2, \dots, x_p or linear functions thereof. Let $A_2' = (a_2, \dots, a_p)$ be a constant vector where a_j , $j = 2, \dots, p$ are real scalar constants. Letting $X'_{(2)} = (x_2, \dots, x_p)$, a linear function of $X_{(2)}$ is $u = A_2' X_{(2)} = a_2 x_2 + \dots + a_p x_p$. Then, the mean value and variance of this linear function are $E[u] = E[A_2' X_{(2)}] = A_2' \mu_{(2)}$ and $\text{Var}(u) = \text{Var}(A_2' X_{(2)}) = A_2' \Sigma_{22} A_2$ where $\mu'_{(2)} = (\mu_2, \dots, \mu_p) = E[X_{(2)}]$ and Σ_{22} is the covariance matrix associated with $X_{(2)}$, which is available from the partitioning of Σ specified in the previous subsection. Let us determine the correlation between x_1 , the variable being predicted, and u , a linear function of the variables being utilized to predict x_1 , denoted by $\rho_{1,u}$, that is,

$$\rho_{1,u} = \frac{\text{Cov}(x_1, u)}{\sqrt{\text{Var}(x_1) \text{Var}(u)}},$$

where $\text{Cov}(x_1, u) = E[(x_1 - E(x_1))(u - E(u))] = E[(x_1 - E(x_1))(X_{(2)} - E(X_{(2)}))' A_2] = \text{Cov}(x_1, X_{(2)}) A_2 = \Sigma_{12} A_2$, $\text{Var}(x_1) = \sigma_{11}$, $\text{Var}(u) = A_2' \Sigma_{22} A_2 > O$. Letting $\Sigma_{22}^{-\frac{1}{2}}$ be the positive definite square root of Σ_{22} , we can write $\Sigma_{12} A_2 = (\Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}) (\Sigma_{22}^{\frac{1}{2}} A_2)$. Then, on applying Cauchy-Schwartz' inequality, we may write $\Sigma_{12} A_2 = (\Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}) (\Sigma_{22}^{\frac{1}{2}} A_2) \leq \sqrt{(\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})(A_2' \Sigma_{22} A_2)}$. Thus,

$$\begin{aligned} \rho_{1,u} &\leq \frac{\sqrt{(\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})(A_2' \Sigma_{22} A_2)}}{\sqrt{(\sigma_{11})(A_2' \Sigma_{22} A_2)}} = \frac{\sqrt{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}}{\sqrt{\sigma_{11}}}, \text{ that is,} \\ \rho_{1,u}^2 &\leq \frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11}} = \rho_{1.(2\dots p)}^2. \end{aligned} \tag{5.6.8}$$

This establishes the following result:

Theorem 5.6.1. *The multiple correlation coefficient $\rho_{1.(2\dots p)}$ of x_1 on x_2, \dots, x_p represents the maximum correlation between x_1 and an arbitrary linear function of x_2, \dots, x_p .*

This shows that if we consider the joint variation of x_1 and (x_2, \dots, x_p) , this scale-free joint variation, namely the correlation, is maximum when the scale-free covariance, which constitutes a scale-free measure of joint variation, is the multiple correlation coefficient. Correlation measures a scale-free joint scatter in the variables involved, in this case x_1 and (x_2, \dots, x_p) . Correlation does not measure general relationships between the variables; counterexamples are provided in Mathai and Haubold (2017b). Hence “maximum correlation” should be interpreted as maximum joint scale-free variation or joint scatter in the variables.

For the next property, we will use the following two basic results on conditional expectations, referring also to Mathai and Haubold (2017b). Let x and y be two real scalar random variables having a joint distribution. Then,

$$E[y] = E[E(y|x)] \tag{i}$$

whenever the expected values exist, where the inside expectation is taken in the conditional space of y , given x , for all x , that is, $E_{y|x}(y|x)$, and the outside expectation is taken in the marginal space of x , that is $E_x(x)$. The other result states that

$$\text{Var}(y) = \text{Var}(E[y|x]) + E[\text{Var}(y|x)] \tag{ii}$$

where it is assumed that the expected value of the conditional variance and the variance of the conditional expectation exist. Situations where the results stated in

(i) and (ii) are applicable or not applicable are described and illustrated in Mathai and Haubold (2017b). In result (i), x can be a scalar, vector or matrix variable. Now, let us examine the problem of predicting x_1 on the basis of x_2, \dots, x_p . What is the “best” predictor function of x_2, \dots, x_p for predicting x_1 , “best” being construed as *in the minimum mean square sense*. If $\phi(x_2, \dots, x_p)$ is an arbitrary predictor, then at given values of x_2, \dots, x_p , ϕ is a constant. Consider the squared distance $(x_1 - b)^2$ between x_1 and $b = \phi(x_2, \dots, x_p|x_2, \dots, x_p)$ or b is ϕ at given values of x_2, \dots, x_p . Then, “minimum in the mean square sense” means to minimize the expected value of $(x_1 - b)^2$ over all b or $\min E(x_1 - b)^2$. We have already established in Mathai and Haubold (2017b) that the minimizing value of b is $b = E[x_1]$ at given x_2, \dots, x_p or the conditional expectation of x_1 , given x_2, \dots, x_p or $b = E[x_1|x_2, \dots, x_p]$. Hence, this “best” predictor is also called *the regression of x_1 on (x_2, \dots, x_p)* or $E[x_1|x_2, \dots, x_p]$ = the regression of x_1 on x_2, \dots, x_p , or the best predictor of x_1 based on x_2, \dots, x_p . Note that, in general, for any scalar variable y and a constant a ,

$$\begin{aligned} E[y - a]^2 &= E[y - E(y) + E(y) - a]^2 = E[(y - E(y))^2 - 2E[(y - E(y))(E(y) - a)] \\ &\quad + E[(E(y) - a)^2] = \text{Var}(y) + 0 + [E(y) - a]^2. \end{aligned} \quad (iii)$$

As the only term on the right-hand side containing a is $[E(y) - a]^2$, the minimum is attained when this term is zero since it is a non-negative constant, zero occurring when $a = E[y]$. Thus, $E[y - a]^2$ is minimized when $a = E[y]$. If $a = \phi(X_{(2)})$ at given value of $X_{(2)}$, then the best predictor of x_1 , based on $X_{(2)}$ is $E[x_1|X_{(2)}]$ or the regression of x_1 on $X_{(2)}$. Let us determine what happens when $E[x_1|X_{(2)}]$ is a linear function in $X_{(2)}$. Let the linear function be $b_0 + b_2x_2 + \dots + b_px_p = b_0 + B'_2X_{(2)}$, $B'_2 = (b_2, \dots, b_p)$, where b_0, b_2, \dots, b_p are real constants [Note that only real variables and real constants are considered in this section]. That is, for some constant b_0 ,

$$E[x_1|X_{(2)}] = b_0 + b_2x_2 + \dots + b_px_p. \quad (iv)$$

Taking expectation with respect to x_1, x_2, \dots, x_p in (iv), it follows from (i) that the left-hand side becomes $E[x_1]$, the right side being $b_0 + b_2E[x_2] + \dots + b_pE[x_p]$; subtracting this from (iv), we have

$$E[x_1|X_{(2)}] - E[x_1] = b_2(x_2 - E[x_2]) + \dots + b_p(x_p - E[x_p]). \quad (v)$$

Multiplying both sides of (v) by $x_j - E[x_j]$ and taking expectations throughout, the right-hand side becomes $b_2\sigma_{2j} + \dots + b_p\sigma_{pj}$ where $\sigma_{ij} = \text{Cov}(x_i, x_j)$, $i \neq j$, and it is the variance of x_j when $i = j$. The left-hand side is $E[(x_j - E(x_j))(E[x_1|x_{(2)}] - E(x_1))] = E[E(x_1x_j|X_{(2)})] - E(x_j)E(x_1) = E[x_1x_j] - E(x_1)E(x_j) = \text{Cov}(x_1, x_j)$. Three properties were utilized in the derivation, namely (i), the fact that $\text{Cov}(u, v) = E[(u - E(u))(v - E(v))] = E[u(v - E(v))] = E[v(u - E(u))]$ and $\text{Cov}(u, v) = E(uv) - E(u)E(v)$. As well, $\text{Var}(u) = E[u - E(u)]^2 = E[u(u - E(u))]$ as long as the

second order moments exist. Thus, we have the following by combining all the linear equations for $j = 2, \dots, p$:

$$\Sigma_{21} = \Sigma_{22}b \Rightarrow b = \Sigma_{22}^{-1}\Sigma_{21} \text{ or } b' = \Sigma_{12}\Sigma_{22}^{-1} \quad (5.6.9)$$

when Σ_{22} is nonsingular, which is the case as it was assumed that $\Sigma_{22} > O$. Now, the best predictor of x_1 based on a linear function of $X_{(2)}$ or the best predictor in the class of all linear functions of $X_{(2)}$ is

$$E[x_1|X_{(2)}] = b'X_{(2)} = \Sigma_{12}\Sigma_{22}^{-1}X_{(2)}. \quad (5.6.10)$$

Let us consider the correlation between x_1 and its best linear predictor based on $X_{(2)}$ or the correlation between x_1 and the linear regression of x_1 on $X_{(2)}$. Observe that $\text{Cov}(x_1, \Sigma_{12}\Sigma_{22}^{-1}X_{(2)}) = \Sigma_{12}\Sigma_{22}^{-1}\text{Cov}(X_{(2)}, x_1) = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $\Sigma_{21} = \Sigma'_{12}$. Consider the variance of the best linear predictor: $\text{Var}(b'X_{(2)}) = b'\text{Cov}(X_{(2)})b = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Thus, the square of the correlation between x_1 and its best linear predictor or the linear regression on $X_{(2)}$, denoted by $\rho_{x_1, b'X_{(2)}}^2$, is the following:

$$\rho_{x_1, b'X_{(2)}}^2 = \frac{[\text{Cov}(x_1, b'X_{(2)})]^2}{\text{Var}(x_1)\text{Var}(b'X_{(2)})} = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11}} = \rho_{1.(2\dots p)}^2. \quad (5.6.11)$$

Hence, the following result:

Theorem 5.6.2. *The multiple correlation $\rho_{1.(2\dots p)}$ between x_1 and x_2, \dots, x_p is also the correlation between x_1 and its best linear predictor or x_1 and its linear regression on x_2, \dots, x_p .*

Observe that normality has not been assumed for obtaining all of the above properties. Thus, the results hold for any population for which moments of order two exist. However, in the case of a nonsingular normal population, that is, $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, it follows from equation (3.3.5), that for $r = 1$, $E[x_1|X_{(2)}] = \Sigma_{12}\Sigma_{22}^{-1}X_{(2)}$ when $E[X_{(2)}] = \mu_{(2)} = O$ and $E(x_1) = \mu_1 = 0$; otherwise, $E[x_1|X_{(2)}] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_{(2)} - \mu_{(2)})$.

5.6.4. Distributional aspects of the sample multiple correlation coefficient

From (5.6.7), we have

$$1 - r_{1.(2\dots p)}^2 = 1 - \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}} = \frac{s_{11} - S_{12}S_{22}^{-1}S_{21}}{s_{11}} = \frac{|S|}{|S_{22}s_{11}|}, \quad (5.6.12)$$

which can be established from the expansion of the determinant $|S| = |S_{22}| |S_{11} - S_{12}S_{22}^{-1}S_{21}|$, which is available from Sect. 1.3. In our case S_{11} is 1×1 and hence we denote it as s_{11} and then $|S_{11} - S_{12}S_{22}^{-1}S_{21}|$ is $s_{11} - S_{12}S_{22}^{-1}S_{21}$ which is 1×1 . Let $u = 1 - r_{1.(2\dots p)}^2 = \frac{|S|}{|S_{22}|s_{11}}$. We can compute arbitrary moments of u by integrating out over the density of S , namely the Wishart density with $m = n - 1$ degrees of freedom when the population is Gaussian, where n is the sample size. That is, for arbitrary h ,

$$E[u^h] = \frac{1}{2^{\frac{mp}{2}} |\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} \int_{S>O} u^h |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1}S)} dS. \quad (i)$$

Note that $u^h = |S|^h |S_{22}|^{-h} s_{11}^{-h}$. Among the three factors $|S|^h$, $|S_{22}|^{-h}$ and s_{11}^{-h} , $|S_{22}|^{-h}$ and s_{11}^{-h} are creating problems. We will replace these by equivalent integrals so that the problematic part be shifted to the exponent. Consider the identities

$$s_{11}^{-h} = \frac{1}{\Gamma(h)} \int_{x=0}^{\infty} x^{h-1} e^{-s_{11}x} dx, \quad x > 0, \quad s_{11} > 0, \quad \Re(h) > 0 \quad (ii)$$

$$|S_{22}|^{-h} = \frac{1}{\Gamma_{p_2}(h)} \int_{X_2>O} |X_2|^{h - \frac{p_2+1}{2}} e^{-\text{tr}(S_{22}X_2)} dX_2, \quad (iii)$$

for $X_2 > O$, $S_{22} > O$, $\Re(h) > \frac{p_2-1}{2}$ where $X_2 > O$ is a $p_2 \times p_2$ real positive definite matrix, $p_2 = p - 1$, $p_1 = 1$, $p_1 + p_2 = p$. Then, excluding $-\frac{1}{2}$, the exponent in (i) becomes the following:

$$\text{tr}(\Sigma^{-1}S) + 2s_{11}x + 2\text{tr}(S_{22}X_2) = \text{tr}[S(\Sigma^{-1} + 2Z)], \quad Z = \begin{bmatrix} x & O \\ O & X_2 \end{bmatrix}. \quad (iv)$$

Noting that $(\Sigma^{-1} + 2Z) = \Sigma^{-1}(I + 2\Sigma Z)$, we are now in a position to integrate out S from (i) by using a real matrix-variate gamma integral, denoting the constant part in (i) as c_1 :

$$\begin{aligned} E[u^h] &= c_1 \frac{1}{\Gamma(h)\Gamma_{p_2}(h)} \int_{x=0}^{\infty} x^{h-1} \int_{X_2>O} |X_2|^{h - \frac{p_2+1}{2}} \\ &\quad \times \left[\int_{S>O} |S|^{\frac{m}{2} + h - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}[S(\Sigma^{-1} + 2Z)]} dS \right] dx \wedge dX_2 \\ &= c_1 2^{p(\frac{m}{2} + h)} \Gamma_p(m/2 + h) \int_{x=0}^{\infty} \int_{X_2>O} x^{h-1} |X_2|^{h - \frac{p_2+1}{2}} |\Sigma^{-1} + 2Z|^{-(\frac{m}{2} + h)} dx \wedge dX_2 \\ &= \frac{c_1 2^{p(\frac{m}{2} + h)}}{\Gamma(h)\Gamma_{p_2}(h)} \Gamma_p(m/2 + h) |\Sigma|^{\frac{m}{2} + h} \int_{x=0}^{\infty} \int_{X_2>O} x^{h-1} |X_2|^{h - \frac{p_2+1}{2}} \\ &\quad \times |I + 2\Sigma Z|^{-(\frac{m}{2} + h)} dx \wedge dX_2. \quad (5.6.13) \end{aligned}$$

The integral in (5.6.13) can be evaluated for a general Σ , which will produce the non-null density of $1 - r_{1.(2...p)}^2$ —non-null in the sense that the population multiple correlation $\rho_{1.(2...p)} \neq 0$. However, if $\rho_{1.(2...p)} = 0$, which we call the null case, the determinant part in (5.6.13) splits into two factors, one depending only on x and the other only involving X_2 . So letting $H_o: \rho_{1.(2...p)} = 0$,

$$E[u^h | H_o] = \frac{c_1 2^{p(\frac{m}{2}+h)}}{\Gamma(h)\Gamma_{p_2}(h)} \Gamma_p(m/2 + h) |\Sigma|^{\frac{m}{2}+h} \int_0^\infty x^{h-1} [1 + 2\sigma_{11}x]^{-(\frac{m}{2}+h)} dx$$

$$\times \int_{X_2 > 0} |X_2|^{h-\frac{p_2+1}{2}} |I + 2\Sigma_{22}X_2|^{-(\frac{m}{2}+h)} dX_2. \quad (v)$$

But the x -integral gives $\frac{\Gamma(h)\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2}+h)}(2\sigma_{11})^{-h}$ for $\Re(h) > 0$ and the X_2 -integral gives $\frac{\Gamma_{p_2}(h)\Gamma_{p_2}(\frac{m}{2})}{\Gamma_{p_2}(\frac{m}{2}+h)}|2\Sigma_{22}|^{-h}$ for $\Re(h) > \frac{p_2-1}{2}$. Substituting all these in (v), we note that all the factors containing 2 and $\Sigma, \sigma_{11}, \Sigma_{22}$ cancel out, and then by using the fact that

$$\frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma(\frac{m}{2} + h)\Gamma_{p-1}(\frac{m}{2} + h)} = \pi^{\frac{p-1}{2}} \frac{\Gamma(\frac{m}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{m}{2} + h)},$$

we have the following expression for the h -th null moment of u :

$$E[u^h | H_o] = \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} - \frac{p-1}{2})} \frac{\Gamma(\frac{m}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{m}{2} + h)}, \quad \Re(h) > -\frac{m}{2} + \frac{p-1}{2}, \quad (5.6.14)$$

which happens to be the h -th moment of a real scalar type-1 beta random variable with the parameters $(\frac{m}{2} - \frac{p-1}{2}, \frac{p-1}{2})$. Since h is arbitrary, this h -th moment uniquely determines the distribution, thus the following result:

Theorem 5.6.3. *When the population has a p -variate Gaussian distribution with the parameters μ and $\Sigma > O$, and the population multiple correlation coefficient $\rho_{1.(2...p)} = 0$, the sample multiple correlation coefficient $r_{1.(2...p)}$ is such that $u = 1 - r_{1.(2...p)}^2$ is distributed as a real scalar type-1 beta random variable with the parameters $(\frac{m}{2} - \frac{p-1}{2}, \frac{p-1}{2})$, and thereby $v = \frac{u}{1-u} = \frac{1-r_{1.(2...p)}^2}{r_{1.(2...p)}^2}$ is distributed as a real scalar type-2 beta random variable with the parameters $(\frac{m}{2} - \frac{p-1}{2}, \frac{p-1}{2})$ and $w = \frac{1-u}{u} = \frac{r_{1.(2...p)}^2}{1-r_{1.(2...p)}^2}$ is distributed as a real scalar type-2 beta random variable with the parameters $(\frac{p-1}{2}, \frac{m}{2} - \frac{p-1}{2})$ whose density is*

$$f_w(w) = \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} - \frac{p-1}{2})\Gamma(\frac{p-1}{2})} w^{\frac{p-1}{2}-1} (1+w)^{-(\frac{m}{2})}, \quad 0 \leq w < \infty, \quad (5.6.15)$$

and zero elsewhere.

As F -tables are available, we may conveniently express the above real scalar type-2 beta density in terms of an F -density. It suffices to make the substitution $w = \frac{p-1}{m-p+1}F$ where F is a real F random variable having $p-1$ and $m-p+1$ degrees of freedom, that is, an $F_{p-1, m-p+1}$ random variable, with $m = n-1$, n being the sample size. The density of this F random variable, denoted by $f_F(F)$, is the following:

$$f_F(F) = \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{p-1}{2})\Gamma(\frac{m}{2} - \frac{p-1}{2})} \left(\frac{p-1}{m-p+1}\right)^{\frac{p-1}{2}} F^{\frac{p-1}{2}-1} \left(1 + \frac{p-1}{m-p+1}F\right)^{-\frac{m}{2}}, \quad (5.6.16)$$

whenever $0 \leq F < \infty$, and zero elsewhere. In the above simplification, observe that $\frac{(p-1)/2}{(\frac{m}{2} - \frac{p-1}{2})} = \frac{p-1}{m-p+1}$. Then, for taking a decision with respect to testing the hypothesis

$H_o : \rho_{1.(2\dots p)} = 0$, first compute $F_{p-1, m-p+1} = \frac{m-p+1}{p-1}w$, $w = \frac{r_{1.(2\dots p)}^2}{1-r_{1.(2\dots p)}^2}$. Then, reject H_o if the observed $F_{p-1, m-p+1} \geq F_{p-1, m-p+1, \alpha}$ for a given α . This will be a test at significance level α or, equivalently, a test whose critical region's size is α . The non-null distribution for evaluating the power of this likelihood ratio test can be determined by evaluating the integral in (5.6.13) and identifying the distribution through the uniqueness property of arbitrary moments.

Note 5.6.1. By making use of Theorem 5.6.3 as a starting point and exploiting various results connecting real scalar type-1 beta, type-2 beta, F and gamma variables, one can obtain numerous results on the distributional aspects of certain functions involving the sample multiple correlation coefficient.

5.6.5. The partial correlation coefficient

Partial correlation is a concept associated with the correlation between residuals in two variables after removing the effects of linear regression on a set of other variables. Consider the real vector $X' = (x_1, x_2, x_3, \dots, x_p) = (x_1, x_2, X'_3)$, $X'_3 = (x_3, \dots, x_p)$ where x_1, \dots, x_p are all real scalar variables. Let the covariance matrix of X be $\Sigma > O$ and let it be partitioned as follows:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ X_{(3)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \Sigma_{13} \\ \sigma_{21} & \sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}, \quad X_{(3)} \text{ being } (p-2) \times 1,$$

where σ_{11} , σ_{12} , σ_{21} , σ_{22} are 1×1 , Σ_{13} and Σ_{23} are $1 \times (p-2)$, $\Sigma_{31} = \Sigma'_{13}$, $\Sigma_{32} = \Sigma'_{23}$ and Σ_{33} is $(p-2) \times (p-2)$. Let $E[X] = O$ without any loss of generality. Consider the problem of predicting x_1 by using a linear function of $X_{(3)}$. Then, the regression of x_1 on $X_{(3)}$ is $E[x_1|X_{(3)}] = \Sigma_{13}\Sigma_{33}^{-1}X_{(3)}$ from (5.6.10), and the residual part, after removing

this regression from x_1 is $e_1 = x_1 - \Sigma_{13}\Sigma_{33}^{-1}X_{(3)}$. Similarly, the linear regression of x_2 on $X_{(3)}$ is $E[x_2|X_{(3)}] = \Sigma_{23}\Sigma_{33}^{-1}X_{(3)}$ and the residual in x_2 after removing the effect of $X_{(3)}$ is $e_2 = x_2 - \Sigma_{23}\Sigma_{33}^{-1}X_{(3)}$. What are then the variances of e_1 and e_2 , the covariance between e_1 and e_2 , and the scale-free covariance, namely the correlation between e_1 and e_2 ? Since e_1 and e_2 are all linear functions of the variables involved, we can utilize the expressions for variances of linear functions and covariance between linear functions, a basic discussion of such results being given in Mathai and Haubold (2017b). Thus,

$$\begin{aligned}\text{Var}(e_1) &= \text{Var}(x_1) + \text{Var}(\Sigma_{13}\Sigma_{33}^{-1}X_{(3)}) - 2\text{Cov}(x_1, \Sigma_{13}\Sigma_{33}^{-1}X_{(3)}) \\ &= \sigma_{11} + \Sigma_{13}\Sigma_{33}^{-1}\text{Cov}(X_{(3)})\Sigma_{33}^{-1}\Sigma_{31} - 2\text{Cov}(x_1, \Sigma_{33}^{-1}X_{(3)}) \\ &= \sigma_{11} + \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} - 2\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31} = \sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31}.\end{aligned}\quad (i)$$

It can be similarly shown that

$$\text{Var}(e_2) = \sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32} \quad (ii)$$

$$\text{Cov}(e_1, e_2) = \sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}.\quad (iii)$$

Then, the correlation between the residuals e_1 and e_2 , which is called the partial correlation between x_1 and x_2 after removing the effects of linear regression on $X_{(3)}$ and is denoted by $\rho_{12.(3\dots p)}$, is such that

$$\rho_{12.(3\dots p)}^2 = \frac{[\sigma_{12} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}]^2}{[\sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{31}][\sigma_{22} - \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{32}]}. \quad (5.6.17)$$

In the above simplifications, we have for instance used the fact that $\Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32} = \Sigma_{23}\Sigma_{33}^{-1}\Sigma_{31}$ since both are real 1×1 and one is the transpose of the other.

The corresponding sample partial correlation coefficient between x_1 and x_2 after removing the effects of linear regression on $X_{(3)}$, denoted by $r_{12.(3\dots p)}$, is such that:

$$r_{12.(3\dots p)}^2 = \frac{[s_{12} - S_{13}S_{33}^{-1}S_{32}]^2}{[s_{11} - S_{13}S_{33}^{-1}S_{31}][s_{22} - S_{23}S_{33}^{-1}S_{32}]} \quad (5.6.18)$$

where the sample sum of products matrix S is partitioned correspondingly, that is,

$$S = \begin{bmatrix} s_{11} & s_{12} & S_{13} \\ s_{21} & s_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix}, \quad S_{33} \text{ being } (p-2) \times (p-2), \quad (5.6.19)$$

and s_{11} , s_{12} , s_{21} , s_{22} being 1×1 . In all the above derivations, we did not use any assumption of an underlying Gaussian population. The results hold for any general population as long as product moments up to second order exist. However, if we assume a p -variate nonsingular Gaussian population, then we can obtain some interesting results on the distributional aspects of the sample partial correlation, as was done in the case of the sample multiple correlation. Such results will not be herein considered.

Exercises 5.6

5.6.1. Let the $p \times p$ real positive definite matrix W be distributed as $W \sim W_p(m, \Sigma)$ with $\Sigma = I$. Consider the partitioning $W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$ where W_{11} is $r \times r$, $r < p$. Evaluate explicitly the normalizing constant in the density of W by first integrating out (1): W_{11} , (2): W_{22} , (3): W_{12} .

5.6.2. Repeat Exercise 5.6.1 for the complex case.

5.6.3. Let the $p \times p$ real positive definite matrix W have a real Wishart density with degrees of freedom $m \geq p$ and parameter matrix $\Sigma > O$. Consider the transformation $W = TT'$ where T is lower triangular with positive diagonal elements. Evaluate the densities of the t_{jj} 's and the t_{ij} 's, $i > j$ if (1): $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$, (2): $\Sigma > O$ is a general matrix.

5.6.4. Repeat Exercise 5.6.3 for the complex case. In the complex case, the diagonal elements in T are real and positive.

5.6.5. Let $S \sim W_p(m, \Sigma)$, $\Sigma > O$. Compute the density of S^{-1} in the real case, and repeat for the complex case.

5.7. Distributions of Products and Ratios of Matrix-variate Random Variables

In the real scalar case, one can easily interpret products and ratios of real scalar variables, whether these are random or mathematical variables. However, when it comes to matrices, products and ratios are to be carefully defined. Let X_1 and X_2 be independently distributed $p \times p$ real symmetric and positive definite matrix-variate random variables with density functions $f_1(X_1)$ and $f_2(X_2)$, respectively. By definition, f_1 and f_2 are respectively real-valued scalar functions of the matrices X_1 and X_2 . Due to statistical independence of X_1 and X_2 , their joint density, denoted by $f(X_1, X_2)$, is the product of the marginal densities, that is, $f(X_1, X_2) = f_1(X_1)f_2(X_2)$. Let us define a ratio and a product of matrices. Let $U_2 = X_2^{\frac{1}{2}}X_1X_2^{\frac{1}{2}}$ and $U_1 = X_2^{\frac{1}{2}}X_1^{-1}X_2^{\frac{1}{2}}$ be called the symmetric product

and symmetric ratio of the matrices X_1 and X_2 , where $X_2^{\frac{1}{2}}$ denotes the positive definite square root of the positive definite matrix X_2 . Let us consider the product U_2 first. We could have also defined a product by interchanging X_1 and X_2 . When it comes to ratios, we could have considered the ratios X_1 to X_2 as well as X_2 to X_1 . Nonetheless, we will start with U_1 and U_2 as defined above.

5.7.1. The density of a product of real matrices

Consider the transformation $U_2 = X_2^{\frac{1}{2}} X_1 X_2^{\frac{1}{2}}$, $V = X_2$. Then, it follows from Theorem 1.6.5 that:

$$dX_1 \wedge dX_2 = |V|^{-\frac{p+1}{2}} dU_2 \wedge dV. \tag{5.7.1}$$

Letting the joint density of U_2 and V be denoted by $g(U_2, V)$ and the marginal density of U_2 , by $g_2(U_2)$, we have

$$\begin{aligned} f_1(X_1) f_2(X_2) dX_1 \wedge dX_2 &= |V|^{-\frac{p+1}{2}} f_1(V^{-\frac{1}{2}} U_2 V^{-\frac{1}{2}}) f_2(V) dU_2 \wedge dV \\ g_2(U_2) &= \int_V |V|^{-\frac{p+1}{2}} f_1(V^{-\frac{1}{2}} U_2 V^{-\frac{1}{2}}) f_2(V) dV, \end{aligned} \tag{5.7.2}$$

$g_2(U_2)$ being referred to as the density of the symmetric product U_2 of the matrices X_1 and X_2 . For example, letting X_1 and X_2 be independently distributed two-parameter matrix-variate gamma random variables with the densities

$$f_{3j}(X_j) = \frac{|B_j|^{\alpha_j}}{\Gamma_p(\alpha_j)} |X_j|^{\alpha_j - \frac{p+1}{2}} e^{-\text{tr}(B_j X_j)}, \quad j = 1, 2, \tag{i}$$

for $B_j > O$, $X_j > O$, $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, and zero elsewhere, we have

$$g_2(U_2) = c |U_2|^{\alpha_1 - \frac{p+1}{2}} \int_{V>O} |V|^{\alpha_2 - \alpha_1 - \frac{p+1}{2}} e^{-\text{tr}(B_2 V + B_1 V^{-\frac{1}{2}} U_2 V^{-\frac{1}{2}})} dV, \tag{5.7.3}$$

where c is the product of the normalizing constants of the densities specified in (i). On comparing (5.7.3) with the Krätzel integral defined in the real scalar case in Chap. 2, as well as in Mathai (2012) and Mathai and Haubold (1988, 2011a, 2017,a), it is seen that (5.7.3) can be regarded as a real matrix-variate analogue of Krätzel’s integral. One could also obtain the real matrix-variate version of the inverse Gaussian density from the integrand.

As another example, let $f_1(X_1)$ be a real matrix-variate type-1 beta density as previously defined in this chapter, whose parameters are $(\gamma + \frac{p+1}{2}, \alpha)$ with $\Re(\alpha) > \frac{p-1}{2}$, $\Re(\gamma) > -1$, its density being given by

$$f_4(X_1) = \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})\Gamma_p(\alpha)} |X_1|^\gamma |I - X_1|^{\alpha - \frac{p+1}{2}} \quad (ii)$$

for $O < X_1 < I$, $\Re(\gamma) > -1$, $\Re(\alpha) > \frac{p-1}{2}$, and zero elsewhere. Letting $f_2(X_2) = f(X_2)$ be any other density, the density of U_2 is then

$$\begin{aligned} g_2(U_2) &= \int_V |V|^{-\frac{p+1}{2}} f_1(V^{-\frac{1}{2}}U_2V^{-\frac{1}{2}})f_2(V)dV \\ &= \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})\Gamma_p(\alpha)} \int_V |V|^{-\frac{p+1}{2}} |V^{-\frac{1}{2}}U_2V^{-\frac{1}{2}}|^\gamma |I - V^{-\frac{1}{2}}U_2V^{-\frac{1}{2}}|^{\alpha - \frac{p+1}{2}} f(V)dV \\ &= \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})} \frac{|U_2|^\gamma}{\Gamma_p(\alpha)} \int_{V>U_2>O} |V|^{-\alpha-\gamma} |V - U_2|^{\alpha - \frac{p+1}{2}} f(V)dV \\ &= \frac{\Gamma_p(\gamma + \frac{p+1}{2} + \alpha)}{\Gamma_p(\gamma + \frac{p+1}{2})} K_{2,U_2,\gamma}^{-\alpha} f \end{aligned} \quad (5.7.4)$$

where

$$K_{2,U_2,\gamma}^{-\alpha} f = \frac{|U_2|^\gamma}{\Gamma_p(\alpha)} \int_{V>U_2>O} |V|^{-\alpha-\gamma} |V - U_2|^{\alpha - \frac{p+1}{2}} f(V)dV, \quad \Re(\alpha) > \frac{p-1}{2}, \quad (5.7.5)$$

is called the real matrix-variate Erdélyi-Kober right-sided or second kind fractional integral of order α and parameter γ as for $p = 1$, that is, in the real scalar case, (5.7.5) corresponds to the Erdélyi-Kober fractional integral of the second kind of order α and parameter γ . This connection of the density of a symmetric product of matrices to a fractional integral of the second kind was established by Mathai (2009, 2010) and further papers.

5.7.2. M-convolution and fractional integral of the second kind

Mathai (1997) referred to the structure in (5.7.2) as the M-convolution of a product where f_1 and f_2 need not be statistical densities. Actually, they could be any function provided the integral exists. However, if f_1 and f_2 are statistical densities, this M-convolution of a product can be interpreted as the density of a symmetric product. Thus, a physical interpretation to an M-convolution of a product is provided in terms of statistical densities. We have seen that (5.7.2) is connected to a fractional integral when f_1 is a real

matrix-variate type-1 beta density and f_2 is an arbitrary density. From this observation, one can introduce a general definition for a fractional integral of the second kind in the real matrix-variate case. Let

$$f_1(X_1) = \phi_1(X_1) \frac{|I - X_1|^{\alpha - \frac{p+1}{2}}}{\Gamma_p(\alpha)}, \Re(\alpha) > \frac{p-1}{2}, \tag{iii}$$

and $f_2(X_2) = \phi_2(X_2)f(X_2)$ where ϕ_1 and ϕ_2 are specified functions and f is an arbitrary function. Then, consider the M-convolution of a product, again denoted by $g_2(U_2)$:

$$g_2(U_2) = \int_V |V|^{-\frac{p+1}{2}} \phi_1(V^{-\frac{1}{2}}U_2V^{-\frac{1}{2}}) \frac{|I - V^{-\frac{1}{2}}U_2V^{-\frac{1}{2}}|^{\alpha - \frac{p+1}{2}}}{\Gamma_p(\alpha)} \times \phi_2(V)f(V)dV, \Re(\alpha) > \frac{p-1}{2}. \tag{5.7.6}$$

The right-hand side (5.7.6) will be called a fractional integral of the second kind of order α in the real matrix-variate case. By letting $p = 1$ and specifying ϕ_1 and ϕ_2 , one can obtain all the fractional integrals of the second kind of order α that have previously been defined by various authors. Hence, for a general p , one has the corresponding real matrix-variate cases. For example, on letting $\phi_1(X_1) = |X_1|^\nu$ and $\phi_2(X_2) = 1$, one has Erdélyi-Kober fractional integral of the second kind of (5.7.5) in the real matrix-variate case as for $p = 1$, it is the Erdélyi-Kober fractional integral of the second kind of order α . Letting $\phi_1(X_1) = 1$ and $\phi_2(X_2) = |X_2|^\alpha$, (5.7.6) simplifies to the following integral, again denoted by $g_2(U_2)$:

$$g_2(U_2) = \frac{1}{\Gamma_p(\alpha)} \int_{V>U_2>O} |V - U_2|^{\alpha - \frac{p+1}{2}} f(V)dV. \tag{5.7.7}$$

For $p = 1$, (5.7.7) is Weyl fractional integral of the second kind of order α . Accordingly, (5.7.7) is Weyl fractional integral of the second kind in the real matrix-variate case. For $p = 1$, (5.7.7) is also the Riemann-Liouville fractional integral of the second kind of order α in the real scalar case, if there exists a finite upper bound for V . If V is bounded above by a real positive definite constant matrix $B > O$ in the integral in (5.7.7), then (5.7.7) is Riemann-Liouville fractional integral of the second kind of order α for the real matrix-variate case. Connections to other fractional integrals of the second kind can be established by referring to Mathai and Haubold (2017).

The appeal of fractional integrals of the second kind resides in the fact that they can be given physical interpretations as the density of a symmetric product when f_1 and f_2

are densities or as an M-convolution of products, whether in the scalar variable case or the matrix-variate case, and that in both the real and complex domains.

5.7.3. A pathway extension of fractional integrals

Consider the following modification to the general definition of a fractional integral of the second kind of order α in the real matrix-variate case given in (5.7.6). Let

$$f_1(X_1) = \phi_1(X_1) \frac{1}{\Gamma_p(\alpha)} |I - a(1-q)X_1|^{\frac{\eta}{1-q} - \frac{p+1}{2}}, \quad f_2(X_2) = \phi_2(X_2) f(X_2), \quad (iv)$$

where $\Re(\alpha) > \frac{p-1}{2}$ and $q < 1$, and $a > 0$, $\eta > 0$ are real scalar constants. For all $q < 1$, $g_2(U_2)$ corresponding to $f_1(X_1)$ and $f_2(X_2)$ of (iv) will define a family of fractional integrals of the second kind. Observe that when X_1 and $I - a(1-q)X_1 > O$, then $O < X_1 < \frac{1}{a(1-q)}I$. However, by writing $(1-q) = -(q-1)$ for $q > 1$, one can switch into a type-2 beta form, namely, $I + a(q-1)X_1 > O$ for $q > 1$, which implies that $X_1 > O$ and the fractional nature is lost. As well, when $q \rightarrow 1$,

$$|I + a(q-1)X_1|^{-\frac{\eta}{q-1}} \rightarrow e^{-a\eta \operatorname{tr}(X_1)}$$

which is the exponential form or gamma density form. In this case too, the fractional nature is lost. Thus, through q , one can obtain matrix-variate type-1 and type-2 beta families and a gamma family of functions from (iv). Then q is called the pathway parameter which generates three families of functions. However, the fractional nature of the integrals is lost for the cases $q > 1$ and $q \rightarrow 1$. In the real scalar case, x_1 may have an exponent and making use of $[1 - (1-q)x_1^\delta]^{\alpha-1}$ can lead to interesting fractional integrals for $q < 1$. However, raising X_1 to an exponent δ in the matrix-variate case will fail to produce results of interest as Jacobians will then take inconvenient forms that cannot be expressed in terms of the original matrices; this is for example explained in detail in Mathai (1997) for the case of a squared real symmetric matrix.

5.7.4. The density of a ratio of real matrices

One can define a symmetric ratio in four different ways: $X_2^{\frac{1}{2}}X_1^{-1}X_2^{\frac{1}{2}}$ with $V = X_2$ or $V = X_1$ and $X_1^{\frac{1}{2}}X_2^{-1}X_1^{\frac{1}{2}}$ with $V = X_2$ or $V = X_1$. All these four forms will produce different structures on $f_1(X_1)f_2(X_2)$. Since the form U_1 that was specified in Sect. 5.7 in terms of $X_2^{\frac{1}{2}}X_1^{-1}X_2^{\frac{1}{2}}$ with $V = X_2$ provides connections to fractional integrals of the first kind, we will consider this one whose density, denoted by $g_1(U_1)$, is the following observing that $dX_1 \wedge dX_2 = |V|^{\frac{p+1}{2}} |U_1|^{-(p+1)} dU_1 \wedge dV$:

$$g_1(U_1) = \int_V |V|^{\frac{p+1}{2}} |U_1|^{-(p+1)} f_1(V^{\frac{1}{2}}U_1^{-1}V^{\frac{1}{2}}) f_2(V) dV \quad (5.7.8)$$

provided the integral exists. As in the fractional integral of the second kind in real matrix-variate case, we can give a general definition for a fractional integral of the first kind in the real matrix-variate case as follows: Let $f_1(X_1)$ and $f_2(X_2)$ be taken as in the case of fractional integral of the second kind with ϕ_1 and ϕ_2 as preassigned functions. Then

$$g_1(U_1) = \int_V |V|^{\frac{p+1}{2}} |U_1|^{-(p+1)} \phi_1(V^{\frac{1}{2}} U_1^{-1} V^{\frac{1}{2}}) \times \frac{1}{\Gamma_p(\alpha)} |I - V^{\frac{1}{2}} U_1^{-1} V^{\frac{1}{2}}|^{\alpha - \frac{p+1}{2}} \phi_2(V) f(V) dV, \Re(\alpha) > \frac{p-1}{2}. \tag{5.7.9}$$

As an example, letting

$$\phi_1(X_1) = \frac{\Gamma_p(\gamma + \alpha)}{\Gamma_p(\gamma)} |X_1|^{\gamma - \frac{p+1}{2}} \text{ and } \phi_2(X_2) = 1, \Re(\gamma) > \frac{p-1}{2},$$

we have

$$g_1(U_1) = \frac{\Gamma_p(\gamma + \alpha)}{\Gamma_p(\gamma)} \frac{|U_1|^{-\alpha - \gamma}}{\Gamma_p(\alpha)} \int_{V < U_1} |V|^\gamma |U_1 - V|^{\alpha - \frac{p+1}{2}} f(V) dV = \frac{\Gamma_p(\gamma + \alpha)}{\Gamma_p(\gamma)} K_{1,U_1,\gamma}^{-\alpha} f \tag{5.7.10}$$

for $\Re(\alpha) > \frac{p-1}{2}, \Re(\gamma) > \frac{p-1}{2}$, where

$$K_{1,U_1,\gamma}^{-\alpha} f = \frac{|U_1|^{-\alpha - \gamma}}{\Gamma_p(\alpha)} \int_{V < U_1} |V|^\gamma |U_1 - V|^{\alpha - \frac{p+1}{2}} f(V) dV \tag{5.7.11}$$

for $\Re(\alpha) > \frac{p-1}{2}, \Re(\gamma) > \frac{p-1}{2}$, is Erdélyi-Kober fractional integral of the first kind of order α and parameter γ in the real matrix-variate case. Since for $p = 1$ or in the real scalar case, $K_{1,u_1,\gamma}^{-\alpha} f$ is Erdélyi-Kober fractional integral of order α and parameter γ , the first author referred to $K_{1,U_1,\gamma}^{-\alpha} f$ in (5.7.11) as Erdélyi-Kober fractional integral of the first kind of order α in the real matrix-variate case.

By specializing ϕ_1 and ϕ_2 in the real scalar case, that is, for $p = 1$, one can obtain all the fractional integrals of the first kind of order α that have been previously introduced in the literature by various authors. One can similarly derive the corresponding results on fractional integrals of the first kind in the real matrix-variate case. Before concluding this section, we will consider one more special case. Let

$$\phi_1(X_1) = |X_1|^{-\alpha - \frac{p+1}{2}} \text{ and } \phi_2(X_2) = |X_2|^\alpha.$$

In this case, $g_1(U_1)$ is not a statistical density but it is the M-convolution of a ratio. Under the above substitutions, $g_1(U_1)$ of (5.7.9) becomes

$$g_1(U_1) = \frac{1}{\Gamma_p(\alpha)} \int_{V < U_1} |U_1 - V|^{\alpha - \frac{p+1}{2}} f(V) dV, \Re(\alpha) > \frac{p-1}{2}. \quad (5.7.12)$$

For $p = 1$, (5.7.12) is Weyl fractional integral of the first kind of order α ; accordingly the first author refers to (5.7.12) as Weyl fractional integral of the first kind of order α in the real matrix-variate case. Since we are considering only real positive definite matrices here, there is a natural lower bound for the integral or the integral is over $O < V < U_1$. When there is a specific lower bound, such as $O < V$, then for $p = 1$, (5.7.12) is called the Riemann-Liouville fractional integral of the first kind of order α . Hence (5.7.12) will be referred to as the Riemann-Liouville fractional integral of the first kind of order α in the real matrix-variate case.

Example 5.7.1. Let X_1 and X_2 be independently distributed $p \times p$ real positive definite gamma matrix-variate random variables whose densities are

$$f_j(X_j) = \frac{1}{\Gamma_p(\alpha_j)} |X_j|^{\alpha_j - \frac{p+1}{2}} e^{-\text{tr}(X_j)}, \quad X_j > O, \Re(\alpha_j) > \frac{p-1}{2}, \quad j = 1, 2,$$

and zero elsewhere. Show that the densities of the symmetric ratios of matrices $U_1 = X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}}$ and $U_2 = X_2^{\frac{1}{2}} X_1^{-1} X_2^{\frac{1}{2}}$ are identical.

Solution 5.7.1. Observe that for $p = 1$ that is, in the real scalar case, both U_1 and U_2 are the ratio of real scalar variables $\frac{x_2}{x_1}$ but in the matrix-variate case U_1 and U_2 are different matrices. Hence, we cannot expect the densities of U_1 and U_2 to be the same. They will happen to be identical because of a property called functional symmetry of the gamma densities. Consider U_1 and let $V = X_1$. Then, $X_2 = V^{\frac{1}{2}} U_1 V^{\frac{1}{2}}$ and $dX_1 \wedge dX_2 = |V|^{\frac{p+1}{2}} dV \wedge dU_1$. Due to the statistical independence of X_1 and X_2 , their joint density is $f_1(X_1) f_2(X_2)$ and the joint density of U_1 and V is $|V|^{\frac{p+1}{2}} f_1(V) f_2(V^{\frac{1}{2}} U_1 V^{\frac{1}{2}})$, the marginal density of U_1 , denoted by $g_1(U_1)$, being the following:

$$g_1(U_1) = \frac{|U_1|^{\alpha_2 - \frac{p+1}{2}}}{\Gamma_p(\alpha_1) \Gamma_p(\alpha_2)} \int_{V > O} |V|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(V + V^{\frac{1}{2}} U_1 V^{\frac{1}{2}})} dV.$$

The exponent of e can be written as follows:

$$-\text{tr}(V) - \text{tr}(V^{\frac{1}{2}} U_1 V^{\frac{1}{2}}) = -\text{tr}(V) - \text{tr}(V U_1) = -\text{tr}(V(I + U_1)) = -\text{tr}[(I + U_1)^{\frac{1}{2}} V (I + U_1)^{\frac{1}{2}}].$$

Letting $Y = (I + U_1)^{\frac{1}{2}}V(I + U_1)^{\frac{1}{2}} \Rightarrow dY = |I + U_1|^{\frac{p+1}{2}}dV$. Then carrying out the integration in $g_1(U_1)$, we obtain the following density function:

$$g_1(U_1) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)}|U_1|^{\alpha_2 - \frac{p+1}{2}}|I + U_1|^{-(\alpha_1 + \alpha_2)}, \tag{i}$$

which is a real matrix-variate type-2 beta density with the parameters (α_2, α_1) . The original conditions $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, remain the same, no additional conditions being needed. Now, consider U_2 and let $V = X_2$ so that $X_1 = V^{\frac{1}{2}}U_2^{-1}V^{\frac{1}{2}} \Rightarrow dX_1 \wedge dX_2 = |V|^{\frac{p+1}{2}}|U_2|^{-(p+1)}dV \wedge dU_2$. The marginal density of U_2 is then:

$$g_2(U_2) = \frac{|U_2|^{-\alpha_1 + \frac{p+1}{2}}|U_2|^{-(p+1)}}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} \int_{V>O} |V|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}[V + V^{\frac{1}{2}}U_2^{-1}V^{\frac{1}{2}}]} \tag{ii}$$

As previously explained, the exponent in (ii) can be simplified to $-\text{tr}[(I + U_2^{-1})^{\frac{1}{2}}V(I + U_2^{-1})^{\frac{1}{2}}]$, which once integrated out yields $\Gamma_p(\alpha_1 + \alpha_2)|I + U_2^{-1}|^{-(\alpha_1 + \alpha_2)}$. Then,

$$|U_2|^{-\alpha_1 + \frac{p+1}{2}}|U_2|^{-(p+1)}|I + U_2^{-1}|^{-(\alpha_1 + \alpha_2)} = |U_2|^{\alpha_2 - \frac{p+1}{2}}|I + U_2|^{-(\alpha_1 + \alpha_2)}. \tag{iii}$$

It follows from (i),(ii) and (iii) that $g_1(U_1) = g_2(U_2)$. Thus, the densities of U_1 and U_2 are indeed one and the same, as had to be proved.

5.7.5. A pathway extension of first kind integrals, real matrix-variate case

As in the case of fractional integral of the second kind, we can also construct a pathway extension of the first kind integrals in the real matrix-variate case. Let

$$f_1(X_1) = \frac{\phi_1(X_1)}{\Gamma_p(\alpha)}|I - a(1 - q)X_1|^{\alpha - \frac{p+1}{2}}, \alpha = \frac{\eta}{1 - q}, \Re(\alpha) > \frac{p - 1}{2}, \tag{5.7.13}$$

and $f_2(X_2) = \phi_2(X_2)f(X_2)$ for the scalar parameters $a > 0$, $\eta > 0$, $q < 1$. When $q < 1$, (5.7.13) remains in the generalized type-1 beta family of functions. However, when $q > 1$, f_1 switches to the generalized type-2 beta family of functions and when $q \rightarrow 1$, (5.6.13) goes into a gamma family of functions. Since $X_1 > O$ for $q > 1$ and $q \rightarrow 1$, the fractional nature is lost in those instances. Hence, only the case $q < 1$ is relevant in this subsection.

For various values of $q < 1$, one has a family of fractional integrals of the first kind coming from (5.7.13). For details on the concept of pathway, the reader may refer to Mathai

(2005) and later papers. With the function $f_1(X_1)$ as specified in (5.7.13) and the corresponding $f_2(X_2) = \phi_2(X_2)f(X_2)$, one can write down the M-convolution of a ratio, $g_1(U_1)$, corresponding to (5.7.8). Thus, we have the pathway extended form of $g_1(U_1)$.

5.7a. Density of a Product and Integrals of the Second Kind

The discussion in this section parallels that in the real matrix-variate case. Hence, only a summarized treatment will be provided. With respect to the density of a product when \tilde{f}_1 and \tilde{f}_2 are matrix-variate gamma densities in the complex domain, the results are parallel to those obtained in the real matrix-variate case. Hence, we will consider an extension of fractional integrals to the complex matrix-variate cases. Matrices in the complex domain will be denoted with a tilde. Let \tilde{X}_1 and \tilde{X}_2 be independently distributed Hermitian positive definite complex matrix-variate random variables whose densities are $\tilde{f}_1(\tilde{X}_1)$ and $\tilde{f}_2(\tilde{X}_2)$, respectively. Let $\tilde{U}_2 = \tilde{X}_2^{\frac{1}{2}}\tilde{X}_1\tilde{X}_2^{\frac{1}{2}}$ and $\tilde{U}_1 = \tilde{X}_2^{\frac{1}{2}}\tilde{X}_1^{-1}\tilde{X}_2^{\frac{1}{2}}$ where $\tilde{X}_2^{\frac{1}{2}}$ denotes the Hermitian positive definite square root of the Hermitian positive definite matrix \tilde{X}_2 . Statistical densities are real-valued scalar functions whether the argument matrix is in the real or complex domain.

5.7a.1. Density of a product and fractional integral of the second kind, complex case

Let us consider the transformation $(\tilde{X}_1, \tilde{X}_2) \rightarrow (\tilde{U}_2, \tilde{V})$ and $(\tilde{X}_1, \tilde{X}_2) \rightarrow (\tilde{U}_1, \tilde{V})$, the Jacobians being available from Chap. 1 or Mathai (1997). Then,

$$d\tilde{X}_1 \wedge d\tilde{X}_2 = \begin{cases} |\det(\tilde{V})|^{-p} d\tilde{U}_2 \wedge d\tilde{V} \\ |\det(\tilde{V})|^p |\det(\tilde{U}_1)|^{-2p} d\tilde{U}_1 \wedge d\tilde{V}. \end{cases} \quad (5.7a.1)$$

When f_1 and f_2 are statistical densities, the density of the product, denoted by $\tilde{g}_2(\tilde{U}_2)$, is the following:

$$\tilde{g}_2(\tilde{U}_2) = \int_{\tilde{V}} |\det(\tilde{V})|^{-p} f_1(\tilde{V}^{-\frac{1}{2}}\tilde{U}_2\tilde{V}^{-\frac{1}{2}}) f_2(\tilde{V}) d\tilde{V} \quad (5.7a.2)$$

where $|\det(\cdot)|$ is the absolute value of the determinant of (\cdot) . If f_1 and f_2 are not statistical densities, (5.7a.1) will be called the M-convolution of the product. As in the real matrix-variate case, we will give a general definition of a fractional integral of order α of the second kind in the complex matrix-variate case. Let

$$\tilde{f}_1(\tilde{X}_1) = \phi_1(\tilde{X}_1) \frac{1}{\tilde{\Gamma}_p(\alpha)} |\det(I - \tilde{X}_1)|^{\alpha-p}, \quad \Re(\alpha) > p - 1,$$

and $f_2(\tilde{X}_2) = \phi_2(\tilde{X}_2)\tilde{f}(\tilde{X}_2)$ where ϕ_1 and ϕ_2 are specified functions and f is an arbitrary function. Then, (5.7a.2) becomes

$$\begin{aligned} \tilde{g}_2(\tilde{U}_2) &= \int_{\tilde{V}} |\det(\tilde{V})|^{-p} \phi_1(\tilde{V}^{-\frac{1}{2}}\tilde{U}_2\tilde{V}^{-\frac{1}{2}}) \\ &\quad \times \frac{1}{\tilde{\Gamma}_p(\alpha)} |\det(I - \tilde{V}^{-\frac{1}{2}}\tilde{U}_2\tilde{V}^{-\frac{1}{2}})|^{\alpha-p} \phi_2(\tilde{V}) f(\tilde{V}) d\tilde{V} \end{aligned} \tag{5.7a.3}$$

for $\Re(\alpha) > p - 1$. As an example, let

$$\phi_1(\tilde{X}_1) = \frac{\tilde{\Gamma}_p(\gamma + p + \alpha)}{\tilde{\Gamma}_p(\gamma + p)} |\det(\tilde{X}_1)|^\gamma \text{ and } \phi_2(\tilde{X}_2) = 1.$$

Observe that $\tilde{f}_1(\tilde{X}_1)$ has now become a complex matrix-variate type-1 beta density with the parameters $(\gamma + p, \alpha)$ so that (5.7a.3) can be expressed as follows:

$$\begin{aligned} \tilde{g}_2(\tilde{U}_2) &= \frac{\tilde{\Gamma}_p(\gamma + p + \alpha)}{\tilde{\Gamma}_p(\gamma + p)} \frac{|\det(\tilde{U}_2)|^\gamma}{\tilde{\Gamma}_p(\alpha)} \int_{\tilde{V} > \tilde{U}_2 > O} |\det(\tilde{V})|^{-\alpha-\gamma} |\det(\tilde{V} - \tilde{U}_2)|^{\alpha-p} \tilde{f}(\tilde{V}) d\tilde{V} \\ &= \frac{\tilde{\Gamma}_p(\gamma + p + \alpha)}{\tilde{\Gamma}_p(\gamma + p)} \tilde{K}_{2, \tilde{U}_2, \gamma}^{-\alpha} f \end{aligned} \tag{5.7a.4}$$

where

$$\tilde{K}_{2, \tilde{U}_2, \gamma}^{-\alpha} f = \frac{|\det(\tilde{U}_2)|^\gamma}{\tilde{\Gamma}_p(\alpha)} \int_{\tilde{V} > \tilde{U}_2 > O} |\det(\tilde{V})|^{-\alpha-\gamma} |\det(\tilde{V} - \tilde{U}_2)|^{\alpha-p} f(\tilde{V}) d\tilde{V} \tag{5.7a.5}$$

is Erdélyi-Kober fractional integral of the second kind of order α in the complex matrix-variate case, which is defined for $\Re(\alpha) > p - 1, \Re(\gamma) > -1$. The extension of fractional integrals to complex matrix-variate cases was introduced in Mathai (2013). As a second example, let

$$\phi_1(\tilde{X}_1) = 1 \text{ and } \phi_2(\tilde{X}_2) = |\det(\tilde{V})|^\alpha.$$

In that case, (5.7a.3) becomes

$$\tilde{g}_2(\tilde{U}_2) = \int_{\tilde{V} > \tilde{U}_2 > O} |\det(\tilde{V} - \tilde{U}_2)|^{\alpha-p} f(V) d\tilde{V}, \Re(\alpha) > p - 1. \tag{5.7a.6}$$

The integral (5.7a.6) is Weyl fractional integral of the second kind of order α in the complex matrix-variate case. If V is bounded above by a Hermitian positive definite constant

matrix $B > O$, then (5.7a.6) is a Riemann-Liouville fractional integral of the second kind of order α in the complex matrix-variate case.

A pathway extension parallel to that developed in the real matrix-variate case can be similarly obtained. Accordingly, the details of the derivation are omitted.

5.7a.2. Density of a ratio and fractional integrals of the first kind, complex case

We will now derive the density of the symmetric ratio \tilde{U}_1 defined in Sect. 5.7a. If \tilde{f}_1 and \tilde{f}_2 are statistical densities, then the density of \tilde{U}_1 , denoted by $\tilde{g}_1(\tilde{U}_1)$, is given by

$$\tilde{g}_1(\tilde{U}_1) = \int_{\tilde{V}} |\det(\tilde{V})|^p |\det(\tilde{U}_1)|^{-2p} \tilde{f}_1(\tilde{V}^{\frac{1}{2}} \tilde{U}^{-1} \tilde{V}^{\frac{1}{2}}) \tilde{f}_2(\tilde{V}) d\tilde{V}, \quad (5.7a.7)$$

provided the integral is convergent. For the general definition, let us take

$$\tilde{f}_1(\tilde{X}_1) = \phi_1(\tilde{X}_1) \frac{1}{\tilde{\Gamma}_p(\alpha)} |\det(I - \tilde{X}_1)|^{\alpha-p}, \quad \Re(\alpha) > p - 1,$$

and $\tilde{f}_2(\tilde{X}_2) = \phi_2(\tilde{X}_2) \tilde{f}(\tilde{X}_2)$ where ϕ_1 and ϕ_2 are specified functions and \tilde{f} is an arbitrary function. Then $\tilde{g}_1(\tilde{U}_1)$ is the following:

$$\begin{aligned} \tilde{g}_1(\tilde{U}_1) &= \int_{\tilde{V}} |\det(\tilde{V})|^p |\det(\tilde{U}_1)|^{-2p} \frac{1}{\tilde{\Gamma}_p(\alpha)} \phi_1(\tilde{V}^{\frac{1}{2}} \tilde{U}_1^{-1} \tilde{V}^{\frac{1}{2}}) \\ &\quad \times |\det(I - \tilde{V}^{\frac{1}{2}} \tilde{U}_1^{-1} \tilde{V}^{\frac{1}{2}})|^{\alpha-p} \phi_2(\tilde{V}) \tilde{f}(\tilde{V}) d\tilde{V}. \end{aligned} \quad (5.7a.8)$$

As an example, let

$$\phi_1(\tilde{X}_1) = \frac{\tilde{\Gamma}_p(\gamma + \alpha)}{\tilde{\Gamma}_p(\gamma)} |\det(\tilde{X}_1)|^{\gamma-p}$$

and $\phi_2 = 1$. Then,

$$\begin{aligned} \tilde{g}_1(\tilde{U}_1) &= \frac{\tilde{\Gamma}_p(\gamma + \alpha)}{\tilde{\Gamma}_p(\gamma)} \frac{|\det(\tilde{U}_1)|^{-\alpha-\gamma}}{\tilde{\Gamma}_p(\alpha)} \int_{O < \tilde{V} < \tilde{U}_1} |\det(\tilde{V})|^\gamma \\ &\quad \times |\det(\tilde{U}_1 - \tilde{V})|^{\alpha-p} \tilde{f}(\tilde{V}) d\tilde{V}, \quad \Re(\alpha) > p - 1 \\ &= \frac{\tilde{\Gamma}_p(\gamma + \alpha)}{\tilde{\Gamma}_p(\gamma)} K_{1, \tilde{U}_1, \gamma}^{-\alpha} \tilde{f} \end{aligned} \quad (5.7a.9)$$

where

$$K_{1, \tilde{U}_1, \gamma}^{-\alpha} f = \frac{|\det(\tilde{U}_1)|^{-\alpha-\gamma}}{\tilde{\Gamma}_p(\alpha)} \int_{V < \tilde{U}_1} |\det(\tilde{V})|^\gamma |\det(\tilde{U}_1 - \tilde{V})|^{\alpha-p} f(\tilde{V}) d\tilde{V} \quad (5.7a.10)$$

for $\Re(\alpha) > p - 1$, is the Erdélyi-Kober fractional integral of the first kind of order α and parameter γ in the complex matrix-variate case. We now consider a second example. On letting $\phi_1(\tilde{X}_1) = |\det(\tilde{X}_1)|^{-\alpha-p}$ and $\phi_2(\tilde{X}_2) = |\det(\tilde{X}_2)|^\alpha$, the density of \tilde{U}_1 is

$$\tilde{g}_1(\tilde{U}_1) = \frac{1}{\tilde{\Gamma}_p(\alpha)} \int_{\tilde{V} < \tilde{U}_1} |\det(\tilde{U}_1 - \tilde{V})|^{\alpha-p} f(\tilde{V}) d\tilde{V}, \quad \Re(\alpha) > p - 1. \quad (5.7a.11)$$

The integral in (5.7a.11) is Weyl’s fractional integral of the first kind of order α in the complex matrix-variate case, denoted by $\tilde{W}_{1, \tilde{U}_1}^{-\alpha} f$. Observe that we are considering only Hermitian positive definite matrices. Thus, there is a lower bound, the integral being over $O < \tilde{V} < \tilde{U}_1$. Hence (5.7a.11) can also represent a Riemann-Liouville fractional integral of the first kind of order α in the complex matrix-variate case with a null matrix as its lower bound. For fractional integrals involving several matrices and fractional differential operators for functions of matrix argument, refer to Mathai (2014a, 2015); for pathway extensions, see Mathai and Haubold (2008, 2011).

Exercises 5.7

All the matrices appearing herein are $p \times p$ real positive definite, when real, and Hermitian positive definite, when in the complex domain. The M-transform of a real-valued scalar function $f(X)$ of the $p \times p$ real matrix X , with the M-transform parameter ρ , is defined as

$$M_f(\rho) = \int_{X > O} |X|^{\rho - \frac{p+1}{2}} f(X) dX, \quad \Re(\rho) > \frac{p-1}{2},$$

whenever the integral is convergent. In the real case, the M-convolution of a product $U_2 = X_2^{\frac{1}{2}} X_1 X_2^{\frac{1}{2}}$ with the corresponding functions $f_1(X_1)$ and $f_2(X_2)$, respectively, is

$$g_2(U_2) = \int_V |V|^{-\frac{p+1}{2}} f_1(V^{-\frac{1}{2}} U_2 V^{-\frac{1}{2}}) f_2(V) dV$$

whenever the integral is convergent. The M-convolution of a ratio in the real case is $g_1(U_1)$. The M-convolution of a product and a ratio in the complex case are $\tilde{g}_2(\tilde{U}_2)$ and $\tilde{g}_1(\tilde{U}_1)$, respectively, as defined earlier in this section. If α is the order of a fractional integral operator operating on f , denoted by $A^{-\alpha} f$, then the semigroup property is that $A^{-\alpha} A^{-\beta} f = A^{-(\alpha+\beta)} f = A^{-\beta} A^{-\alpha} f$.

- 5.7.1.** Show that the M-transform of the M-convolution of a product is the product of the M-transforms of the individual functions f_1 and f_2 , both in the real and complex cases.
- 5.7.2.** What are the M-transforms of the M-convolution of a ratio in the real and complex cases? Establish your assertions.
- 5.7.3.** Show that the semigroup property holds for Weyl's fractional integral of the (1): first kind, (2): second kind, in the real matrix-variate case.
- 5.7.4.** Do (1) and (2) of Exercise 5.7.3 hold in the complex matrix-variate case? Prove your assertion.
- 5.7.5.** Evaluate the M-transforms of the Erdélyi-Kober fractional integral of order α of (1): the first kind, (2): the second kind and state the conditions for their existence.
- 5.7.6.** Repeat Exercise 5.7.5 for (1) Weyl's fractional integral of order α , (2) the Riemann-Liouville fractional integral of order α .
- 5.7.7.** Evaluate the Weyl fractional integral of order α of (a): the first kind, (b): the second kind, in the real matrix-variate case, if possible, if the arbitrary function is (1): $e^{-\text{tr}(X)}$, (2): $e^{\text{tr}(X)}$ and write down the conditions wherever it is evaluated.
- 5.7.8.** Repeat Exercise 5.7.7 for the complex matrix-variate case.
- 5.7.9.** Evaluate the Erdélyi-Kober fractional integral of order α and parameter γ of the (a): first kind, (b): second kind, in the real matrix-variate case, if the arbitrary function is (1): $|X|^\delta$, (2): $|X|^{-\delta}$, wherever possible, and write down the necessary conditions.
- 5.7.10.** Repeat Exercise 5.7.9 for the complex case. In the complex case, $|X| = \text{determinant of } X$, is to be replaced by $|\det(\tilde{X})|$, the absolute value of the determinant of \tilde{X} .

5.8. Densities Involving Several Matrix-variate Random Variables, Real Case

We will start with real scalar variables. The most popular multivariate distribution, apart from the normal distribution, is the Dirichlet distribution, which is a generalization of the type-1 and type-2 beta distributions.

5.8.1. The type-1 Dirichlet density, real scalar case

Let x_1, \dots, x_k be real scalar random variables having a joint density of the form

$$f_1(x_1, \dots, x_k) = c_k x_1^{\alpha_1-1} \dots x_k^{\alpha_k-1} (1 - x_1 - \dots - x_k)^{\alpha_{k+1}-1} \quad (5.8.1)$$

for $\omega = \{(x_1, \dots, x_k) | 0 \leq x_j \leq 1, j = 1, \dots, k, 0 \leq x_1 + \dots + x_k \leq 1\}$, $\Re(\alpha_j) > 0$, $j = 1, \dots, k+1$ and $f_1 = 0$ elsewhere. This is type-1 Dirichlet density where c_k is the normalizing constant. Note that ω describes a simplex and hence the support of f_1 is

the simplex ω . Evaluation of the normalizing constant can be achieved in differing ways. One method relies on the direct integration of the variables, one at a time. For example, integration over x_1 involves two factors $x_1^{\alpha_1-1}$ and $(1 - x_1 - \dots - x_k)^{\alpha_{k+1}-1}$. Let I_1 be the integral over x_1 . Observe that x_1 varies from 0 to $1 - x_2 - \dots - x_k$. Then

$$I_1 = \int_{x_1=0}^{1-x_2-\dots-x_k} x_1^{\alpha_1-1} (1 - x_1 - \dots - x_k)^{\alpha_{k+1}-1} dx_1.$$

But

$$(1 - x_1 - \dots - x_k)^{\alpha_{k+1}-1} = (1 - x_2 - \dots - x_k)^{\alpha_{k+1}-1} \left[1 - \frac{x_1}{1 - x_2 - \dots - x_k} \right]^{\alpha_{k+1}-1}.$$

Make the substitution $y = \frac{x_1}{1-x_2-\dots-x_k} \Rightarrow dx_1 = (1 - x_2 - \dots - x_k)dy$, which enable one to integrate out y by making use of a real scalar type-1 beta integral giving

$$\int_0^1 y^{\alpha_1-1} (1 - y)^{\alpha_{k+1}-1} dy = \frac{\Gamma(\alpha_1)\Gamma(\alpha_{k+1})}{\Gamma(\alpha_1 + \alpha_{k+1})}$$

for $\Re(\alpha_1) > 0, \Re(\alpha_{k+1}) > 0$. Now, proceed similarly by integrating out x_2 from $x_2^{\alpha_2-1} (1 - x_2 - \dots - x_k)^{\alpha_1+\alpha_{k+1}}$, and continue in this manner until x_k is reached. Finally, after canceling out all common gamma factors, one has $\Gamma(\alpha_1) \dots \Gamma(\alpha_{k+1})/\Gamma(\alpha_1 + \dots + \alpha_{k+1})$ for $\Re(\alpha_j) > 0, j = 1, \dots, k + 1$. Thus, the normalizing constant is given by

$$c_k = \frac{\Gamma(\alpha_1 + \dots + \alpha_{k+1})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{k+1})}, \Re(\alpha_j) > 0, j = 1, \dots, k + 1. \tag{5.8.2}$$

Another method for evaluating the normalizing constant c_k consists of making the following transformation:

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= y_2(1 - y_1) \\ x_j &= y_j(1 - y_1)(1 - y_2) \dots (1 - y_{j-1}), j = 2, \dots, k \end{aligned} \tag{5.8.3}$$

It is then easily seen that

$$dx_1 \wedge \dots \wedge dx_k = (1 - y_1)^{k-1} (1 - y_2)^{k-2} \dots (1 - y_{k-1}) dy_1 \wedge \dots \wedge dy_k. \tag{5.8.4}$$

Under this transformation, one has

$$\begin{aligned} 1 - x_1 &= 1 - y_1 \\ 1 - x_1 - x_2 &= (1 - y_1)(1 - y_2) \\ 1 - x_1 - \dots - x_k &= (1 - y_1)(1 - y_2) \dots (1 - y_k). \end{aligned}$$

Then, we have

$$\begin{aligned} & x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1} (1 - x_1 - \cdots - x_k)^{\alpha_{k+1}-1} dx_1 \wedge \cdots \wedge dx_k \\ &= y_1^{\alpha_1-1} \cdots y_k^{\alpha_k-1} (1 - y_1)^{\alpha_2+\cdots+\alpha_{k+1}-1} \\ & \quad \times (1 - y_2)^{\alpha_3+\cdots+\alpha_{k+1}-1} \cdots (1 - y_k)^{\alpha_{k+1}-1} dy_1 \wedge \cdots \wedge dy_k. \end{aligned}$$

Now, all the y_j 's are separated and each one can be integrated out by making use of a type-1 beta integral. For example, the integrals over y_1, y_2, \dots, y_k give the following:

$$\begin{aligned} \int_0^1 y_1^{\alpha_1-1} (1 - y_1)^{\alpha_2+\cdots+\alpha_{k+1}-1} dy_1 &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})} \\ \int_0^1 y_2^{\alpha_2-1} (1 - y_2)^{\alpha_3+\cdots+\alpha_{k+1}-1} dy_2 &= \frac{\Gamma(\alpha_2)\Gamma(\alpha_3 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_2 + \cdots + \alpha_{k+1})} \\ &\vdots \\ \int_0^1 y_k^{\alpha_k-1} (1 - y_k)^{\alpha_{k+1}-1} dy_k &= \frac{\Gamma(\alpha_k)\Gamma(\alpha_{k+1})}{\Gamma(\alpha_k + \alpha_{k+1})} \end{aligned}$$

for $\Re(\alpha_j) > 0$, $j = 1, \dots, k + 1$. Taking the product produces c_k^{-1} .

5.8.2. The type-2 Dirichlet density, real scalar case

Let $x_1 > 0, \dots, x_k > 0$ be real scalar random variables having the joint density

$$f_2(x_1, \dots, x_k) = c_k x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1} (1 + x_1 + \cdots + x_k)^{-(\alpha_1+\cdots+\alpha_{k+1})} \quad (5.8.5)$$

for $x_j > 0$, $j = 1, \dots, k$, $\Re(\alpha_j) > 0$, $j = 1, \dots, k + 1$ and $f_2 = 0$ elsewhere, where c_k is the normalizing constant. This density is known as a type-2 Dirichlet density. It can be shown that the normalizing constant c_k is the same as the one obtained in (5.8.2) for the type-1 Dirichlet distribution. This can be established by integrating out the variables one at a time, starting with x_k or x_1 . This constant can also be evaluated with the help of the following transformation:

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= y_2(1 + y_1) \\ x_j &= y_j(1 + y_1)(1 + y_2) \cdots (1 + y_{j-1}), \quad j = 2, \dots, k, \end{aligned} \quad (5.8.6)$$

whose Jacobian is given by

$$dx_1 \wedge \cdots \wedge dx_k = (1 + y_1)^{k-1} (1 + y_2)^{k-2} \cdots (1 + y_{k-1}) dy_1 \wedge \cdots \wedge dy_k. \quad (5.8.7)$$

5.8.3. Some properties of Dirichlet densities in the real scalar case

Let us determine the h -th moment of $(1 - x_1 - \dots - x_k)$ in a type-1 Dirichlet density:

$$E[1 - x_1 - \dots - x_k]^h = \int_{\omega} (1 - x_1 - \dots - x_k)^h f_1(x_1, \dots, x_k) dx_1 \wedge \dots \wedge dx_k.$$

In comparison with the total integral, the only change is that the parameter α_{k+1} is replaced by $\alpha_{k+1} + h$; thus the result is available from the normalizing constant. That is,

$$E[1 - x_1 - \dots - x_k]^h = \frac{\Gamma(\alpha_{k+1} + h)}{\Gamma(\alpha_{k+1})} \frac{\Gamma(\alpha_1 + \dots + \alpha_{k+1})}{\Gamma(\alpha_1 + \dots + \alpha_{k+1} + h)}. \tag{5.8.8}$$

The additional condition needed is $\Re(\alpha_{k+1} + h) > 0$. Considering the structure of the moment in (5.8.8), $u = 1 - x_1 - \dots - x_k$ is manifestly a real scalar type-1 beta variable with the parameters $(\alpha_{k+1}, \alpha_1 + \dots + \alpha_k)$. This is stated in the following result:

Theorem 5.8.1. *Let x_1, \dots, x_k have a real scalar type-1 Dirichlet density with the parameters $(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Then, $u = 1 - x_1 - \dots - x_k$ has a real scalar type-1 beta distribution with the parameters $(\alpha_{k+1}, \alpha_1 + \dots + \alpha_k)$, and $1 - u = x_1 + \dots + x_k$ has a real scalar type-1 beta distribution with the parameters $(\alpha_1 + \dots + \alpha_k, \alpha_{k+1})$.*

Some parallel results can also be obtained for type-2 Dirichlet variables. Consider a real scalar type-2 Dirichlet density with the parameters $(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Let $v = (1 + x_1 + \dots + x_k)^{-1}$. Then, when taking the h -th moment of v , that is $E[v^h]$, we see that the only change is that α_{k+1} becomes $\alpha_{k+1} + h$. Accordingly, v has a real scalar type-1 beta distribution with the parameters $(\alpha_{k+1}, \alpha_1 + \dots + \alpha_k)$. Thus, $1 - v = \frac{x_1 + \dots + x_k}{1 + x_1 + \dots + x_k}$ is a type-1 beta random variables with the parameters interchanged. Hence the following result:

Theorem 5.8.2. *Let x_1, \dots, x_k have a real scalar type-2 Dirichlet density with the parameters $(\alpha_1, \dots, \alpha_k; \alpha_{k+1})$. Then $v = (1 + x_1 + \dots + x_k)^{-1}$ has a real scalar type-1 beta distribution with the parameters $(\alpha_{k+1}, \alpha_1 + \dots + \alpha_k)$ and $1 - v = \frac{x_1 + \dots + x_k}{1 + x_1 + \dots + x_k}$ has a real scalar type-1 beta distribution with the parameters $(\alpha_1 + \dots + \alpha_k, \alpha_{k+1})$.*

Observe that the joint product moments $E[x_1^{h_1} \dots x_k^{h_k}]$ can be determined both in the cases of real scalar type-1 Dirichlet and type-2 Dirichlet densities. This can be achieved by considering the corresponding normalizing constants. Since an arbitrary product moment will uniquely determine the corresponding distribution, one can show that all subsets of variables from the set $\{x_1, \dots, x_k\}$ are again real scalar type-1 Dirichlet and real scalar type-2 Dirichlet distributed, respectively; to identify the marginal joint density of a subset

under consideration, it suffices to set the complementary set of h_j 's equal to zero. Type-1 and type-2 Dirichlet densities enjoy many properties, some of which are mentioned in the exercises. As well, there exist several types of generalizations of the type-1 and type-2 Dirichlet models. The first author and his coworkers have developed several such models, one of which was introduced in connection with certain reliability problems.

5.8.4. Some generalizations of the Dirichlet models

Let the real scalar variables x_1, \dots, x_k have a joint density of the following type, which is a generalization of the type-1 Dirichlet density:

$$g_1(x_1, \dots, x_k) = b_k x_1^{\alpha_1-1} (1-x_1)^{\beta_1} x_2^{\alpha_2-1} (1-x_1-x_2)^{\beta_2} \dots \\ \times x_k^{\alpha_k-1} (1-x_1-\dots-x_k)^{\beta_k+\alpha_{k+1}-1} \quad (5.8.9)$$

for $(x_1, \dots, x_k) \in \omega$, $\Re(\alpha_j) > 0$, $j = 1, \dots, k+1$, as well as other necessary conditions to be stated later, and $g_1 = 0$ elsewhere, where b_k denotes the normalizing constant. This normalizing constant can be evaluated by integrating out the variables one at a time or by making the transformation (5.8.3) and taking into account its associated Jacobian as specified in (5.8.4). Under the transformation (5.8.3), y_1, \dots, y_k will be independently distributed with y_j having a type-1 beta density with the parameters (α_j, γ_j) , $\gamma_j = \alpha_{j+1} + \dots + \alpha_{k+1} + \beta_j + \dots + \beta_k$, $j = 1, \dots, k$, which yields the normalizing constant

$$b_k = \prod_{j=1}^k \frac{\Gamma(\alpha_j + \gamma_j)}{\Gamma(\alpha_j)\Gamma(\gamma_j)} \quad (5.8.10)$$

for $\Re(\alpha_j) > 0$, $j = 1, \dots, k+1$, $\Re(\gamma_j) > 0$, $j = 1, \dots, k$, where

$$\gamma_j = \alpha_{j+1} + \dots + \alpha_{k+1} + \beta_j + \dots + \beta_k, \quad j = 1, \dots, k. \quad (5.8.11)$$

Arbitrary moments $E[x_1^{h_1} \dots x_k^{h_k}]$ are available from the normalizing constant b_k by replacing α_j by $\alpha_j + h_j$ for $j = 1, \dots, k$ and then taking the ratio. It can be observed from this arbitrary moment that all subsets of the type (x_1, \dots, x_j) have a density of the type specified in (5.8.9). For other types of subsets, one has initially to rearrange the variables and the corresponding parameters by bringing them to the first j positions and then utilize the previous result on subsets.

The following model corresponding to (5.8.9) for the type-2 Dirichlet model was introduced by the first author:

$$g_2(x_1, \dots, x_k) = a_k x_1^{\alpha_1-1} (1+x_1)^{-\beta_1} x_2^{\alpha_2-1} (1+x_1+x_2)^{-\beta_2} \dots \\ \times x_k^{\alpha_k-1} (1+x_1+\dots+x_k)^{-(\alpha_1+\dots+\alpha_{k+1})-\beta_k} \quad (5.8.12)$$

for $x_j > 0, j = 1, \dots, k, \Re(\alpha_j) > 0, j = 1, \dots, k + 1$, as well as other necessary conditions to be stated later, and $g_2 = 0$ elsewhere. In order to evaluate the normalizing constant a_k , one can use the transformation (5.8.6) and its associated Jacobian given in (5.8.7). Then, y_1, \dots, y_k become independently distributed real scalar type-2 beta variables with the parameters (α_j, δ_j) , where

$$\delta_j = \alpha_1 + \dots + \alpha_{j-1} + \alpha_{k+1} + \beta_j + \dots + \beta_k \tag{5.8.13}$$

for $\Re(\alpha_j) > 0, j = 1, \dots, k + 1, \Re(\delta_j) > 0, j = 1, \dots, k$. Other generalizations are available in the literature.

5.8.5. A pseudo Dirichlet model

In the type-1 Dirichlet model, the support is the previously described simplex ω . We will now consider a model, which was recently introduced by the first author, wherein the variables can vary freely in a hypercube. Let us begin with the case $k = 2$. Consider the model

$$g_{12}(x_1, x_2) = c_{12} x_2^{\alpha_1} (1-x_1)^{\alpha_1-1} (1-x_2)^{\alpha_2-1} (1-x_1x_2)^{-(\alpha_1+\alpha_2-1)}, 0 \leq x_j \leq 1, \tag{5.8.14}$$

for $\Re(\alpha_j) > 0, j = 1, 2$, and $g_{12} = 0$ elsewhere. In this case, the variables are free to vary within the unit square. Let us evaluate the normalizing constant c_{12} . For this purpose, let us expand the last factor by making use of the binomial expansion since $0 < x_1x_2 < 1$. Then,

$$(1 - x_1x_2)^{-(\alpha_1+\alpha_2-1)} = \sum_{k=0}^{\infty} \frac{(\alpha_1 + \alpha_2 - 1)_k}{k!} x_1^k x_2^k \tag{i}$$

where for example the Pochhammer symbol $(a)_k$ stands for

$$(a)_k = a(a + 1) \dots (a + k - 1), a \neq 0, (a)_0 = 1.$$

Integral over x_1 gives

$$\int_0^1 x_1^k (1 - x_1)^{\alpha_1-1} dx_1 = \frac{\Gamma(k + 1)\Gamma(\alpha_1)}{\Gamma(\alpha_1 + k + 1)}, \Re(\alpha_1) > 0, \tag{ii}$$

and the integral over x_2 yields

$$\int_0^1 x_2^{\alpha_1+k} (1 - x_2)^{\alpha_2-1} dx_2 = \frac{\Gamma(\alpha_1 + k + 1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2 + 1)}. \tag{iii}$$

Taking the product of the right-hand side expressions in (ii) and (iii) and observing that $\Gamma(\alpha_1 + \alpha_2 + k + 1) = \Gamma(\alpha_1 + \alpha_2 + 1)(\alpha_1 + \alpha_2 + 1)_k$ and $\Gamma(k + 1) = (1)_k$, we obtain the following total integral:

$$\begin{aligned} & \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2 + 1)} \sum_{k=0}^{\infty} \frac{(1)_k(\alpha_1 + \alpha_2 - 1)_k}{k!(\alpha_1 + \alpha_2 + 1)_k} \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2 + 1)} {}_2F_1(1, \alpha_1 + \alpha_2 - 1; \alpha_1 + \alpha_2 + 1; 1) \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2 + 1)} \frac{\Gamma(\alpha_1 + \alpha_2 + 1)\Gamma(1)}{\Gamma(\alpha_1 + \alpha_2)\Gamma(2)} \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}, \Re(\alpha_1) > 0, \Re(\alpha_2) > 0, \end{aligned} \quad (5.8.15)$$

where the ${}_2F_1$ hypergeometric function with argument 1 is evaluated with the following identity:

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)} \quad (5.8.16)$$

whenever the gamma functions are defined. Observe that (5.8.15) is a surprising result as it is the total integral coming from a type-1 real beta density with the parameters (α_1, α_2) . Now, consider the general model

$$\begin{aligned} g_{1k}(x_1, \dots, x_k) &= c_{1k}(1 - x_1)^{\alpha_1 - 1} \dots (1 - x_k)^{\alpha_k - 1} x_1^{\alpha_1} \dots \\ &\quad \times x_k^{\alpha_1 + \dots + \alpha_k - 1} (1 - x_1 \dots x_k)^{-(\alpha_1 + \dots + \alpha_k - 1)}, \quad 0 \leq x_j \leq 1, \quad j = 1, \dots, k. \end{aligned} \quad (5.8.17)$$

Proceeding exactly as in the case of $k = 2$, one obtains the total integral as

$$[c_{1k}]^{-1} = \frac{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \dots + \alpha_k)}, \quad \Re(\alpha_j) > 0, \quad j = 1, \dots, k. \quad (5.8.18)$$

This is the total integral coming from a $(k - 1)$ -variate real type-1 Dirichlet model. Some properties of this distribution are pointed out in some of the assigned problems.

5.8.6. The type-1 Dirichlet model in real matrix-variate case

Direct generalizations of the real scalar variable Dirichlet models to real as well as complex matrix-variate cases are possible. The type-1 model will be considered first. Let the $p \times p$ real positive definite matrices X_1, \dots, X_k be such that $X_j > O$, $I - X_j > O$, that is X_j as well as $I - X_j$ are positive definite, for $j = 1, \dots, k$, and, in addition,

$I - X_1 - \dots - X_k > O$. Let $\Omega = \{(X_1, \dots, X_k) \mid O < X_j < I, j = 1, \dots, k, I - X_1 - \dots - X_k > O\}$. Consider the model

$$G_1(X_1, \dots, X_k) = C_k |X_1|^{\alpha_1 - \frac{p+1}{2}} \dots |X_k|^{\alpha_k - \frac{p+1}{2}} \times |I - X_1 - \dots - X_k|^{\alpha_{k+1} - \frac{p+1}{2}}, \quad (X_1, \dots, X_k) \in \Omega, \quad (5.8.19)$$

for $\Re(\alpha_j) > \frac{p-1}{2}, j = 1, \dots, k + 1$, and $G_1 = 0$ elsewhere. The normalizing constant C_k can be determined by using real matrix-variate type-1 beta integrals to integrate the matrices one at the time. We can also evaluate the total integral by means of the following transformation:

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= (I - Y_1)^{\frac{1}{2}} Y_2 (I - Y_1)^{\frac{1}{2}} \\ X_j &= (I - Y_1)^{\frac{1}{2}} \dots (I - Y_{j-1})^{\frac{1}{2}} Y_j (I - Y_{j-1})^{\frac{1}{2}} \dots (I - Y_1)^{\frac{1}{2}}, \quad j = 2, \dots, k. \end{aligned} \quad (5.8.20)$$

The associated Jacobian can then be determined by making use of results on matrix transformations that are provided in Sect. 1.6. Then,

$$dX_1 \wedge \dots \wedge dX_k = |I - Y_1|^{(k-1)(\frac{p+1}{2})} \dots |I - Y_{k-1}|^{\frac{p+1}{2}} dY_1 \wedge \dots \wedge dY_k. \quad (5.8.21)$$

It can be seen that the Y_j 's are independently distributed as real matrix-variate type-1 beta random variables and the product of the integrals gives the following final result:

$$C_k = \frac{\Gamma_p(\alpha_1 + \dots + \alpha_{k+1})}{\Gamma_p(\alpha_1) \dots \Gamma_p(\alpha_{k+1})}, \quad \Re(\alpha_j) > \frac{p-1}{2}, \quad j = 1, \dots, k + 1. \quad (5.8.22)$$

By integrating out the variables one at a time, we can show that the marginal densities of all subsets of $\{X_1, \dots, X_k\}$ also belong to the same real matrix-variate type-1 Dirichlet distribution and single matrices are real matrix-variate type-1 beta distributed. By taking the product moment of the determinants, $E[|X_1|^{h_1} \dots |X_k|^{h_k}]$, one can anticipate the results; however, arbitrary moments of determinants need not uniquely determine the densities of the corresponding matrices. In the real scalar case, one can uniquely identify the density from arbitrary moments, very often under very mild conditions. The result $I - X_1 - \dots - X_k$ has a real matrix-variate type-1 beta distribution can be seen by taking arbitrary moments of the determinant, that is, $E[|I - X_1 - \dots - X_k|^h]$, but evaluating the h -moment of a determinant and then identifying it as the h -th moment of the determinant from a real matrix-variate type-1 beta density is not valid in this case. If one makes a transformation of the type $Y_1 = X_1, \dots, Y_{k-1} = X_{k-1}, Y_k = I - X_1 - \dots - X_k$,

it is seen that $X_k = I - Y_1 - \dots - Y_k$ and that the Jacobian in absolute value is 1. Hence, we end up with a real matrix-variate type-1 Dirichlet density of the same format but whose parameters α_k and α_{k+1} are interchanged. Then, integrating out Y_1, \dots, Y_{k-1} , we obtain a real matrix-variate type-1 beta density with the parameters $(\alpha_{k+1}, \alpha_1 + \dots + \alpha_k)$. Hence the result. When Y_k has a real matrix-variate type-1 beta distribution, we have that $I - Y_k = X_1 + \dots + X_k$ is also a type-1 beta random variable with the parameters interchanged.

The first author and his coworkers have proposed various types of generalizations to the matrix-variate type-1 and type-2 Dirichlet models in the real and complex cases. One of those extensions which is defined in the real domain, is the following:

$$\begin{aligned} G_2(X_1, \dots, X_k) &= C_{1k} |X_1|^{\alpha_1 - \frac{p+1}{2}} |I - X_1|^{\beta_1} |X_2|^{\alpha_2 - \frac{p+1}{2}} \\ &\quad \times |I - X_1 - X_2|^{\beta_2} \dots |X_k|^{\alpha_k - \frac{p+1}{2}} \\ &\quad \times |I - X_1 - \dots - X_k|^{\alpha_{k+1} + \beta_k - \frac{p+1}{2}}, \end{aligned} \quad (5.8.23)$$

for $(X_1, \dots, X_k) \in \Omega$, $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, \dots, k+1$, and $G_2 = 0$ elsewhere. The normalizing constant C_{1k} can be evaluated by integrating variables one at a time or by using the transformation (5.8.20). Under this transformation, the real matrices Y_j 's are independently distributed as real matrix-variate type-1 beta variables with the parameters (α_j, γ_j) , $\gamma_j = \alpha_{j+1} + \dots + \alpha_{k+1} + \beta_j + \dots + \beta_k$. The conditions will then be $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, \dots, k+1$, and $\Re(\gamma_j) > \frac{p-1}{2}$, $j = 1, \dots, k$. Hence

$$C_{1k} = \prod_{j=1}^k \frac{\Gamma_p(\alpha_j + \gamma_j)}{\Gamma_p(\alpha_j) \Gamma_p(\gamma_j)}. \quad (5.8.24)$$

5.8.7. The type-2 Dirichlet model in the real matrix-variate case

The type-2 Dirichlet density in the real matrix-variate case is the following:

$$\begin{aligned} G_3(X_1, \dots, X_k) &= C_k |X_1|^{\alpha_1 - \frac{p+1}{2}} \dots |X_k|^{\alpha_k - \frac{p+1}{2}} \\ &\quad \times |I + X_1 + \dots + X_k|^{-(\alpha_1 + \dots + \alpha_{k+1})}, \quad X_j > O, \quad j = 1, \dots, k, \end{aligned} \quad (5.8.25)$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, \dots, k+1$ and $G_3 = 0$ elsewhere, the normalizing constant C_k being the same as that appearing in the type-1 Dirichlet case. This can be verified, either by integrating matrices one at a time from (5.8.25) or by making the following transformation:

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= (I + Y_1)^{\frac{1}{2}} Y_2 (I + Y_1)^{\frac{1}{2}} \\ X_j &= (I + Y_1)^{\frac{1}{2}} \dots (I + Y_{j-1})^{\frac{1}{2}} Y_j (I + Y_{j-1})^{\frac{1}{2}} \dots (I + Y_1)^{\frac{1}{2}}, \quad j = 2, \dots, k. \end{aligned} \quad (5.8.26)$$

Under this transformation, the Jacobian is as follows:

$$dX_1 \wedge \dots \wedge dX_k = |I + Y_1|^{(k-1)\left(\frac{p+1}{2}\right)} \dots |I + Y_{k-1}|^{\frac{p+1}{2}} dY_1 \wedge \dots \wedge dY_k. \quad (5.8.27)$$

Thus, the Y_j 's are independently distributed real matrix-variate type-2 beta variables and the product of the integrals produces $[C_k]^{-1}$. By integrating matrices one at a time, we can see that all subsets of matrices belonging to $\{X_1, \dots, X_k\}$ will have densities of the type specified in (5.8.25). Several properties can also be established for the model (5.8.25); some of them are included in the exercises.

Example 5.8.1. Evaluate the normalizing constant c explicitly if the function $f(X_1, X_2)$ is a statistical density where the $p \times p$ real matrices $X_j > O$, $I - X_j > O$, $j = 1, 2$, and $I - X_1 - X_2 > O$ where

$$f(X_1, X_2) = c |X_1|^{\alpha_1 - \frac{p+1}{2}} |I - X_1|^{\beta_1} |X_2|^{\alpha_2 - \frac{p+1}{2}} |I - X_1 - X_2|^{\beta_2 - \frac{p+1}{2}}.$$

Solution 5.8.1. Note that

$$|I - X_1 - X_2|^{\beta_2 - \frac{p+1}{2}} = |I - X_1|^{\beta_2 - \frac{p+1}{2}} |I - (I - X_1)^{-\frac{1}{2}} X_2 (I - X_1)^{-\frac{1}{2}}|^{\beta_2 - \frac{p+1}{2}}.$$

Now, letting $Y = (I - X_1)^{-\frac{1}{2}} X_2 (I - X_1)^{-\frac{1}{2}} \Rightarrow dY = |I - X_1|^{-\frac{p+1}{2}} dX_2$, and the integral over X_2 gives the following:

$$|I - X_1|^{\alpha_2 + \beta_2 - \frac{p+1}{2}} \int_{O < Y < I} |Y|^{\alpha_2 - \frac{p+1}{2}} |I - Y|^{\beta_2 - \frac{p+1}{2}} dY = |I - X_1|^{\alpha_2 + \beta_2 - \frac{p+1}{2}} \frac{\Gamma_p(\alpha_2) \Gamma_p(\beta_2)}{\Gamma_p(\alpha_2 + \beta_2)}$$

for $\Re(\alpha_2) > \frac{p-1}{2}$, $\Re(\beta_2) > \frac{p-1}{2}$. Then, the integral over X_1 can be evaluated as follows:

$$\int_{O < X_1 < I} |X_1|^{\alpha_1 - \frac{p+1}{2}} |I - X_1|^{\beta_1 + \beta_2 + \alpha_2 - \frac{p+1}{2}} dX_1 = \frac{\Gamma_p(\alpha_1) \Gamma_p(\beta_1 + \beta_2 + \alpha_2)}{\Gamma_p(\alpha_1 + \alpha_2 + \beta_1 + \beta_2)}$$

for $\Re(\alpha_1) > \frac{p-1}{2}$, $\Re(\beta_1 + \beta_2 + \alpha_2) > \frac{p-1}{2}$. Collecting the results from the integrals over X_2 and X_1 and using the fact that the total integral is 1, we have

$$c = \frac{\Gamma_p(\alpha_2 + \beta_2) \Gamma_p(\alpha_1 + \alpha_2 + \beta_1 + \beta_2)}{\Gamma_p(\alpha_2) \Gamma_p(\beta_2) \Gamma_p(\alpha_1) \Gamma_p(\alpha_2 + \beta_1 + \beta_2)}$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, $\Re(\beta_2) > \frac{p-1}{2}$, and $\Re(\beta_1 + \beta_2 + \alpha_2) > \frac{p-1}{2}$.

The first author and his coworkers have established several generalizations to the type-2 Dirichlet model in (5.8.25). One such model is the following:

$$\begin{aligned} G_4(X_1, \dots, X_k) &= C_{2k} |X_1|^{\alpha_1 - \frac{p+1}{2}} |I + X_1|^{-\beta_1} |X_2|^{\alpha_2 - \frac{p+1}{2}} \\ &\quad \times |I + X_1 + X_2|^{-\beta_2} \dots |X_k|^{\alpha_k - \frac{p+1}{2}} \\ &\quad \times |I + X_1 + \dots + X_k|^{-(\alpha_1 + \dots + \alpha_{k+1} + \beta_k)} \end{aligned} \quad (5.8.28)$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, \dots, k+1$, $X_j > O$, $j = 1, \dots, k$, as well as other necessary conditions to be stated later, and $G_4 = 0$ elsewhere. The normalizing constant C_{2k} can be evaluated by integrating matrices one at a time or by making the transformation (5.8.26). Under this transformation, the Y_j 's are independently distributed real matrix-variate type-2 beta variables with the parameters (α_j, δ_j) , where

$$\delta_j = \alpha_1 + \dots + \alpha_{j-1} + \beta_j + \dots + \beta_k. \quad (5.8.29)$$

The normalizing constant is then

$$G_{2k} = \prod_{j=1}^k \frac{\Gamma_p(\alpha_j + \delta_j)}{\Gamma_p(\alpha_j) \Gamma_p(\delta_j)} \quad (5.8.30)$$

where the δ_j is given in (5.8.29). The marginal densities of the subsets, if taken in the order X_1 , $\{X_1, X_2\}$, and so on, will belong to the same family of densities as that specified by (5.8.28). Several properties of the model (5.8.28) are available in the literature.

5.8.8. A pseudo Dirichlet model

We will now discuss the generalization of the model introduced in Sect. 5.8.5. Consider the density

$$\begin{aligned} G_{1k}(X_1, \dots, X_k) &= C_{1k} |I - X_1|^{\alpha_1 - \frac{p+1}{2}} \dots |I - X_k|^{\alpha_k - \frac{p+1}{2}} \\ &\quad \times |X_2|^{\alpha_1} |X_3|^{\alpha_1 + \alpha_2} \dots |X_k|^{\alpha_1 + \dots + \alpha_{k-1}} \\ &\quad \times |I - X_k^{\frac{1}{2}} \dots X_2^{\frac{1}{2}} X_1 X_2^{\frac{1}{2}} \dots X_k^{\frac{1}{2}}|^{-(\alpha_1 + \dots + \alpha_k - \frac{p+1}{2})}. \end{aligned} \quad (5.8.31)$$

Then, by following steps parallel to those used in the real scalar variable case, one can show that the normalizing constant is given by

$$C_{1k} = \frac{\Gamma_p(\alpha_1 + \dots + \alpha_k) \Gamma_p(p+1)}{\Gamma_p(\alpha_1) \dots \Gamma_p(\alpha_k) [\Gamma_p(\frac{p+1}{2})]^2}. \quad (5.8.32)$$

The binomial expansion of the last factor determinant in (5.8.31) is somewhat complicated as it involves zonal polynomials; this expansion is given in Mathai (1997). Compared to the real scalar case, the only change is the appearance of the constant $\frac{\Gamma_p(p+1)}{[\Gamma_p(\frac{p+1}{2})]^2}$ which

is 1 when $p = 1$. Apart from this constant, the rest is the normalizing constant in a real matrix-variate type-1 Dirichlet model in $k - 1$ variables instead of k variables.

5.8a. Dirichlet Models in the Complex Domain

All the matrices appearing in the remainder of this chapter are $p \times p$ Hermitian positive definite, that is, $\tilde{X}_j = \tilde{X}_j^*$ where an asterisk indicates the conjugate transpose. Complex matrix-variate random variables will be denoted with a tilde. For a complex matrix \tilde{X} , the determinant will be denoted by $\det(\tilde{X})$ and the absolute value of the determinant, by $|\det(\tilde{X})|$. For example, if $\det(\tilde{X}) = a + ib$, a and b being real and $i = \sqrt{-1}$, the absolute value is $|\det(\tilde{X})| = +(a^2 + b^2)^{\frac{1}{2}}$. The type-1 Dirichlet model in the complex domain, denoted by \tilde{G}_1 , is the following:

$$\tilde{G}_1(X_1, \dots, X_k) = \tilde{C}_k |\det(\tilde{X}_1)|^{\alpha_1 - p} \dots |\det(\tilde{X}_k)|^{\alpha_k - p} \times |\det(I - \tilde{X}_1 - \dots - \tilde{X}_k)|^{\alpha_{k+1} - p} \tag{5.8a.1}$$

for $(\tilde{X}_1, \dots, \tilde{X}_k) \in \tilde{\Omega}$, $\tilde{\Omega} = \{(\tilde{X}_1, \dots, \tilde{X}_k) | 0 < \tilde{X}_j < I, j = 1, \dots, k, 0 < \tilde{X}_1 + \dots + \tilde{X}_k < I\}$, $\Re(\alpha_j) > p - 1, j = 1, \dots, k + 1$, and $\tilde{G}_1 = 0$ elsewhere. The normalizing constant \tilde{C}_k can be evaluated by integrating out matrices one at a time with the help of complex matrix-variate type-1 beta integrals. One can also employ a transformation of the type given in (5.8.20) where the real matrices are replaced by matrices in the complex domain and Hermitian positive definite square roots are used. The Jacobian is then as follows:

$$d\tilde{X}_1 \wedge \dots \wedge d\tilde{X}_k = |\det(I - \tilde{Y}_1)|^{(k-1)p} \dots |\det(I - Y_{k-1})|^p d\tilde{Y}_1 \wedge \dots \wedge d\tilde{Y}_k. \tag{5.8a.2}$$

Then \tilde{Y}_j 's are independently distributed as complex matrix-variate type-1 beta variables. On taking the product of the total integrals, one can verify that

$$\tilde{C}_k = \frac{\tilde{\Gamma}_p(\alpha_1 + \dots + \alpha_{k+1})}{\tilde{\Gamma}_p(\alpha_1) \dots \tilde{\Gamma}_p(\alpha_{k+1})} \tag{5.8a.3}$$

where for example $\tilde{\Gamma}_p(\alpha)$ is the complex matrix-variate gamma given by

$$\tilde{\Gamma}_p(\alpha) = \pi^{\frac{p(p-1)}{2}} \Gamma(\alpha) \Gamma(\alpha - 1) \dots \Gamma(\alpha - p + 1), \Re(\alpha) > p - 1. \tag{5.8a.4}$$

The first author and his coworkers have also discussed various types of generalizations to Dirichlet models in complex domain.

5.8a.1. A type-2 Dirichlet model in the complex domain

One can have a model parallel to the type-2 Dirichlet model in the real matrix-variate case. Consider the model

$$\begin{aligned} \tilde{G}_2 = & \tilde{C}_k |\det(\tilde{X}_1)|^{\alpha_1 - p} \cdots |\det(\tilde{X}_k)|^{\alpha_k - p} \\ & \times |\det(I + \tilde{X}_1 + \cdots + \tilde{X}_k)|^{-(\alpha_1 + \cdots + \alpha_{k+1})} \end{aligned} \quad (5.8a.5)$$

for $\tilde{X}_j > O$, $j = 1, \dots, k$, $\Re(\alpha_j) > p - 1$, $j = 1, \dots, k + 1$, and $\tilde{G}_2 = 0$ elsewhere. By integrating out matrices one at a time with the help of complex matrix-variate type-2 integrals or by using a transformation parallel to that provided in (5.7.26) and then integrating out the independently distributed complex type-2 beta variables \tilde{Y}_j 's, we can show that the normalizing constant \tilde{C}_k is the same as that obtained in the complex type-1 Dirichlet case. The first author and his coworkers have given various types of generalizations to the complex type-2 Dirichlet density as well.

Exercises 5.8

5.8.1. By integrating out variables one at a time derive the normalizing constant in the real scalar type-2 Dirichlet case.

5.8.2. By using the transformation (5.8.3), derive the normalizing constant in the real scalar type-1 Dirichlet case.

5.8.3. By using the transformation in (5.8.6), derive the normalizing constant in the real scalar type-2 Dirichlet case.

5.8.4. Derive the normalizing constants for the extended Dirichlet models in (5.8.9) and (5.8.12).

5.8.5. Evaluate $E[x_1^{h_1} \cdots x_k^{h_k}]$ for the model specified in (5.8.12) and state the conditions for its existence.

5.8.6. Derive the normalizing constant given in (5.8.18).

5.8.7. With respect to the pseudo Dirichlet model in (5.8.17), show that the product $u = x_1 \cdots x_k$ is uniformly distributed.

5.8.8. Derive the marginal distribution of (1): x_1 ; (2): (x_1, x_2) ; (3): (x_1, \dots, x_r) , $r < k$, and the conditional distribution of (x_1, \dots, x_r) given (x_{r+1}, \dots, x_k) in the pseudo Dirichlet model in (5.8.17).

5.8.9. Derive the normalizing constant in (5.8.22) by completing the steps in (5.8.22) and then by integrating out matrices one by one.

5.8.10. From the outline given after equation (5.8.22), derive the density of $I - X_1 - \dots - X_k$ and therefrom the density of $X_1 + \dots + X_k$ when (X_1, \dots, X_k) has a type-1 Dirichlet distribution.

5.8.11. Complete the derivation of C_{1k} in (5.8.24) and verify it by integrating out matrices one at a time from the density given in (5.8.23).

5.8.12. Show that $U = (I + X_1 + \dots + X_k)^{-1}$ in the type-2 Dirichlet model in (5.8.25) is a real matrix-variate type-1 beta distributed. As well, specify its parameters.

5.8.13. Evaluate the normalizing constant C_k in (5.8.25) by using the transformation provided in (5.8.26) as well as by integrating out matrices one at a time.

5.8.14. Derive the δ_j in (5.8.29) and thus the normalizing constant C_{2k} in (5.8.28).

5.8.15. For the following model in the complex domain, evaluate C :

$$f(\tilde{X}) = C |\det(\tilde{X}_1)|^{\alpha_1 - p} |\det(I - \tilde{X}_1)|^{\beta_1} |\det(\tilde{X}_2)|^{\alpha_2 - p} |\det(I - \tilde{X}_1 - \tilde{X}_2)|^{\beta_2} \dots \\ \times |\det(I - \tilde{X}_1 - \dots - \tilde{X}_k)|^{\alpha_{k+1} - p + \beta_k}.$$

5.8.16. Evaluate the normalizing constant in the pseudo Dirichlet model in (5.8.31).

5.8.17. In the pseudo Dirichlet model specified in (5.8.31), show that $U = X_k^{\frac{1}{2}} \dots X_2^{\frac{1}{2}} X_1 X_2^{\frac{1}{2}} \dots X_k^{\frac{1}{2}}$ is uniformly distributed.

5.8.18. Show that the normalizing constant in the complex type-2 Dirichlet model specified in (5.8a.5) is the same as the one in the type-1 Dirichlet case. Establish the result by integrating out matrices one by one.

5.8.19. Show that the normalizing constant in the type-2 Dirichlet case in (5.8a.5) is the same as that in the type-1 case. Establish this by using a transformation parallel to (5.8.26) in the complex domain.

5.8.20. Construct a generalized model for the type-2 Dirichlet case for $k = 3$ parallel to the case in (5.8.28) in the complex domain.

References

Mathai, A.M. (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

Mathai, A.M. (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

Mathai, A.M. (1999): *Introduction to Geometrical Probabilities: Distributional Aspects and Applications*, Gordon and Breach, Amsterdam.

Mathai, A.M. (2005): A pathway to matrix-variate gamma and normal densities, *Linear Algebra and its Applications*, **396**, 317–328.

Mathai, A.M. (2009): Fractional integrals in the matrix-variate case and connection to statistical distributions, *Integral Transforms and Special Functions*, **20(12)**, 871–882.

Mathai, A.M. (2010): Some properties of Mittag-Leffler functions and matrix-variate analogues: A statistical perspective, *Fractional Calculus & Applied Analysis*, **13(2)**, 113–132.

Mathai, A.M. (2012): Generalized Krätzel integral and associated statistical densities, *International Journal of Mathematical Analysis*, **6(51)**, 2501–2510.

Mathai, A.M. (2013): Fractional integral operators in the complex matrix-variate case, *Linear Algebra and its Applications*, **439**, 2901–2913.

Mathai, A.M. (2014): Evaluation of matrix-variate gamma and beta integrals, *Applied Mathematics and computations*, **247**, 312–318.

Mathai, A.M. (2014a): Fractional integral operators involving many matrix variables, *Linear Algebra and its Applications*, **446**, 196–215.

Mathai, A.M. (2014b): Explicit evaluations of gamma and beta integrals in the matrix-variate case, *Journal of the Indian Mathematical Society*, **81(3)**, 259–271.

Mathai, A.M. (2015): Fractional differential operators in the complex matrix-variate case, *Linear Algebra and its Applications*, **478**, 200–217.

Mathai, A.M. and H.J. Haubold, H.J. (1988) *Modern Problems in Nuclear and Neutrino Astrophysics*, Akademie-Verlag, Berlin.

Mathai, A.M. and Haubold, H.J. (2008): *Special Functions for Applied Scientists*, Springer, New York.

Mathai, A.M. and Haubold, H.J. (2011): A pathway from Bayesian statistical analysis to superstatistics, *Applied Mathematics and Computations*, **218**, 799–804.

Mathai, A.M. and Haubold, H.J. (2011a): Matrix-variate statistical distributions and fractional calculus, *Fractional Calculus & Applied Analysis*, **24(1)**, 138–155.

Mathai, A.M. and Haubold, H.J. (2017): *Introduction to Fractional Calculus*, Nova Science Publishers, New York.

Mathai, A.M. and Haubold, H.J. (2017a): *Fractional and Multivariable Calculus: Model Building and Optimization*, Springer, New York.

Mathai, A.M. and Haubold, H.J. (2017b): *Probability and Statistics*, De Gruyter, Germany.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 6

Hypothesis Testing and Null Distributions



6.1. Introduction

It is assumed that the readers are familiar with the concept of testing statistical hypotheses on the parameters of a real scalar normal density or independent real scalar normal densities. Those who are not or require a refresher may consult the textbook: Mathai and Haubold (2017) on basic “Probability and Statistics” [De Gruyter, Germany, 2017, free download]. Initially, we will only employ the likelihood ratio criterion for testing hypotheses on the parameters of one or more real multivariate Gaussian (or normal) distributions. All of our tests will be based on a simple random sample of size n from a p -variate nonsingular Gaussian distribution, that is, the $p \times 1$ vectors X_1, \dots, X_n constituting the sample are iid (independently and identically distributed) as $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, when a single real Gaussian population is involved. The corresponding test criterion for the complex Gaussian case will also be mentioned in each section.

In this chapter, we will utilize the following notations. Lower-case letters such as x, y will be used to denote real scalar mathematical or random variables. No distinction will be made between mathematical and random variables. Capital letters such as X, Y will denote real vector/matrix-variate variables, whether mathematical or random. A tilde placed on a letter as for instance $\tilde{x}, \tilde{y}, \tilde{X}$ and \tilde{Y} will indicate that the variables are in the complex domain. No tilde will be used for constant matrices unless the point is to be stressed that the matrix concerned is in the complex domain. The other notations will be identical to those utilized in the previous chapters.

First, we consider certain problems related to testing hypotheses on the parameters of a p -variate real Gaussian population. Only the likelihood ratio criterion, also referred to as λ -criterion, will be utilized. Let L denote the joint density of the sample values in a simple random sample of size n , namely, X_1, \dots, X_n , which are iid $N_p(\mu, \Sigma)$, $\Sigma > O$. Then, as was previously established,

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}, \quad (6.1.1)$$

where $S = \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'$ is the sample sum of products matrix and $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is the sample average, n being the sample size. As well, we have already determined that the maximum likelihood estimators (MLE's) of μ and Σ are $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \frac{1}{n}S$, the sample covariance matrix. Consider the parameter space

$$\Omega = \{(\mu, \Sigma) \mid \Sigma > O, \mu' = (\mu_1, \dots, \mu_p), -\infty < \mu_j < \infty, j = 1, \dots, p\}.$$

The maximum value of L within Ω is obtained by substituting the MLE's of the parameters into L , and since $(\bar{X} - \hat{\mu}) = (\bar{X} - \bar{X}) = O$ and $\text{tr}(\hat{\Sigma}^{-1} S) = \text{tr}(nI_p) = np$,

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |\frac{1}{n}S|^{\frac{n}{2}}} = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}. \quad (6.1.2)$$

Under any given hypothesis on μ or Σ , the parameter space is reduced to a subspace ω in Ω or $\omega \subset \Omega$. For example, if $H_o : \mu = \mu_o$ where μ_o is a given vector, then the parameter space under this null hypothesis reduces to $\omega = \{(\mu, \Sigma) \mid \mu = \mu_o, \Sigma > O\} \subset \Omega$, “null hypothesis” being a technical term used to refer to the hypothesis being tested. The alternative hypothesis against which the null hypothesis is tested, is usually denoted by H_1 . If $\mu = \mu_o$ specifies H_o , then a natural alternative is $H_1 : \mu \neq \mu_o$. One of two things can happen when considering the maximum of the likelihood function under H_o . The overall maximum may occur in ω or it may be attained outside of ω but inside Ω . If the null hypothesis H_o is actually true, then ω and Ω will coincide and the maxima in ω and in Ω will agree. If there are several local maxima, then the overall maximum or supremum is taken. The λ -criterion is defined as follows:

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L}, \quad 0 < \lambda \leq 1. \quad (6.1.3)$$

If the null hypothesis is true, then $\lambda = 1$. Accordingly, an observed value of λ that is close to 0 in a testing situation indicates that the null hypothesis H_o is incorrect and should then be rejected. Hence, the test criterion under the likelihood ratio test is to “reject H_o for $0 < \lambda \leq \lambda_o$ ”, that is, for small values of λ , so that, under H_o , the coverage probability over this interval is equal to the significance level α or the probability of rejecting H_o when H_o is true, that is, $Pr\{0 < \lambda \leq \lambda_o \mid H_o\} = \alpha$ for a pre-assigned α , which is also known as the size of the critical region or the size of the type-1 error. However, rejecting H_o when it is not actually true or when the alternative H_1 is true is a correct decision

whose probability is known as the power of the test and written as $1 - \beta$ where β is the probability of committing a type-2 error or the error of not rejecting H_o when H_o is not true. Thus we have

$$Pr\{0 < \lambda \leq \lambda_o | H_o\} = \alpha \text{ and } Pr\{0 < \lambda \leq \lambda_o | H_1\} = 1 - \beta. \quad (6.1.4)$$

When we preassign $\alpha = 0.05$, we are allowing a tolerance of 5% for the probability of committing the error of rejecting H_o when it is actually true and we say that we have a test at the 5% significance level. Usually, we set α as 0.05 or 0.01. Alternatively, we can allow α to vary and calculate what is known as the p -value when carrying out a test. Such is the principle underlying the likelihood ratio test, the resulting test criterion being referred to as the λ -criterion.

In the complex case, a tilde will be placed above λ and L , (6.1.3) and (6.1.4) remaining essentially the same:

$$\tilde{\lambda} = \frac{\sup_{\omega} \tilde{L}}{\sup_{\Omega} \tilde{L}}, \quad 0 < |\tilde{\lambda}| \leq 1, \quad (6.1a.1)$$

and

$$Pr\{0 < |\tilde{\lambda}| \leq \lambda_o | H_o\} = \alpha, \quad Pr\{0 < |\tilde{\lambda}| \leq \lambda_o | H_1\} = 1 - \beta \quad (6.1a.2)$$

where α is the size or significance level of the test and $1 - \beta$, the power of the test.

6.2. Testing $H_o : \mu = \mu_o$ (Given) When Σ is Known, the Real $N_p(\mu, \Sigma)$ Case

When Σ is known, the only parameter to estimate is μ , its MLE being \bar{X} . Hence, the maximum in Ω is the following:

$$\sup_{\Omega} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (6.2.1)$$

In this case, μ is also specified under the null hypothesis H_o , so that there is no parameter to estimate. Accordingly,

$$\begin{aligned} \sup_{\omega} L &= \frac{e^{-\frac{1}{2} \sum_{j=1}^n (X_j - \mu_o)' \Sigma^{-1} (X_j - \mu_o)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \\ &= \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \end{aligned} \quad (6.2.2)$$

Thus,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = e^{-\frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)}, \quad (6.2.3)$$

and small values of λ correspond to large values of $\frac{n}{2}(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)$. When $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, it has already been established that $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$, $\Sigma > O$. As well, $n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o)$ is the exponent in a p -variate real normal density under H_o , which has already been shown to have a real chisquare distribution with p degrees of freedom or

$$n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o) \sim \chi_p^2.$$

Hence, the test criterion is

$$\text{Reject } H_o \text{ if } n(\bar{X} - \mu_o)' \Sigma^{-1}(\bar{X} - \mu_o) \geq \chi_{p,\alpha}^2, \text{ with } Pr\{\chi_p^2 \geq \chi_{p,\alpha}^2\} = \alpha. \quad (6.2.4)$$

Under the alternative hypothesis, the distribution of the test statistic is a noncentral chisquare with p degrees of freedom and non-centrality parameter $\lambda = \frac{n}{2}(\mu - \mu_o)' \Sigma^{-1}(\mu - \mu_o)$.

Example 6.2.1. For example, suppose that we have a sample of size 5 from a population that has a trivariate normal distribution and let the significance level α be 0.05. Let μ_o , the hypothesized mean value vector specified by the null hypothesis, the known covariance matrix Σ , and the five observation vectors X_1, \dots, X_5 be the following:

$$\mu_o = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix},$$

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}, \quad X_4 = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \quad X_5 = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix},$$

the inverse of Σ having been evaluated via elementary transformations. The sample average, $\frac{1}{5}(X_1 + \dots + X_5)$ denoted by \bar{X} , is

$$\bar{X} = \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix},$$

and

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix}.$$

For testing H_o , the following test statistic has to be evaluated:

$$n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) = \frac{5}{5^2} [4 \quad 4 \quad 8] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 8 \end{bmatrix} = \frac{40}{5} = 8.$$

As per our criterion, H_o should be rejected if $8 \geq \chi_{p,\alpha}^2$. Since $\chi_{p,\alpha}^2 = \chi_{3,0.05}^2 = 7.81$, this critical value being available from a chisquare table, $H_o : \mu = \mu_o$ should be rejected at the specified significance level. Moreover, in this case, the p -value is $Pr\{\chi_3^2 \geq 8\} \approx 0.035$, which can be evaluated by interpolation from the percentiles provided in a chi-square table or by making use of statistical packages such as R.

6.2.1. Paired variables and linear functions

Let Y_1, \dots, Y_k be $p \times 1$ vectors having their own p -variate distributions which are not known. However, suppose that a certain linear function $X = a_1 Y_1 + \dots + a_k Y_k$ is known to have a p -variate real Gaussian distribution with mean value vector $E[X] = \mu$ and covariance matrix $\text{Cov}(X) = \Sigma$, $\Sigma > O$, that is, $X = a_1 Y_1 + \dots + a_k Y_k \sim N_p(\mu, \Sigma)$, $\Sigma > O$, where a_1, \dots, a_k are fixed known scalar constants. An example of this type is $X = Y_1 - Y_2$ where Y_1 consists of measurements on p attributes before subjecting those attributes to a certain process, such as administering a drug to a patient, and Y_2 consists of the measurements on the same attributes after the process is completed. We would like to examine the difference $Y_1 - Y_2$ to study the effect of the process on these characteristics. If it is reasonable to assume that this difference $X = Y_1 - Y_2$ is $N_p(\mu, \Sigma)$, $\Sigma > O$, then we could test hypotheses on $E[X] = \mu$. When Σ is known, the general problem reduces to that discussed in Sect. 6.2. Assuming that we have iid variables on Y_1, \dots, Y_k , we would evaluate the corresponding values of X , which produces iid variables on X , that is, a simple random sample of size n from $X = a_1 Y_1 + \dots + a_k Y_k$. Thus, when Σ is known, letting $u = n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) \sim \chi_p^2$ where \bar{X} denote the sample average, the test would be carried out as follows at significance level α :

$$\text{Reject } H_o : \mu = \mu_o \text{ (specified) when } u \geq \chi_{p,\alpha}^2, \text{ with } Pr\{\chi_p^2 \geq \chi_{p,\alpha}^2\} = \alpha, \quad (6.2.5)$$

the non-null distribution of the test statistic u being a non-central chisquare.

Example 6.2.2. Three variables $x_1 =$ systolic pressure, $x_2 =$ diastolic pressure and $x_3 =$ weight are monitored after administering a drug for the reduction of all these $p = 3$ variables. Suppose that a sample of $n = 5$ randomly selected individuals are given the medication for one week. The following five pairs of observations on each of the three variables were obtained before and after the administration of the medication:

$$\begin{bmatrix} 150, 140 \\ 90, 90 \\ 70, 68 \end{bmatrix}, \begin{bmatrix} 180, 150 \\ 95, 90 \\ 75, 70 \end{bmatrix}, \begin{bmatrix} 160, 160 \\ 85, 80 \\ 70, 65 \end{bmatrix}, \begin{bmatrix} 140, 138 \\ 85, 90 \\ 70, 71 \end{bmatrix}, \begin{bmatrix} 130, 128 \\ 85, 85 \\ 75, 74 \end{bmatrix}.$$

Let X denote the difference, that is, X is equal to the reading before the medication was administered minus the reading after the medication could take effect. The observation vectors on X are then

$$X_1 = \begin{bmatrix} 150 - 140 \\ 90 - 90 \\ 70 - 68 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 2 \end{bmatrix}, X_2 = \begin{bmatrix} 30 \\ 5 \\ 5 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}, X_4 = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

In this case, X_1, \dots, X_5 are observations on iid variables. We are going to assume that these iid variables are coming from a population whose distribution is $N_3(\mu, \Sigma)$, $\Sigma > O$, where Σ is known. Let the sample average $\bar{X} = \frac{1}{5}(X_1 + \dots + X_5)$, the hypothesized mean value vector specified by the null hypothesis $H_o : \mu = \mu_o$, and the known covariance matrix Σ be as follows:

$$\bar{X} = \frac{1}{5} \begin{bmatrix} 44 \\ 5 \\ 12 \end{bmatrix}, \mu_o = \begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Let us evaluate $\bar{X} - \mu_o$ and $n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o)$ which are needed for testing the hypothesis $H_o : \mu = \mu_o$:

$$\begin{aligned} \bar{X} - \mu_o &= \frac{1}{5} \begin{bmatrix} 44 \\ 5 \\ 12 \end{bmatrix} - \begin{bmatrix} 8 \\ 0 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} \\ n(\bar{X} - \mu_o)' \Sigma^{-1} (\bar{X} - \mu_o) &= \frac{5}{5^2} [4 \ 5 \ 2] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} = 8.4. \end{aligned}$$

Let us test H_o at the significance level $\alpha = 0.05$. The critical value which can readily be found in a chisquare table is $\chi_{p, \alpha}^2 = \chi_{3, 0.05}^2 = 7.81$. As per our criterion, we reject H_o if $8.4 \geq \chi_{p, \alpha}^2$; since $8.4 > 7.81$, we reject H_o . The p -value in this case is $Pr\{\chi_p^2 \geq 8.4\} = Pr\{\chi_3^2 \geq 8.4\} \approx 0.04$.

6.2.2. Independent Gaussian populations

Let $Y_j \sim N_p(\mu_{(j)}, \Sigma_j)$, $\Sigma_j > O$, $j = 1, \dots, k$, and let these k populations be independently distributed. Assume that a simple random sample of size n_j from Y_j is available for $j = 1, \dots, k$; then these samples can be represented by the p -vectors Y_{jq} , $q = 1, \dots, n_j$, which are iid as Y_{j1} , for $j = 1, \dots, k$. Consider a given linear

function $X = a_1Y_1 + \cdots + a_kY_k$ where X is $p \times 1$ and the Y_j 's are taken in a given order. Let $U = a_1\bar{Y}_1 + \cdots + a_k\bar{Y}_k$ where $\bar{Y}_j = \frac{1}{n_j} \sum_{q=1}^{n_j} Y_{jq}$ for $j = 1, \dots, k$. Then $E[U] = a_1\mu_{(1)} + \cdots + a_k\mu_{(k)} = \mu$ (say), where a_1, \dots, a_k are given real scalar constants. The covariance matrix in U is $\text{Cov}(U) = \frac{a_1^2}{n_1} \Sigma_1 + \cdots + \frac{a_k^2}{n_k} \Sigma_k = \frac{1}{n} \Sigma$ (say), where n is a symbol. Consider the problem of testing hypotheses on μ when Σ is known or when $a_j, \Sigma_j, j = 1, \dots, k$, are known. Let $H_o : \mu = \mu_o$ (specified), in the sense $\mu_{(j)}$ is a known vector for $j = 1, \dots, k$, when Σ is known. Then, under H_o , all the parameters are known and the standardized U is observable, the test statistic being

$$\sum_{j=1}^k a_j n_j (\bar{Y}_j - \mu_{(j)})' \Sigma_j^{-1} (\bar{Y}_j - \mu_{(j)}) \sim \sum_{j=1}^k a_j \chi_p^{2(j)} \quad (6.2.6)$$

where $\chi_p^{2(j)}, j = 1, \dots, k$, denote independent chisquares random variables, each having p degrees of freedom. However, since this is a linear function of independent chisquare variables, even the null distribution is complicated. Thus, only the case of two independent populations will be examined.

Consider the problem of testing the hypothesis $\mu_1 - \mu_2 = \delta$ (a given vector) when there are two independent normal populations sharing a common covariance matrix Σ (known). Then U is $U = \bar{Y}_1 - \bar{Y}_2$ with $E[U] = \mu_1 - \mu_2 = \delta$ (given) under H_o and $\text{Cov}(U) = (\frac{1}{n_1} + \frac{1}{n_2}) \Sigma = \frac{n_1+n_2}{n_1 n_2} \Sigma$, the test statistic, denoted by v , being

$$v = \frac{n_1 n_2}{n_1 + n_2} (U - \delta)' \Sigma^{-1} (U - \delta) = \frac{n_1 n_2}{n_1 + n_2} (\bar{Y}_1 - \bar{Y}_2 - \delta)' \Sigma^{-1} (\bar{Y}_1 - \bar{Y}_2 - \delta) \sim \chi_p^2. \quad (6.2.7)$$

The resulting test criterion is

$$\text{Reject } H_o \text{ if the observed value of } v \geq \chi_{p, \alpha}^2 \text{ with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} = \alpha. \quad (6.2.8)$$

Example 6.2.3. Let $Y_1 \sim N_3(\mu_{(1)}, \Sigma)$ and $Y_2 \sim N_3(\mu_{(2)}, \Sigma)$ represent independently distributed normal populations having a known common covariance matrix Σ . The null hypothesis is $H_o : \mu_{(1)} - \mu_{(2)} = \delta$ where δ is specified. Denote the observation vectors on Y_1 and Y_2 by $Y_{1j}, j = 1, \dots, n_1$ and $Y_{2j}, j = 1, \dots, n_2$, respectively, and let the sample sizes be $n_1 = 4$ and $n_2 = 5$. Let those observation vectors be

$$Y_{11} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, Y_{12} = \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix}, Y_{13} = \begin{bmatrix} 7 \\ 8 \\ 7 \end{bmatrix}, Y_{14} = \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix} \text{ and}$$

$$Y_{21} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, Y_{22} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}, Y_{23} = \begin{bmatrix} 7 \\ 10 \\ 8 \end{bmatrix}, Y_{24} = \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix}, Y_{25} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix},$$

and the common covariance matrix Σ be

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Let the hypothesized vector under $H_0 : \mu_{(1)} - \mu_{(2)} = \delta$ be $\delta' = (1, 1, 2)$. In order to test this null hypothesis, the following quantities must be evaluated:

$$\begin{aligned} \bar{Y}_1 &= \frac{1}{n_1}(Y_{11} + \cdots + Y_{1n_1}) = \frac{1}{4}(Y_{11} + Y_{12} + Y_{13} + Y_{14}), \\ \bar{Y}_2 &= \frac{1}{n_2}(Y_{21} + \cdots + Y_{2n_2}) = \frac{1}{5}(Y_{21} + \cdots + Y_{25}), \\ U &= \bar{Y}_1 - \bar{Y}_2, \quad v = \frac{n_1 n_2}{n_1 + n_2}(U - \delta)' \Sigma^{-1}(U - \delta). \end{aligned}$$

They are

$$\begin{aligned} \bar{Y}_1 &= \frac{1}{4} \left\{ \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 7 \end{bmatrix} + \begin{bmatrix} 8 \\ 10 \\ 12 \end{bmatrix} \right\} = \frac{1}{4} \begin{bmatrix} 22 \\ 24 \\ 27 \end{bmatrix}, \\ \bar{Y}_2 &= \frac{1}{5} \left\{ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 7 \\ 10 \\ 8 \end{bmatrix} + \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 20 \\ 20 \\ 21 \end{bmatrix}, \\ U &= \bar{Y}_1 - \bar{Y}_2 = \frac{1}{4} \begin{bmatrix} 22 \\ 24 \\ 27 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 20 \\ 20 \\ 21 \end{bmatrix} = \begin{bmatrix} 1.50 \\ 2.00 \\ 2.55 \end{bmatrix}. \end{aligned}$$

Then,

$$\begin{aligned} U - \delta &= \begin{bmatrix} 1.50 \\ 2.00 \\ 2.55 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.50 \\ 1.00 \\ 0.55 \end{bmatrix}. \\ v &= \frac{n_1 n_2}{n_1 + n_2} (U - \delta)' \Sigma^{-1} (U - \delta) \\ &= \frac{(4)(5)}{9} [0.50 \quad 1.00 \quad 0.55] \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.50 \\ 1.00 \\ 0.55 \end{bmatrix} \\ &= 1.3275 \times \frac{20}{9} = 2.95. \end{aligned}$$

Let us test H_o at the significance level $\alpha = 0.05$. The critical value which is available from a chisquare table is $\chi_{p, \alpha}^2 = \chi_{3, 0.05}^2 = 7.81$. As per our criterion, we reject H_o if $2.95 \geq \chi_{p, \alpha}^2$; however, since $2.95 < 7.81$, we cannot reject H_o . The p -value in this case is $Pr\{\chi_p^2 \geq 2.95\} = Pr\{\chi_3^2 \geq 2.95\} \approx 0.096$, which can be determined by interpolation.

6.2a. Testing $H_o : \mu = \mu_o$ (given) When Σ is Known, Complex Gaussian Case

The derivation of the λ -criterion in the complex domain is parallel to that provided for the real case. In the parameter space,

$$\sup_{\Omega} \tilde{L} = \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{S})}}{\pi^{np} |\det(\Sigma)|^n} \quad (6.2a.1)$$

and under $H_o : \mu = \mu_o$, a given vector,

$$\sup_{\omega} \tilde{L} = \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{S}) - n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)}}{\pi^{np} |\det(\Sigma)|^n}. \quad (6.2a.2)$$

Accordingly,

$$\tilde{\lambda} = \frac{\sup_{\omega} \tilde{L}}{\sup_{\Omega} \tilde{L}} = e^{-n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)}. \quad (6.2a.3)$$

Here as well, small values of $\tilde{\lambda}$ correspond to large values of $\tilde{y} \equiv n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o)$, which has a real gamma distribution with the parameters ($\alpha = p$, $\beta = 1$) or a chisquare distribution with p degrees of freedom in the complex domain as described earlier so that $2\tilde{y}$ has a real chisquare distribution having $2p$ degrees of freedom. Thus, a real chisquare table can be utilized for testing the null hypothesis H_o , the criterion being

$$\text{Reject } H_o \text{ if } 2n(\bar{\bar{X}} - \mu_o)^* \Sigma^{-1} (\bar{\bar{X}} - \mu_o) \geq \chi_{2p, \alpha}^2, \text{ with } Pr\{\chi_{2p}^2 \geq \chi_{2p, \alpha}^2\} = \alpha. \quad (6.2a.4)$$

The test criteria as well as the decisions are parallel to those obtained for the real case in the situations of paired values and in the case of independent populations. Accordingly, such test criteria and associated decisions will not be further discussed.

Example 6.2a.1. Let $p = 2$ and the 2×1 complex vector $\tilde{X} \sim \tilde{N}_2(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} = \tilde{\Sigma}^* > O$, with $\tilde{\Sigma}$ assumed to be known. Consider the null hypothesis $H_o : \tilde{\mu} = \tilde{\mu}_o$ where $\tilde{\mu}_o$ is specified. Let the known $\tilde{\Sigma}$ and the specified $\tilde{\mu}_o$ be the following where $i = \sqrt{-1}$:

$$\tilde{\mu}_o = \begin{bmatrix} 1+i \\ 1-2i \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \Rightarrow \tilde{\Sigma}^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix}, \quad \tilde{\Sigma} = \tilde{\Sigma}^* > O.$$

Let the general $\tilde{\mu}$ and general \tilde{X} be represented as follows for $p = 2$:

$$\tilde{\mu} = \begin{bmatrix} \mu_1 + i\nu_1 \\ \mu_2 + i\nu_2 \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} x_1 + iy_1 \\ x_2 + iy_2 \end{bmatrix}$$

so that, for the given $\tilde{\Sigma}$,

$$\det(\tilde{\Sigma}) = (2)(3) - (1+i)(1-i) = 6 - (1^2 + 1^2) = 4 = \det(\tilde{X}^*) = |\det(\tilde{X})|.$$

The exponent of the general density for $p = 2$, excluding -1 , is the form $(\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})$. Further,

$$[(\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})]^* = (\tilde{X} - \tilde{\mu})^* \tilde{\Sigma}^{-1} (\tilde{X} - \tilde{\mu})$$

since both $\tilde{\Sigma}$ and $\tilde{\Sigma}^{-1}$ are Hermitian. Thus, the exponent, which is 1×1 , is real and negative definite. The explicit form, excluding -1 , for $p = 2$ and the given covariance matrix $\tilde{\Sigma}$, is the following:

$$Q = \frac{1}{4} \{ 3[(x_1 - \mu_1)^2 + (y_1 - \nu_1)^2] + 2[(x_2 - \mu_2)^2 + (y_2 - \nu_2)^2] \\ + 2[(x_1 - \mu_1)(x_2 - \mu_2) + (y_1 - \nu_1)(y_2 - \nu_2)] \},$$

and the general density for $p = 2$ and this $\tilde{\Sigma}$ is of the following form:

$$f(\tilde{X}) = \frac{1}{4\pi^2} e^{-Q}$$

where the Q is as previously given. Let the following be an observed sample of size $n = 4$ from a $\tilde{N}_2(\tilde{\mu}_o, \tilde{\Sigma})$ population whose associated covariance matrix $\tilde{\Sigma}$ is as previously specified:

$$\tilde{X}_1 = \begin{bmatrix} 1 \\ 1+i \end{bmatrix}, \quad \tilde{X}_2 = \begin{bmatrix} 2-3i \\ i \end{bmatrix}, \quad \tilde{X}_3 = \begin{bmatrix} 2+i \\ 3 \end{bmatrix}, \quad \tilde{X}_4 = \begin{bmatrix} i \\ 2-i \end{bmatrix}.$$

Then,

$$\bar{\tilde{X}} = \frac{1}{4} \left\{ \begin{bmatrix} 1 \\ 1+i \end{bmatrix} + \begin{bmatrix} 2-3i \\ i \end{bmatrix} + \begin{bmatrix} 2+i \\ 3 \end{bmatrix} + \begin{bmatrix} i \\ 2-i \end{bmatrix} \right\} = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix}$$

$$\bar{\tilde{X}} - \tilde{\mu}_o = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix} - \begin{bmatrix} 1+i \\ 1-2i \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1-5i \\ 2+9i \end{bmatrix},$$

$$\begin{aligned}
& 2n(\bar{X} - \tilde{\mu}_o)' \tilde{\Sigma}^{-1} (\bar{X} - \tilde{\mu}_o) \\
&= (2)(4) \frac{1}{4^2} [1 + 5i \quad 2 - 9i] \frac{1}{4} \begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix} \begin{bmatrix} 1 - 5i \\ 2 + 9i \end{bmatrix} \\
&= \frac{1}{8} \{3(1 + 5i)(1 - 5i) + 2(2 - 9i)(2 + 9i) - (1 + i)(1 + 5i)(2 + 9i) \\
&\quad - (1 - i)(2 - 9i)(1 - 5i)\} \\
&= \frac{1}{8} \{3 \times 26 + 2 \times 85 + 2 \times 62\} = 46.5.
\end{aligned}$$

Let us test the stated null hypothesis at the significance level $\alpha = 0.05$. Since $\chi_{2p, \alpha}^2 = \chi_{4, 0.05}^2 = 9.49$ and $46.5 > 9.49$, we reject H_o . In this case, the p -value is $Pr\{\chi_{2p}^2 \geq 46.5\} = Pr\{\chi_4^2 \geq 46.5\} \approx 0$.

6.2.3. Test involving a subvector of a mean value vector when Σ is known

Let the $p \times 1$ vector $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, and the X_j 's be independently distributed. Let the joint density of X_j , $j = 1, \dots, n$, be denoted by L . Then, as was previously established,

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \quad (i)$$

where $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ and, letting $\mathbf{X} = (X_1, \dots, X_n)$ of dimension $p \times n$ and $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$, $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$. Let \bar{X} , Σ^{-1} and μ be partitioned as follows:

$$\bar{X} = \begin{bmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \end{bmatrix}, \quad \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}$$

where $\bar{X}^{(1)}$ and $\mu^{(1)}$ are $r \times 1$, $r < p$, and Σ^{11} is $r \times r$. Consider the hypothesis $\mu^{(1)} = \mu_o^{(1)}$ (specified) with Σ known. Thus, this hypothesis concerns only a subvector of the mean value vector, the population covariance matrix being assumed known. In the entire parameter space Ω , μ is estimated by \bar{X} where \bar{X} is the maximum likelihood estimator (MLE) of μ . The maximum of the likelihood function in the entire parameter space is then

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (ii)$$

Let us now determine the MLE of $\mu^{(2)}$, which is the only unknown quantity under the null hypothesis. To this end, we consider the following expansion:

$$\begin{aligned} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) &= [(\bar{X}^{(1)} - \mu_o^{(1)})', (\bar{X}^{(2)} - \mu^{(2)})'] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \bar{X}^{(1)} - \mu_o^{(1)} \\ \bar{X}^{(2)} - \mu^{(2)} \end{bmatrix} \\ &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{11} (\bar{X}^{(1)} - \mu_o^{(1)}) + (\bar{X}^{(2)} - \mu^{(2)})' \Sigma^{22} (\bar{X}^{(2)} - \mu^{(2)}) \\ &\quad + 2(\bar{X}^{(2)} - \mu^{(2)})' \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}). \end{aligned} \quad (iii)$$

Noting that there are only two terms involving $\mu^{(2)}$ in (iii), we have

$$\begin{aligned} \frac{\partial}{\partial \mu^{(2)}} \ln L = O &\Rightarrow O - 2\Sigma^{22}(\bar{X}^{(2)} - \mu^{(2)}) - 2\Sigma^{21}(\bar{X}^{(1)} - \mu_o^{(1)}) = O \\ &\Rightarrow \hat{\mu}^{(2)} = \bar{X}^{(2)} + (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}). \end{aligned}$$

Then, substituting this MLE $\hat{\mu}^{(2)}$ in the various terms in (iii), we have the following:

$$\begin{aligned} (\bar{X}^{(2)} - \hat{\mu}^{(2)})' \Sigma^{22} (\bar{X}^{(2)} - \hat{\mu}^{(2)}) &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) \\ 2(\bar{X}^{(2)} - \hat{\mu}^{(2)})' \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) &= -2(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21} (\bar{X}^{(1)} - \mu_o^{(1)}) \Rightarrow \\ (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) &= (\bar{X}^{(1)} - \mu_o^{(1)})' [\Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}] (\bar{X}^{(1)} - \mu_o^{(1)}) \\ &= (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}), \end{aligned}$$

since, as established in Sect. 1.3, $\Sigma_{11}^{-1} = \Sigma^{11} - \Sigma^{12} (\Sigma^{22})^{-1} \Sigma^{21}$. Thus, the maximum of L under the null hypothesis is given by

$$\max_{H_o} L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)})}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}},$$

and the λ -criterion is then

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = e^{-\frac{n}{2} (\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)})}. \quad (6.2.1)$$

Hence, we reject H_o for small values of λ or for large values of $n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) \sim \chi_r^2$ since the expected value and covariance matrix of $\bar{X}^{(1)}$ are respectively $\mu_o^{(1)}$ and Σ_{11}/n . Accordingly, the criterion can be enunciated as follows:

$$\text{Reject } H_o : \mu^{(1)} = \mu_o^{(1)} \text{ (given) if } u \equiv n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) \geq \chi_{r, \alpha}^2 \quad (6.2.2)$$

with $Pr\{\chi_r^2 \geq \chi_{r, \alpha}^2\} = \alpha$. In the complex Gaussian case, the corresponding $2\tilde{u}$ will be distributed as a real chisquare random variable having $2r$ degrees of freedom; thus, the

criterion will consist of rejecting the corresponding null hypothesis whenever the observed value of $2\tilde{u} \geq \chi_{2r, \alpha}^2$.

Example 6.2.4. Let the 4×1 vector X have a real normal distribution $N_4(\mu, \Sigma)$, $\Sigma > O$. Consider the hypothesis that part of μ is specified. For example, let the hypothesis H_o and Σ be the following:

$$H_o : \mu = \mu_o = \begin{bmatrix} 1 \\ -1 \\ \mu_3 \\ \mu_4 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} = \Sigma' > O, X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \equiv \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix}.$$

Since we are specifying the first two parameters in μ , the hypothesis can be tested by computing the distribution of $X^{(1)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Observe that $X^{(1)} \sim N_2(\mu^{(1)}, \Sigma_{11})$, $\Sigma_{11} > O$ where

$$\mu^{(1)} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \mu_o^{(1)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, H_o : \mu^{(1)} = \mu_o^{(1)}, \Sigma_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \Sigma_{11}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Let the observed vectors from the original $N_4(\mu, \Sigma)$ population be

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 4 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ 2 \end{bmatrix}, X_3 = \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \end{bmatrix}, X_4 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 4 \end{bmatrix}.$$

Then the observations corresponding to the subvector $X^{(1)}$, denoted by $X_j^{(1)}$, are the following:

$$X_1^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2^{(1)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3^{(1)} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, X_4^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, X_5^{(1)} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

In this case, the sample size $n = 5$ and the sample mean, denoted by $\bar{X}^{(1)}$, is

$$\begin{aligned} \bar{X}^{(1)} &= \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \Rightarrow \\ \bar{X}^{(1)} - \mu_o^{(1)} &= \frac{1}{5} \begin{bmatrix} 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -1 \\ 8 \end{bmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} n(\bar{X}^{(1)} - \mu_o^{(1)})' \Sigma_{11}^{-1} (\bar{X}^{(1)} - \mu_o^{(1)}) &= \frac{5}{5^2} [-1 \quad 8] \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 8 \end{bmatrix} \\ &= \frac{1}{15} (146) = 9.73. \end{aligned}$$

If $9.73 > \chi_{2,\alpha}^2$, then we would reject $H_o^{(1)} : \mu^{(1)} = \mu_o^{(1)}$. Let us test this hypothesis at the significance level $\alpha = 0.01$. Since $\chi_{2,0.01}^2 = 9.21$, we reject the null hypothesis. In this instance, the p -value, which can be determined from a chisquare table, is $Pr\{\chi_2^2 \geq 9.73\} \approx 0.007$.

6.2.4. Testing $\mu_1 = \dots = \mu_p$, with Σ known, real Gaussian case

Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, and the X_j be independently distributed. Letting $\mu' = (\mu_1, \dots, \mu_p)$, consider the hypothesis

$$H_o : \mu_1 = \mu_2 = \dots = \mu_p = \nu,$$

where ν , the common μ_j is unknown. This implies that $\mu_i - \mu_j = 0$ for all i and j . Consider the $p \times 1$ vector J of unities, $J' = (1, \dots, 1)$ and then take any non-null vector that is orthogonal to J . Let A be such a vector so that $A'J = 0$. Actually, $p - 1$ linearly independent such vectors are available. For example, if p is even, then take $1, -1, \dots, 1, -1$ as the elements of A and, when p is odd, one can start with $1, -1, \dots, 1, -1$ and take the last three elements as $1, -2, 1$, or the last element as 0 , that is,

$$J = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad \text{for } p \text{ even} \quad \text{and} \quad A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -2 \\ 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 0 \end{bmatrix} \quad \text{for } p \text{ odd}.$$

When the last element of the vector A is zero, we are simply ignoring the last element in X_j . Let the $p \times 1$ vector $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, and the X_j 's be independently distributed. Let the scalar $y_j = A'X_j$ and the $1 \times n$ vector $Y = (y_1, \dots, y_n) = (A'X_1, \dots, A'X_n) = A'(X_1, \dots, X_n) = A'\mathbf{X}$, where the $p \times n$ matrix $\mathbf{X} = (X_1, \dots, X_n)$. Let $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n) = A'\frac{1}{n}(X_1 + \dots + X_n) = A'\bar{X}$. Then, $\sum_{j=1}^n (y_j - \bar{y})(y_j - \bar{y})' = A' \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' A$ where $\sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = S =$

the sample sum of products matrix in the X_j 's, where $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ the $p \times n$ matrix whose columns are all equal to \bar{X} . Thus, one has $\sum_{j=1}^n (y_j - \bar{y})^2 = A'SA$. Consider the hypothesis $\mu_1 = \dots = \mu_p = \nu$. Then, $A'\mu = \nu A'J = \nu \mathbf{0} = \mathbf{0}$ under H_o . Since $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, we have $y_j \sim N_1(A'\mu, A'\Sigma A)$, $A'\Sigma A > 0$. Under H_o , $y_j \sim N_1(0, A'\Sigma A)$, $j = 1, \dots, n$, the y_j 's being independently distributed. Consider the joint density of y_1, \dots, y_n , denoted by L :

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2A'\Sigma A}(y_j - A'\mu)^2}}{(2\pi)^{\frac{1}{2}}[A'\Sigma A]^{\frac{1}{2}}}. \quad (i)$$

Since Σ is known, the only unknown quantity in L is μ . Differentiating $\ln L$ with respect to μ and equating the result to a null vector, we have

$$\sum_{j=1}^n (y_j - A'\hat{\mu}) = 0 \Rightarrow \sum_{j=1}^n y_j - nA'\hat{\mu} = 0 \Rightarrow \bar{y} - A'\hat{\mu} = 0 \Rightarrow A'(\bar{X} - \hat{\mu}) = 0.$$

However, since A is a fixed known vector and the equation holds for arbitrary \bar{X} , $\hat{\mu} = \bar{X}$. Hence the maximum of L , in the entire parameter space $\Omega = \mu$, is the following:

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2A'\Sigma A} \sum_{j=1}^n [A'(X_j - \bar{X})]^2}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}} = \frac{e^{-\frac{1}{2A'\Sigma A} A'SA}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}}. \quad (ii)$$

Now, noting that under H_o , $A'\mu = 0$, we have

$$\max_{H_o} L = \frac{e^{-\frac{1}{2A'\Sigma A} \sum_{j=1}^n A'X_jX_j'A}}{(2\pi)^{\frac{np}{2}} [A'\Sigma A]^{\frac{n}{2}}}. \quad (iii)$$

From (i) to (iii), the λ -criterion is as follows, observing that $A'(\sum_{j=1}^n X_jX_j')A = \sum_{j=1}^n A'(X_j - \bar{X})(X_j - \bar{X})'A + nA'(\bar{X}\bar{X}'A) = A'SA + nA'\bar{X}\bar{X}'A$:

$$\lambda = e^{-\frac{n}{2A'\Sigma A} A'\bar{X}\bar{X}'A}. \quad (6.2.3)$$

But since $\sqrt{\frac{n}{A'\Sigma A}} A'\bar{X} \sim N_1(0, 1)$ under H_o , we may test this null hypothesis either by using the standard normal variable or a chisquare variable as $\frac{n}{A'\Sigma A} A'\bar{X}\bar{X}'A \sim \chi_1^2$ under H_o . Accordingly, the criterion consists of rejecting H_o

$$\text{when } \left| \sqrt{\frac{n}{A'\Sigma A}} A'\bar{X} \right| \geq z_{\frac{\alpha}{2}}, \text{ with } Pr\{z \geq z_{\beta}\} = \beta, z \sim N_1(0, 1)$$

or

$$\text{when } u \equiv \frac{n}{A'\Sigma A} (A'\bar{X}\bar{X}'A) \geq \chi_{1, \alpha}^2, \text{ with } Pr\{\chi_1^2 \geq \chi_{1, \alpha}^2\} = \alpha, u \sim \chi_1^2. \quad (6.2.4)$$

Example 6.2.5. Consider a 4-variate real Gaussian vector $X \sim N_4(\mu, \Sigma)$, $\Sigma > O$ with Σ as specified in Example 6.2.4 and the null hypothesis that the individual components of the mean value vector μ are all equal, that is,

$$\Sigma = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}, \quad H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4 \equiv v \text{ (say)}, \quad \text{with } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}.$$

Let L be a 4×1 constant vector such that $L' = (1, -1, 1, -1)$. Then, under H_o , $L'\mu = 0$ and $u = L'X$ is univariate normal; more specifically, $u \sim N_1(0, L'\Sigma L)$ where

$$L'\Sigma L = [1 \quad -1 \quad 1 \quad -1] \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = 7 \Rightarrow u \sim N_1(0, 7).$$

Let the observation vectors be the same as those used in Example 6.2.4 and let $u_j = L'X_j$, $j = 1, \dots, 5$. Then, the five independent observations from $u \sim N_1(0, 7)$ are the following:

$$u_1 = L'X_1 = [1 \quad -1 \quad 1 \quad -1] \begin{bmatrix} 1 \\ 0 \\ 2 \\ 4 \end{bmatrix} = -1, \quad u_2 = L' \begin{bmatrix} -1 \\ 1 \\ 1 \\ 2 \end{bmatrix} = -3, \quad u_3 = L' \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \end{bmatrix} = -3,$$

$$u_4 = L' \begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix} = -3, \quad u_5 = L' \begin{bmatrix} 2 \\ -1 \\ 0 \\ 4 \end{bmatrix} = -1,$$

the average $\bar{u} = \frac{1}{5}(u_1 + \dots + u_5) = \frac{1}{5}(-1 - 3 - 3 - 3 - 1)$ being equal to $-\frac{11}{5}$. Then, the standardized sample mean $z = \frac{\sqrt{n}}{\sigma_u}(\bar{u} - 0) \sim N_1(0, 1)$. Let us test the null hypothesis at the significance level $\alpha = 0.05$. Referring to a $N_1(0, 1)$ table, the required critical value, denoted by $z_{\frac{\alpha}{2}} = z_{0.025}$ is 1.96. Therefore, we reject H_o in favor of the alternative hypothesis that at least two components of μ are unequal at significance level α if the observed value of

$$|z| = \left| \frac{\sqrt{n}}{\sigma_u}(\bar{u} - 0) \right| \geq 1.96.$$

Since the observed value of $|z|$ is $|\frac{\sqrt{5}}{\sqrt{7}}(-\frac{7}{5} - 0)| = \sqrt{1.4} = 1.18$ is less than 1.96, we do not reject H_o at the 5% significance level. Letting $z \sim N_1(0, 1)$, the p -value in this case is $Pr\{|z| \geq 1.18\} = 0.238$, this quantile being available from a standard normal table.

In the complex case, proceeding in a parallel manner to the real case, the lambda criterion will be the following:

$$\tilde{\lambda} = e^{-\frac{n}{A^* \tilde{\Sigma} A} A^* \tilde{X} \tilde{X}^* A} \quad (6.2a.5)$$

where an asterisk indicates the conjugate transpose. Letting $\tilde{u} = \frac{2n}{A^* \tilde{\Sigma} A} (A^* \tilde{X} \tilde{X}^* A)$, it can be shown that under H_o , \tilde{u} is distributed as a real chisquare random variable having 2 degrees of freedom. Accordingly, the criterion will be as follows:

$$\text{Reject } H_o \text{ if the observed } \tilde{u} \geq \chi_{2, \alpha}^2 \text{ with } Pr\{\chi_2^2 \geq \chi_{2, \alpha}^2\} = \alpha. \quad (6.2a.6)$$

Example 6.2a.2. When $p > 2$, the computations become quite involved in the complex case. Thus, we will let $p = 2$ and consider the bivariate complex $\tilde{N}_2(\tilde{\mu}, \tilde{\Sigma})$ distribution that was specified in Example 6.2a.1, assuming that $\tilde{\Sigma}$ is as given therein, the same set of observations being utilized as well. In this case, the null hypothesis is $H_o : \tilde{\mu}_1 = \tilde{\mu}_2$, the parameters and sample average being

$$\tilde{\mu} = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}, \quad \tilde{X} = \frac{1}{4} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix}.$$

Letting $L' = (1, -1)$, $L' \tilde{\mu} = 0$ under H_o , and

$$\tilde{u} = L' \tilde{X} = \frac{1}{4} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 5-i \\ 6+i \end{bmatrix} = -\frac{1}{4}(1+2i); \quad (L' \tilde{X})^* (L' \tilde{X}) = \frac{1}{16}(1-2i)(1+2i) = \frac{5}{16};$$

$$L' \tilde{\Sigma} L = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 3; \quad v = \frac{2n}{L' \tilde{\Sigma} L} [(L' \tilde{X})^* (L' \tilde{X})] = \frac{8}{3} \times \frac{5}{16} = \frac{5}{6}.$$

The criterion consists of rejecting H_o if the observed value of $v \geq \chi_{2, \alpha}^2$. Letting the significance level of the test be $\alpha = 0.05$, the critical value is $\chi_{2, 0.05}^2 = 5.99$, which is readily available from a chisquare table. The observed value of v being $\frac{5}{6} < 5.99$, we do not reject H_o . In this case, the p -value is $Pr\{\chi_2^2 \geq \frac{5}{6}\} \approx 0.318$.

6.2.5. Likelihood ratio criterion for testing $H_o : \mu_1 = \dots = \mu_p, \Sigma$ known

Consider again, $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, with the X_j 's being independently distributed and Σ , assumed known. Letting the joint density of X_1, \dots, X_n be denoted by L , then, as determined earlier,

$$L = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \quad (i)$$

where n is the sample size and S is the sample sum of products matrix. In the entire parameter space

$$\Omega = \{(\mu, \Sigma) \mid \Sigma > O \text{ known, } \mu' = (\mu_1, \dots, \mu_p)\},$$

the MLE of μ is \bar{X} = the sample average. Then

$$\max_{\Omega} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}. \quad (ii)$$

Consider the following hypothesis on $\mu' = (\mu_1, \dots, \mu_p)$:

$$H_o : \mu_1 = \dots = \mu_p = \nu, \nu \text{ is unknown.}$$

Then, the MLE of μ under H_o is $\hat{\mu} = J\hat{\nu} = J\frac{1}{p}J'\bar{X}$, $J' = (1, \dots, 1)$. This $\hat{\nu}$ is in fact the sum of all observations on all components of \bar{X}_j , $j = 1, \dots, n$, divided by np , which is identical to the sum of all the coordinates of \bar{X} divided by p or $\hat{\mu} = \frac{1}{p}JJ'\bar{X}$. In order to evaluate the maximum of L under H_o , it suffices to substitute $\hat{\mu}$ to μ in (i). Accordingly, the λ -criterion is

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = e^{-\frac{n}{2}(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu})}. \quad (6.2.5)$$

Thus, we reject H_o for small values of λ or for large values of $w \equiv n(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu})$. Let us determine the distribution of w . First, note that

$$\bar{X} - \hat{\mu} = \bar{X} - \frac{1}{p}JJ'\bar{X} = (I_p - \frac{1}{p}JJ')\bar{X},$$

and let

$$\begin{aligned} w &= n(\bar{X} - \hat{\mu})' \Sigma^{-1}(\bar{X} - \hat{\mu}) = n\bar{X}'(I - \frac{1}{p}JJ')\Sigma^{-1}(I - \frac{1}{p}JJ')\bar{X} \quad (iii) \\ &= (\bar{X} - \mu)'(I - \frac{1}{p}JJ')\Sigma^{-1}(I - \frac{1}{p}JJ')(\bar{X} - \mu) \end{aligned}$$

since $J'(I - \frac{1}{p}JJ') = O$, $\mu = \nu J$ being the true mean value of the $N_p(\mu, \Sigma)$ distribution. Observe that $\sqrt{n}(\bar{X} - \mu) \sim N_p(O, \Sigma)$, $\Sigma > O$, and that $\frac{1}{p}JJ'$ is idempotent. Since $I - \frac{1}{p}JJ'$ is also idempotent and its rank is $p - 1$, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that

$$I - \frac{1}{p}JJ' = P' \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} P.$$

Letting $U = P\sqrt{n}(\bar{X} - \hat{\mu})$, with $U' = (u_1, \dots, u_{p-1}, u_p)$, $U \sim N_p(O, P\Sigma P')$. Now, on noting that

$$U' \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} = (u_1, \dots, u_{p-1}, 0),$$

we have

$$n(\bar{X} - \hat{\mu})' \Sigma^{-1} (\bar{X} - \hat{\mu}) = [U_1', 0] P \Sigma^{-1} P' \begin{bmatrix} U_1 \\ 0 \end{bmatrix} = U_1' B^{-1} U_1, \quad U_1' = (u_1, \dots, u_{p-1}),$$

B being the covariance matrix associated with U_1 , so that $U_1 \sim N_{p-1}(O, B)$, $B > O$. Thus, $U_1' B^{-1} U_1 \sim \chi_{p-1}^2$, a real scalar chisquare random variable having $p - 1$ degrees of freedom. Hence, upon evaluating

$$w = \bar{X}' \left(I - \frac{1}{p} J J' \right) \Sigma^{-1} \left(I - \frac{1}{p} J J' \right) \bar{X},$$

one would reject $H_o : \mu_1 = \dots = \mu_p = \nu$, ν unknown, whenever the observed value of

$$w \geq \chi_{p-1, \alpha}^2, \quad \text{with } Pr\{\chi_{p-1}^2 \geq \chi_{p-1, \alpha}^2\} = \alpha. \quad (6.2.6)$$

Observe that the degrees of freedom of this chisquare variable, that is, $p - 1$, coincides with the number of parameters being restricted by H_o .

Example 6.2.6. Consider the trivariate real Gaussian population $X \sim N_3(\mu, \Sigma)$, $\Sigma > O$, as already specified in Example 6.2.1 with the same Σ and the same observed sample vectors for testing $H_o : \mu' = (\nu, \nu, \nu)$, namely,

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \bar{X} = \frac{1}{5} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

The following test statistic has to be evaluated for $p = 3$:

$$w = \bar{X}' \left(I - \frac{1}{p} J J' \right) \Sigma^{-1} \left(I - \frac{1}{p} J J' \right) \bar{X}, \quad J' = (1, 1, 1).$$

We have to evaluate the following quantities in order to determine the value of w :

$$\begin{aligned} \frac{1}{3}JJ'\Sigma^{-1} &= \frac{1}{3} \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix}, \quad \Sigma^{-1}\frac{1}{3}JJ' = \frac{1}{3} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \\ \frac{1}{3}JJ'\Sigma^{-1}\frac{1}{3}JJ' &= \frac{1}{9} \begin{bmatrix} \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & \frac{3}{2} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \\ \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 1 & 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ + \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{7}{6} \\ 0 & -\frac{7}{6} & \frac{7}{6} \end{bmatrix} = (I - \frac{1}{3}JJ')\Sigma^{-1}(I - \frac{1}{3}JJ'). \end{aligned}$$

Thus,

$$\begin{aligned} w &= \bar{X}'(I - \frac{1}{3}JJ')\Sigma^{-1}(I - \frac{1}{3}JJ')\bar{X}, \quad J' = (1, 1, 1) \\ &= \frac{1}{5^2} [9 \quad 4 \quad 3] \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{7}{6} \\ 0 & -\frac{7}{6} & \frac{7}{6} \end{bmatrix} \begin{bmatrix} 9 \\ 4 \\ 3 \end{bmatrix} \\ &= \frac{1}{5^2} \left[9^2 \frac{1}{3} + 4^2 \frac{3}{2} + 3^2 \frac{7}{6} - \frac{2}{3} (9)(4) - \frac{14}{6} (4)(3) \right] \\ &= 0.38. \end{aligned}$$

We reject H_0 whenever $w \geq \chi_{p-1, \alpha}^2$. Letting the significance level be $\alpha = 0.05$, the tabulated critical value is $\chi_{p-1, \alpha}^2 = \chi_{2, 0.05}^2 = 5.99$, and since $0.38 < 5.99$, we do not reject the null hypothesis. In this instance, the p -value is $Pr\{\chi_2^2 \geq 0.38\} \approx 0.32$.

6.3. Testing $H_0 : \mu = \mu_0$ (given) When Σ is Unknown, Real Gaussian Case

In this case, both μ and Σ are unknown in the entire parameter space Ω ; however, $\mu = \mu_0$ known while Σ is still unknown in the subspace ω . The MLE under Ω is the same as that obtained in Sect. 6.1.1., that is,

$$\sup_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}. \quad (6.3.1)$$

When $\mu = \mu_o$, Σ is estimated by $\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (X_j - \mu_o)(X_j - \mu_o)'$. As shown in Sect. 3.5, $\hat{\Sigma}$ can be reexpressed as follows:

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{n} \sum_{j=1}^n (X_j - \mu_o)(X_j - \mu_o) = \frac{1}{n} S + (\bar{X} - \mu_o)(\bar{X} - \mu_o)' \\ &= \frac{1}{n} [S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'].\end{aligned}$$

Then, under the null hypothesis, we have

$$\sup_{\omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'|^{\frac{n}{2}}}. \quad (6.3.2)$$

Thus,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{|S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'|^{\frac{n}{2}}}.$$

On applying results on the determinants of partitioned matrices which were obtained in Sect. 1.3, we have the following equivalent representations of the denominator:

$$\begin{aligned}\left| \begin{bmatrix} S & n(\bar{X} - \mu_o) \\ -(\bar{X} - \mu_o)' & 1 \end{bmatrix} \right| &= [1] |S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'| \\ &= |S| |1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)|,\end{aligned}$$

that is,

$$|S + n(\bar{X} - \mu_o)(\bar{X} - \mu_o)'| = |S| [1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)],$$

which yields the following simplified representation of the likelihood ratio statistic:

$$\lambda = \frac{1}{[1 + n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)]^{\frac{n}{2}}}. \quad (6.3.3)$$

Small values of λ correspond to large values of $u \equiv n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$, which is connected to Hotelling's T_n^2 statistic. Hence the criterion is the following: "Reject H_o for large values of u ". The distribution of u can be derived by making use of the independence of the sample mean and sample sum of products matrix and the densities of these quantities. An outline of the derivation is provided in the next subsection.

6.3.1. The distribution of the test statistic

Let us examine the distribution of $u = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$. We have already established in Theorems 3.5.3, that S and \bar{X} are independently distributed in the case of a real p -variate nonsingular Gaussian $N_p(\mu, \Sigma)$ population. It was also determined in Corollary 3.5.1 that the distribution of the sample average \bar{X} is a p -variate real Gaussian vector with the parameters μ and $\frac{1}{n}\Sigma$, $\Sigma > O$ and in the continuing discussion, it is shown that the distribution of S is a matrix-variate Wishart with $m = n - 1$ degrees of freedom, where n is the sample size and parameter matrix $\Sigma > O$. Hence the joint density of S and \bar{X} , denoted by $f(S, \bar{X})$, is the product of the marginal densities. Letting $\Sigma = I$, this joint density is given by

$$f(S, \bar{X}) = \frac{n^{\frac{p}{2}}}{(2\pi)^{\frac{p}{2}} 2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S) - \frac{n}{2}\text{tr}((\bar{X} - \mu)(\bar{X} - \mu)')}, \quad m = n - 1. \quad (i)$$

Note that it is sufficient to consider the case $\Sigma = I$. Due to the presence of S^{-1} in $u = (\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$, the effect of any scaling matrix on X_j will disappear. If X_j goes to $A^{\frac{1}{2}}X_j$ for any constant positive definite matrix A then S^{-1} will go to $A^{-\frac{1}{2}}S^{-1}A^{-\frac{1}{2}}$ and thus u will be free of A .

Letting $Y = S^{-\frac{1}{2}}(\bar{X} - \mu)$ for fixed S , $Y \sim N_p(O, S^{-1}/n)$, so that the conditional density of Y , given S , is

$$g(Y|S) = \frac{n^{\frac{p}{2}} |S|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} e^{-\frac{n}{2}\text{tr}(SY Y')}.$$

Thus, the joint density of S and Y , denoted by $f_1(S, Y)$, is

$$f_1(S, Y) = \frac{n^{\frac{p}{2}}}{(2\pi)^{\frac{p}{2}} 2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} |S|^{\frac{m+1}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S[I + nYY'])}, \quad m = n - 1. \quad (ii)$$

On integrating out S from (ii) by making use of a matrix-variate gamma integral, we obtain the following marginal density of Y , denoted by $f_2(Y)$:

$$f_2(Y)dY = \frac{n^{\frac{p}{2}}}{(\pi)^{\frac{p}{2}}} \frac{\Gamma_p(\frac{m+1}{2})}{\Gamma_p(\frac{m}{2})} |I + nYY'|^{-\frac{(m+1)}{2}} dY, \quad m = n - 1. \quad (iii)$$

However, $|I + nYY'| = 1 + nY'Y$, which can be established by considering two representations of the determinant

$$\begin{vmatrix} I & -\sqrt{n}Y \\ \sqrt{n}Y' & 1 \end{vmatrix},$$

similarly to what was done in Sect. 6.3 to obtain the likelihood ratio statistic given in (6.3.3). As well, it can easily be shown that

$$\frac{\Gamma_p(\frac{m+1}{2})}{\Gamma_p(\frac{m}{2})} = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m+1}{2} - \frac{p}{2})}$$

by expanding the matrix-variate gamma functions. Now, letting $s = Y'Y$, it follows from Theorem 4.2.3 that $dY = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} ds$. Thus, the density of s , denoted by $f_3(s)$, is

$$f_3(s) ds = \frac{n^{\frac{p}{2}} \Gamma(\frac{m+1}{2})}{\Gamma(\frac{m+1}{2} - \frac{p}{2}) \Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} (1 + ns)^{-(\frac{m+1}{2})} ds \quad (6.3.4)$$

$$= \frac{n^{\frac{p}{2}} \Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2} - \frac{p}{2}) \Gamma(\frac{p}{2})} s^{\frac{p}{2}-1} (1 + ns)^{-(\frac{n}{2})} ds, \quad m = n - 1, \quad (6.3.5)$$

for $n = p + 1, p + 2, \dots, 0 \leq s < \infty$, and zero elsewhere. It can then readily be seen from (6.3.5) that $ns = nY'Y = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) = u$ is distributed as a real scalar type-2 beta random variable whose parameters are $(\frac{p}{2}, \frac{n}{2} - \frac{p}{2})$, $n = p + 1, \dots$. Thus, the following result:

Theorem 6.3.1. *Consider a real p -variate normal population $N_p(\mu, \Sigma)$, $\Sigma > O$, and a simple random sample of size n from this normal population, $X_j \sim N_p(\mu, \Sigma)$, $j = 1, \dots, n$, the X_j 's being independently distributed. Let the $p \times n$ matrix $\mathbf{X} = (X_1, \dots, X_n)$ be the sample matrix and the p -vector $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ denote the sample average. Let $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$ be a $p \times n$ matrix whose columns are all equal to \bar{X} , and $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ be the sample sum of products matrix. Then, $u = n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu)$ has a real scalar type-2 beta distribution with the parameters $(\frac{p}{2}, \frac{n}{2} - \frac{p}{2})$, so that $u \sim \frac{p}{n-p} F_{p, n-p}$ where $F_{p, n-p}$ denotes a real F random variable whose degrees of freedoms are p and $n - p$.*

Hence, in order to test the hypothesis $H_o : \mu = \mu_o$, the likelihood ratio statistic gives the test criterion: Reject H_o for large values of $u = n(\bar{X} - \mu_o)'S^{-1}(\bar{X} - \mu_o)$, which is equivalent to rejecting H_o for large values of an F -random variable having p and $n - p$ degrees of freedom where $F_{p, n-p} = \frac{n-p}{p} u = \frac{n-p}{p} n(\bar{X} - \mu_o)'S^{-1}(\bar{X} - \mu_o)$, that is,

$$\begin{aligned} \text{reject } H_o \text{ if } \frac{n-p}{p} u = F_{p, n-p} \geq F_{p, n-p, \alpha}, \\ \text{with } \alpha = Pr\{F_{p, n-p} \geq F_{n, n-p, \alpha}\} \end{aligned} \quad (6.3.6)$$

at a given significance level α where $u = n(\bar{X} - \mu_o)' S^{-1}(\bar{X} - \mu_o) \sim \frac{p}{n-p} F_{p, n-p}$, n being the sample size.

Example 6.3.1. Consider a trivariate real Gaussian vector $X \sim N_3(\mu, \Sigma)$, $\Sigma > O$, where Σ is unknown. We would like to test the following hypothesis on μ : $H_o : \mu = \mu_o$, with $\mu'_o = (1, 1, 1)$. Consider the following simple random sample of size $n = 5$ from this $N_3(\mu, \Sigma)$ population:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

so that

$$\begin{aligned} \bar{X} &= \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, X_1 - \bar{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix}, X_2 - \bar{X} = \frac{1}{5} \begin{bmatrix} 4 \\ -2 \\ -9 \end{bmatrix}, \\ X_3 - \bar{X} &= \frac{1}{5} \begin{bmatrix} -6 \\ 3 \\ 6 \end{bmatrix}, X_4 - \bar{X} = \frac{1}{5} \begin{bmatrix} -11 \\ 3 \\ 6 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} 9 \\ -7 \\ -4 \end{bmatrix}. \end{aligned}$$

Let $\mathbf{X} = [X_1, \dots, X_5]$ the 3×5 sample matrix and $\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \dots, \bar{X}]$ be the 3×5 matrix of sample means. Then,

$$\begin{aligned} \mathbf{X} - \bar{\mathbf{X}} &= [X_1 - \bar{X}, \dots, X_5 - \bar{X}] = \frac{1}{5} \begin{bmatrix} 4 & 4 & -6 & -11 & 9 \\ 3 & -2 & 3 & 3 & -7 \\ 1 & -9 & 6 & 6 & -4 \end{bmatrix}, \\ S &= (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \frac{1}{5^2} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix}. \end{aligned}$$

Let $S = \frac{1}{5^2} A$. In order to evaluate the test statistic, we need $S^{-1} = 25A^{-1}$. To obtain the correct inverse without any approximation, we will use the transpose of the cofactor matrix divided by the determinant. The determinant of A , $|A|$, as obtained in terms of the elements of the first row and the corresponding cofactors is equal to 531250. The matrix of cofactors, denoted by $\text{Cof}(A)$, which is symmetric in this case, is the following:

$$\text{Cof}(A) = \begin{bmatrix} 6375 & 4250 & 4250 \\ 4250 & 17000 & -4250 \\ 4250 & -4250 & 9500 \end{bmatrix} \Rightarrow S^{-1} = \frac{25}{531250} \text{Cof}(A).$$

The null hypothesis is $H_o : \mu = \mu_o = (1, 1, 1)'$, so that

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix},$$

the observed value of the test statistic being

$$\begin{aligned} w &= \frac{n-p}{p} n (\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o) = \frac{(5-3)}{3} 5 \frac{25}{5^2} [4 \ 3 \ 1] A^{-1} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix} \\ &= \frac{2}{3} 5 \frac{25}{5^2} \frac{1}{531250} [(4)^2(6375) + (3)^2(17000) + (1)^2(9500) \\ &\quad + 2(4)(3)(4250) + 2(4)(1)(4250) - 2(3)(1)(4250)] \\ &= \frac{2}{3} 5 \frac{25}{5^2} \frac{1}{531250} [375000] = 2.35. \end{aligned}$$

The test statistic w under the null hypothesis is F -distributed, that is, $w \sim F_{p, n-p}$. Let us test H_o at the significance level $\alpha = 0.05$. Since the critical value as obtained from an F -table is $F_{p, n-p, \alpha} = F_{3, 2, 0.05} = 19.2$ and $2.35 < 19.2$, we do not reject H_o .

Note 6.3.1. If S is replaced by $\frac{1}{n-1}S$, an unbiased estimator for Σ , then the test statistic $\frac{1}{n-1}n(\bar{X} - \mu_o)'[\frac{1}{n-1}S]^{-1}(\bar{X} - \mu_o) = \frac{T_n^2}{n-1}$ where T_n^2 denotes Hotelling's T^2 statistic, which for $p = 1$ corresponds to the square of a Student- t statistic having $n-1$ degrees of freedom.

Since u as defined in Theorem 6.3.1 is distributed as a type-2 beta random variable with the parameters $(\frac{p}{2}, \frac{n-p}{2})$, we have the following results: $\frac{1}{u}$ is type-2 beta distributed with the parameters $(\frac{n-p}{2}, \frac{p}{2})$, $\frac{u}{1+u}$ is type-1 beta distributed with the parameters $(\frac{p}{2}, \frac{n-p}{2})$, and $\frac{1}{1+u}$ is type-1 beta distributed with the parameters $(\frac{n-p}{2}, \frac{p}{2})$, n being the sample size.

6.3.2. Paired values or linear functions when Σ is unknown

Let Y_1, \dots, Y_k be $p \times 1$ vectors having their own distributions which are unknown. However, suppose that it is known that a certain linear function $X = a_1 Y_1 + \dots + a_k Y_k$ has a p -variate real Gaussian $N_p(\mu, \Sigma)$ distribution with $\Sigma > O$. We would like to test hypotheses of the type $E[X] = a_1 \mu_{(1)}^{(o)} + \dots + a_k \mu_{(k)}^{(o)}$ where the $\mu_j^{(o)}$'s, $j = 1, \dots, k$, are specified. Since we do not know the distributions of Y_1, \dots, Y_k , let us convert the iid variables on Y_j , $j = 1, \dots, k$, to iid variables on X_j , say X_1, \dots, X_n , $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, where Σ is unknown. First, the observations on Y_1, \dots, Y_k are transformed into observations on the X_j 's. The problem then involves a single normal population whose covariance

matrix is unknown. An example of this type is Y_1 representing a $p \times 1$ vector before a certain process, such as administering a drug to a patient; in this instance, Y_1 could consist of measurements on p characteristics observed in a patient. Observations on Y_2 will then be the measurements on the same p characteristics after the process such as after administering the drug to the patient. Then $Y_{2q} - Y_{1q} = X_q$ will represent the variable corresponding to the difference in the measurements on the q -th characteristic. Let the hypothesis be $H_0 : \mu = \mu_o$ (given), Σ being unknown. Note that once the observations on X_j are taken, then the individual $\mu_{(j)}$'s are irrelevant as they no longer are of any use. Once the X_j 's are determined, one can compute the sum of products matrix S in X_j . In this case, the test statistic is $u = n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$, which is distributed as a type-2 beta with parameters $(\frac{p}{2}, \frac{n-p}{2})$. Then, $u \sim \frac{p}{n-p} F$ where F is an F random variable having p and $n - p$ degrees of freedom, that is, an $F_{p, n-p}$ random variable, n being the sample size. Thus, the test criterion is applied as follows: Determine the observed value of u and the corresponding observed value of $F_{p, n-p}$ that is, $\frac{n-p}{p} u$, and then

$$\text{reject } H_0 \text{ if } \frac{n-p}{p} u \geq F_{p, n-p, \alpha}, \text{ with } Pr\{F_{p, n-p} \geq F_{p, n-p, \alpha}\} = \alpha. \quad (6.3.7)$$

Example 6.3.2. Five motivated individuals were randomly selected and subjected to an exercise regimen for a month. The exercise program promoters claim that the subjects can expect a weight loss of 5 kg as well as a 2-in. reduction in lower stomach girth by the end of the month period. Let Y_1 and Y_2 denote the two component vectors representing weight and girth before starting the routine and at the end of the exercise program, respectively. The following are the observations on the five individuals:

$$(Y_1, Y_2) = \begin{bmatrix} (85, 85) \\ (40, 41) \end{bmatrix}, \begin{bmatrix} (80, 70) \\ (40, 45) \end{bmatrix}, \begin{bmatrix} (75, 73) \\ (36, 36) \end{bmatrix}, \begin{bmatrix} (70, 71) \\ (38, 38) \end{bmatrix}, \begin{bmatrix} (70, 68) \\ (35, 34) \end{bmatrix}.$$

Obviously, Y_1 and Y_2 are dependent variables having a joint distribution. We will assume that the difference $X = Y_1 - Y_2$ has a real Gaussian distribution, that is, $X \sim N_2(\mu, \Sigma)$, $\Sigma > O$. Under this assumption, the observations on X are

$$X_1 = \begin{bmatrix} 85 - 85 \\ 40 - 41 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, X_2 = \begin{bmatrix} 10 \\ -5 \end{bmatrix}, X_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Let $\mathbf{X} = [X_1, X_2, \dots, X_5]$ and $\mathbf{X} - \bar{\mathbf{X}} = [X_1 - \bar{X}, \dots, X_5 - \bar{X}]$, both being 2×5 matrix. The observed sample average \bar{X} , the claim of the exercise routine promoters $\mu = \mu_o$ as

well as other relevant quantities are as follows:

$$\begin{aligned}\bar{X} &= \frac{1}{5}(X_1 + \cdots + X_5) = \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix}, \mu_o = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \\ X_1 - \bar{X} &= \begin{bmatrix} 0 \\ -1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -13 \\ 0 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 37 \\ -20 \end{bmatrix}, X_3 - \bar{X} = \frac{1}{5} \begin{bmatrix} -3 \\ 5 \end{bmatrix}, X_4 - \bar{X} = \frac{1}{5} \begin{bmatrix} -18 \\ 5 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} -3 \\ 10 \end{bmatrix};\end{aligned}$$

$$\mathbf{X} - \bar{\mathbf{X}} = [X_1 - \bar{X}, \dots, X_5 - \bar{X}] = \frac{1}{5} \begin{bmatrix} -13 & 37 & -3 & -18 & -3 \\ 0 & -20 & 5 & 5 & 10 \end{bmatrix};$$

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \frac{1}{5^2} \begin{bmatrix} -13 & 37 & -3 & -18 & -3 \\ 0 & -20 & 5 & 5 & 10 \end{bmatrix} \begin{bmatrix} -13 & 0 \\ 37 & -20 \\ -3 & 5 \\ -18 & 5 \\ -3 & 10 \end{bmatrix}$$

$$= \frac{1}{5^2} \begin{bmatrix} 1880 & -875 \\ -875 & 550 \end{bmatrix} = \frac{1}{5^2} A, A = \begin{bmatrix} 1880 & -875 \\ -875 & 550 \end{bmatrix} \Rightarrow S^{-1} = 25A^{-1};$$

$$|A| = 1880(550) - (875)^2 = 214975; \text{Cof}(A) = \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix}; A^{-1} = \frac{\text{Cof}(A)}{|A|};$$

$$A^{-1} = \frac{1}{268375} \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix}, S^{-1} = 25A^{-1};$$

$$\bar{X} - \mu_o = \frac{1}{5} \begin{bmatrix} 13 \\ -5 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \end{bmatrix} = -\frac{1}{5} \begin{bmatrix} 12 \\ 15 \end{bmatrix}.$$

The test statistic being $w = \frac{(n-p)}{p} n(\bar{X} - \mu_o)' S^{-1} (\bar{X} - \mu_o)$, its observed value is

$$\begin{aligned}v &= \frac{(5-2)}{2} 5 \frac{5^2}{5^2} \frac{1}{268375} [12 \quad 15] \begin{bmatrix} 550 & 875 \\ 875 & 1880 \end{bmatrix} \begin{bmatrix} 12 \\ 15 \end{bmatrix} \\ &= \frac{3}{2} \frac{5}{268375} [(12)^2(550) + (15)^2(1880) + 2(12)(15)(875)] = 22.84.\end{aligned}$$

Letting the significance level of the test be $\alpha = 0.05$, the critical value is $F_{p, n-p, \alpha} = F_{2, 3, 0.05} = 9.55$. Since $22.84 > 9.55$, H_o is thus rejected.

6.3.3. Independent Gaussian populations

Consider k independent p -variate real Gaussian populations whose individual distribution is $N_p(\mu_{(j)}, \Sigma_j)$, $\Sigma_j > O$, $j = 1, \dots, k$. Given simple random samples of sizes

n_1, \dots, n_k from these k populations, we may wish to test a hypothesis on a given linear functions of the mean values, that is, $H_o : a_1\mu_{(1)} + \dots + a_k\mu_{(k)} = \mu_0$ where a_1, \dots, a_k are known constants and μ_0 is a given quantity under the null hypothesis. We have already discussed this problem for the case of known covariance matrices. When the Σ_j 's are all unknown or some of them are known while others are not, the MLE's of the unknown covariance matrices turn out to be the respective sample sums of products matrices divided by the corresponding sample sizes. This will result in a linear function of independent Wishart matrices whose distribution proves challenging to determine, even for the null case.

Special case of two independent Gaussian populations

Consider the special case of two independent real Gaussian populations having identical covariance matrices. that is, let the populations be $Y_{1q} \sim N_p(\mu_{(1)}, \Sigma)$, $\Sigma > O$, the Y_{1q} 's, $q = 1, \dots, n_1$, being iid, and $Y_{2q} \sim N_p(\mu_{(2)}, \Sigma)$, $\Sigma > O$, the Y_{2q} 's, $q = 1, \dots, n_2$, being iid. Let the sample $p \times n_1$ and $p \times n_2$ matrices be denoted as $\mathbf{Y}_1 = (Y_{11}, \dots, Y_{1n_1})$ and $\mathbf{Y}_2 = (Y_{21}, \dots, Y_{2n_2})$ and let the sample averages be $\bar{Y}_j = \frac{1}{n_j}(Y_{j1} + \dots + Y_{jn_j})$, $j = 1, 2$. Let $\bar{\mathbf{Y}}_j = (\bar{Y}_j, \dots, \bar{Y}_j)$, a $p \times n_j$ matrix whose columns are equal to \bar{Y}_j , $j = 1, 2$, and let

$$S_j = (\mathbf{Y}_j - \bar{\mathbf{Y}}_j)(\mathbf{Y}_j - \bar{\mathbf{Y}}_j)', \quad j = 1, 2,$$

be the corresponding sample sum of products matrices. Then, S_1 and S_2 are independently distributed as Wishart matrices having $n_1 - 1$ and $n_2 - 1$ degrees of freedom, respectively. As the sum of two independent $p \times p$ real or complex matrices having matrix-variate gamma distributions with the same scale parameter matrix is again gamma distributed with the shape parameters summed up and the same scale parameter matrix, we observe that since the two populations are independently distributed, $S_1 + S_2 \equiv S$ has a Wishart distribution having $n_1 + n_2 - 2$ degrees of freedom. We now consider a hypothesis of the type $\mu_{(1)} = \mu_{(2)}$. In order to do away with the unknown common mean value, we may consider the real p -vector $U = \bar{Y}_1 - \bar{Y}_2$, so that $E(U) = O$ and $\text{Cov}(U) = \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma = (\frac{1}{n_1} + \frac{1}{n_2})\Sigma = \frac{n_1+n_2}{n_1n_2}\Sigma$. The MLE of this pooled covariance matrix is $\frac{1}{n_1+n_2}S$ where S is Wishart distributed with $n_1 + n_2 - 2$ degrees of freedom. Then, following through the steps included in Sect. 6.3.1 with the parameter m now being $n_1 + n_2 - 2$, the power of S will become $\frac{(n_1+n_2-2+1)}{2} - \frac{p+1}{2}$ when integrating out S . Letting the null hypothesis be $H_o : E[Y_1] - E[Y_2] = \delta$ (specified), such as $\delta = 0$, the function resulting from integrating out S is

$$c \left[\frac{(n_1 + n_2)}{n_1 n_2} u \right]^{\frac{p}{2}-1} \left[1 + \frac{(n_1 + n_2)}{n_1 n_2} u \right]^{-\frac{1}{2}(n_1+n_2-1)} \quad (6.3.8)$$

where c is the normalizing constant, so that $w = \frac{(n_1+n_2)}{n_1n_2}(\bar{Y}_1 - \bar{Y}_2 - \delta)'S^{-1}(\bar{Y}_1 - \bar{Y}_2 - \delta)$ is distributed as a type-2 beta with the parameters $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$. Writing $w = \frac{p}{n_1+n_2-1-p}F_{p, n_1+n_2-1-p}$, this F is seen to be an F statistic having p and $n_1 + n_2 - 1 - p$ degrees of freedom. We will state these results as theorems.

Theorem 6.3.2. *Let the $p \times p$ real positive definite matrices X_1 and X_2 be independently distributed as real matrix-variate gamma random variables with densities*

$$f_j(X_j) = \frac{|B|^{\alpha_j}}{\Gamma_p(\alpha_j)} |X_j|^{\alpha_j - \frac{p+1}{2}} e^{-\text{tr}(BX_j)}, \quad B > O, \quad X_j > O, \quad \Re(\alpha_j) > \frac{p-1}{2}, \quad (6.3.9)$$

$j=1,2$, and zero elsewhere. Then, as can be seen from (5.2.6), the Laplace transform associated with X_j or that of f_j , denoted as $L_{X_j}(*T)$, is

$$L_{X_j}(*T) = |I + B^{-1} *T|^{-\alpha_j}, \quad I + B^{-1} *T > O, \quad j = 1, 2. \quad (i)$$

Accordingly, $U_1 = X_1 + X_2$ has a real matrix-variate gamma density with the parameters $(\alpha_1 + \alpha_2, B)$ whose associated Laplace transform is

$$L_{U_1}(*T) = |I + B^{-1} *T|^{-(\alpha_1+\alpha_2)}, \quad (ii)$$

and $U_2 = a_1X_1 + a_2X_2$ has the Laplace transform

$$L_{U_2}(*T) = |I + a_1B^{-1} *T|^{-\alpha_1} |I + a_2B^{-1} *T|^{-\alpha_2}, \quad (iii)$$

whenever $I + a_jB^{-1} *T > O, \quad j = 1, 2$, where a_1 and a_2 are real scalar constants.

It follows from (ii) that $X_1 + X_2$ is also real matrix-variate gamma distributed. When $a_1 \neq a_2$, it is very difficult to invert (iii) in order to obtain the corresponding density. This can be achieved by expanding one of the determinants in (iii) in terms of zonal polynomials, say the second one, after having first taken $|I + a_1B^{-1} *T|^{-(\alpha_1+\alpha_2)}$ out as a factor in this instance.

Theorem 6.3.3. *Let $Y_j \sim N_p(\mu_{(j)}, \Sigma), \quad \Sigma > O, \quad j = 1, 2$, be independent p -variate real Gaussian distributions sharing the same covariance matrix. Given a simple random sample of size n_1 from Y_1 and a simple random sample of size n_2 from Y_2 , let the sample averages be denoted by \bar{Y}_1 and \bar{Y}_2 and the sample sums of products matrices, by S_1 and S_2 , respectively. Consider the hypothesis $H_o : \mu_{(1)} - \mu_{(2)} = \delta$ (given). Letting $S = S_1 + S_2$ and*

$$w = \frac{(n_1 + n_2)}{n_1n_2}(\bar{Y}_1 - \bar{Y}_2 - \delta)'S^{-1}(\bar{Y}_1 - \bar{Y}_2 - \delta), \quad w \sim \frac{p}{n_1 + n_2 - 1 - p}F_{p, n_1+n_2-1-p} \quad (iv)$$

where F_{p, n_1+n_2-1-p} denotes an F distribution with p and $n_1 + n_2 - 1 - p$ degrees of freedom, or equivalently, w is distributed as a type-2 beta variable with the parameters $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$. We reject the null hypothesis H_0 if $\frac{n_1+n_2-1-p}{p} w \geq F_{p, n_1+n_2-1-p, \alpha}$ with

$$\Pr\{F_{p, n_1+n_2-1-p} \geq F_{p, n_1+n_2-1-p, \alpha}\} = \alpha \quad (vi)$$

at a given significance level α .

Theorem 6.3.4. Let w be as defined in Theorem 6.3.3. Then $w_1 = \frac{1}{w}$ is a real scalar type-2 beta variable with the parameters $(\frac{n_1+n_2-1-p}{2}, \frac{p}{2})$; $w_2 = \frac{w}{1+w}$ is a real scalar type-1 beta variable with the parameters $(\frac{p}{2}, \frac{(n_1+n_2-1-p)}{2})$; $w_3 = \frac{1}{1+w}$ is a real scalar type-1 beta variable with the parameters $(\frac{n_1+n_2-1-p}{2}, \frac{p}{2})$.

Those last results follow from the connections between real scalar type-1 and type-2 beta random variables. Results parallel to those appearing in (i) to (vi) and stated Theorems 6.3.1–6.3.4 can similarly be obtained for the complex case.

Example 6.3.3. Consider two independent populations whose respective distributions are $N_2(\mu_{(1)}, \Sigma_1)$ and $N_2(\mu_{(2j)}, \Sigma_2)$, $\Sigma_j > O$, $j = 1, 2$, and samples of sizes $n_1 = 4$ and $n_2 = 5$ from these two populations, respectively. Let the population covariance matrices be identical with $\Sigma_1 = \Sigma_2 = \Sigma$, the common covariance matrix being unknown, and let the observed sample vectors from the first population, $X_j \sim N_2(\mu_{(1)}, \Sigma)$, be

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Denoting the sample mean from the first population by \bar{X} and the sample sum of products matrix by S_1 , we have

$$\bar{X} = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \text{ and } S_1 = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})', \mathbf{X} = [X_1, X_2, X_3, X_4], \bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}],$$

the observations on these quantities being the following:

$$\begin{aligned} X_1 - \bar{X} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 \\ -2 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{4} \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \quad X_3 - \bar{X} = \frac{1}{4} \begin{bmatrix} 4 \\ 6 \end{bmatrix}, \quad X_4 - \bar{X} = \frac{1}{4} \begin{bmatrix} -4 \\ -6 \end{bmatrix}; \\ \mathbf{X} - \bar{\mathbf{X}} &= \frac{1}{4} \begin{bmatrix} 4 & -4 & 4 & -4 \\ -2 & 2 & 6 & -6 \end{bmatrix}, \quad S_1 = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \\ S_1 &= \frac{1}{4^2} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix}. \end{aligned}$$

Let the sample vectors from the second population denoted as Y_1, \dots, Y_5 be

$$Y_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Y_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Y_3 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad Y_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad Y_5 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Then, the sample average and the deviation vectors are the following:

$$\begin{aligned} \bar{Y} &= \frac{1}{5}[Y_1 + \dots + Y_5] = \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \\ Y_1 - \bar{Y} &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} -5 \\ 3 \end{bmatrix}, \quad Y_2 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \\ Y_3 - \bar{Y} &= \frac{1}{5} \begin{bmatrix} 0 \\ -7 \end{bmatrix}, \quad Y_4 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad Y_5 - \bar{Y} = \frac{1}{5} \begin{bmatrix} 5 \\ 3 \end{bmatrix}, \\ \mathbf{Y} - \bar{\mathbf{Y}} &= \frac{1}{5} \begin{bmatrix} -5 & 0 & 0 & 0 & 5 \\ 3 & -2 & -7 & 3 & 3 \end{bmatrix}, \quad S_2 = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})'; \\ S_2 &= \frac{1}{5^2} \begin{bmatrix} 50 & 0 \\ 0 & 80 \end{bmatrix}; \quad S = S_1 + S_2 = \frac{1}{16} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} + \frac{1}{25} \begin{bmatrix} 50 & 0 \\ 0 & 80 \end{bmatrix} \\ &= \begin{bmatrix} 6.00 & 2.00 \\ 2.00 & 8.20 \end{bmatrix} \Rightarrow S^{-1} = \frac{\text{Cof}(S)}{|S|} \\ |S| &= 45.20; \quad S^{-1} = \frac{1}{45.20} \begin{bmatrix} 8.20 & -2.00 \\ -2.00 & 6.00 \end{bmatrix}. \end{aligned}$$

Letting the null hypothesis be

$$H_0 : \mu_{(1)} - \mu_{(2)} = \delta = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\bar{X} - \bar{Y} - \delta = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} = - \begin{bmatrix} 2.0 \\ 0.9 \end{bmatrix}; \quad n_1 = 4, \quad n_2 = 5.$$

Thus, test statistic is $u \sim F_{p, n_1+n_2-1-p}$ where

$$\begin{aligned} u &= \frac{(n_1 + n_2 - 1 - p)(n_1 + n_2)}{p n_1 n_2} (\bar{X} - \bar{Y} - \delta)' S^{-1} (\bar{X} - \bar{Y} - \delta) \\ &= \frac{(4 + 5 - 1 - 2)(4 + 5)}{2} \frac{1}{(4)(5) 45.2} \begin{bmatrix} -2.0 & -0.9 \end{bmatrix} \begin{bmatrix} 8.2 & -2.0 \\ -2.0 & 6.0 \end{bmatrix} \begin{bmatrix} -2.0 \\ -0.9 \end{bmatrix} \\ &\approx 0.91. \end{aligned}$$

Let us test H_0 at the 5% significance level. Since the required critical value is $F_{p, n_1+n_2-1-p, \alpha} = F_{2, 6, 0.05} = 5.14$ and $0.91 < 5.14$, the null hypothesis is not rejected.

6.3.4. Testing $\mu_1 = \dots = \mu_p$ when Σ is unknown in the real Gaussian case

Let the $p \times 1$ vector $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, the X_j 's being independently distributed. Let the $p \times 1$ vector of unities be denoted by J or $J' = (1, \dots, 1)$, and let A be a vector that is orthogonal to J so that $A'J = 0$. For example, we can take

$$A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -1 \end{bmatrix} \quad \text{when } p \text{ is even, } A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ 1 \\ -2 \\ 1 \end{bmatrix} \quad \text{or } A = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ -1 \\ 0 \end{bmatrix} \quad \text{when } p \text{ is odd.}$$

If the last component of A is zero, we are then ignoring the last component of X_j . Let $y_j = A'X_j$, $j = 1, \dots, n$, and the y_j 's be independently distributed. Then $y_j \sim N_1(A'\mu, A'\Sigma A)$, $A'\Sigma A > O$, is a univariate normal variable with mean value $A'\mu$ and variance $A'\Sigma A$. Consider the $p \times n$ sample matrix comprising the X_j 's, that is, $\mathbf{X} = (X_1, \dots, X_n)$. Let the sample average of the X_j 's be $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ and $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$. Then, the sample sum of products matrix $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$. Consider the $1 \times n$ vector $Y = (y_1, \dots, y_n) = (A'X_1, \dots, A'X_n) = A'\mathbf{X}$, $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n) = A'\bar{X}$, $\sum_{j=1}^n (y_j - \bar{y})^2 = A'(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'A = A'SA$. Let the null hypothesis be $H_0 : \mu_1 = \dots = \mu_p = \nu$, where ν is unknown, $\mu' = (\mu_1, \dots, \mu_p)$. Thus, H_0 is $A'\mu = \nu A'J = 0$. The joint density of y_1, \dots, y_n , denoted by L , is then

$$\begin{aligned}
L &= \prod_{j=1}^n \frac{e^{-\frac{1}{2A'\Sigma A}(y_j - A'\mu)^2}}{(2\pi)^{\frac{1}{2}}[A'\Sigma A]^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2(A'\Sigma A)}\sum_{j=1}^n (y_j - A'\mu)^2}}{(2\pi)^{\frac{n}{2}}[A'\Sigma A]^{\frac{n}{2}}} \\
&= \frac{e^{-\frac{1}{2A'\Sigma A}\{A'SA + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A\}}}{(2\pi)^{\frac{n}{2}}[A'\Sigma A]^{\frac{n}{2}}} \tag{i}
\end{aligned}$$

where

$$\begin{aligned}
\sum_{j=1}^n (y_j - A'\mu)^2 &= \sum_{j=1}^n (y_j - \bar{y} + \bar{y} - A'\mu)^2 = \sum_{j=1}^n (y_j - \bar{y})^2 + n(\bar{y} - A'\mu)^2 \\
&= \sum_{j=1}^n A'(X_j - \bar{X})(X_j - \bar{X})'A + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A \\
&= A'SA + nA'(\bar{X} - \mu)(\bar{X} - \mu)'A.
\end{aligned}$$

Let us determine the MLE's of μ and Σ . We have

$$\ln L = -\frac{n}{2}(2\pi) - \frac{n}{2} \ln A'\Sigma A - \frac{1}{2A'\Sigma A}[A'SA + n(A'(\bar{X} - \mu)(\bar{X} - \mu)'A)].$$

On differentiating $\ln L$ with respect to μ_1 and equating the result to zero, we have

$$\begin{aligned}
\frac{\partial}{\partial \mu_1} \ln L = 0 &\Rightarrow nA' \left[\frac{\partial}{\partial \mu_1} \{ \bar{X}\bar{X}' - \bar{X}\mu' - \mu\bar{X}' + \mu\mu' \} \right] A = O \\
&\Rightarrow nA'[-\bar{X}[1, 0, \dots, 0] - \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \bar{X}' + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mu' + \mu[1, 0, \dots, 0]]A = O \\
&\Rightarrow 2a_1 A'(\bar{X} - \mu) = 0 \Rightarrow \hat{\mu} = \bar{X} \tag{ii}
\end{aligned}$$

since the equation holds for each μ_j , $j = 1, \dots, p$, and $A' = (a_1, \dots, a_p)$, $a_j \neq 0$, $j = 1, \dots, p$, A being fixed. As well, $(\bar{X} - \mu)(\bar{X} - \mu)' = \bar{X}\bar{X}' - \bar{X}\mu' - \mu\bar{X}' + \mu\mu'$. Now, consider differentiating $\ln L$ with respect to an element of Σ , say, σ_{11} , at $\hat{\mu} = \bar{X}$:

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{11}} \ln L &= 0 \\
&\Rightarrow -\frac{2n a_1^2 \sigma_{11}}{2A'\Sigma A} + \frac{A'SA}{2(A'\Sigma A)^2} (2a_1^2 \sigma_{11}) = 0 \\
&\Rightarrow A'\hat{\Sigma}A = \frac{1}{n} A'SA
\end{aligned}$$

for each element in Σ and hence $\hat{\Sigma} = \frac{1}{n}S$. Thus,

$$\max_{\Omega} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{(A'SA)^{\frac{n}{2}}}. \quad (iii)$$

Under H_o , $A'\mu = 0$ and consequently the maximum under H_o is the following:

$$\max_{H_o} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{[A'(S + n\bar{X}\bar{X}')A]^{\frac{n}{2}}}. \quad (iv)$$

Accordingly, the λ -criterion is

$$\lambda = \frac{(A'SA)^{\frac{n}{2}}}{[A'(S + n\bar{X}\bar{X}')A]^{\frac{n}{2}}} = \frac{1}{[1 + n\frac{A'\bar{X}\bar{X}'A}{A'SA}]^{\frac{n}{2}}}. \quad (6.3.10)$$

We would reject H_o for small values of λ or for large values of $u \equiv nA'\bar{X}\bar{X}'A/A'SA$ where \bar{X} and S are independently distributed. Observing that $S \sim W_p(n-1, \Sigma)$, $\Sigma > O$ and $\bar{X} \sim N_p(\mu, \frac{1}{n}\Sigma)$, $\Sigma > O$, we have

$$\frac{n}{A'\Sigma A} A'\bar{X}\bar{X}'A \sim \chi_1^2 \text{ and } \frac{A'SA}{A'\Sigma A} \sim \chi_{n-1}^2.$$

Hence, $(n-1)u$ is a F statistic with 1 and $n-1$ degrees of freedom, and the null hypothesis is to be rejected whenever

$$v \equiv n(n-1) \frac{A'\bar{X}\bar{X}'A}{A'SA} \geq F_{1, n-1, \alpha}, \text{ with } Pr\{F_{1, n-1} \geq F_{1, n-1, \alpha}\} = \alpha. \quad (6.3.11)$$

Example 6.3.4. Consider a real bivariate Gaussian $N_2(\mu, \Sigma)$ population where $\Sigma > O$ is unknown. We would like to test the hypothesis $H_o : \mu_1 = \mu_2$, $\mu' = (\mu_1, \mu_2)$, so that $\mu_1 - \mu_2 = 0$ under this null hypothesis. Let the sample be X_1, X_2, X_3, X_4 , as specified in Example 6.3.3. Let $A' = (1, -1)$ so that $A'\mu = 0$ under H_o . With the same observation vectors as those comprising the first sample in Example 6.3.3, $A'X_1 = (1)$, $A'X_2 = (-2)$, $A'X_3 = (-1)$, $A'X_4 = (0)$. Letting $y = A'X_j$, the observations on y_j are $(1, -2, -1, 0)$ or $A'\mathbf{X} = A'[X_1, X_2, X_3, X_4] = [1, -2, -1, 0]$. The sample sum of products matrix as evaluated in the first part of Example 6.3.3 is

$$S_1 = \frac{1}{16} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \Rightarrow A'S_1A = \frac{1}{16} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{80}{16} = 5.$$

Our test statistic is

$$v = n(n-1) \frac{A'\bar{X}\bar{X}'A}{A'S_1A} \sim F_{1, n-1}, \quad n = 4.$$

Let the significance level be $\alpha = 0.05$. the observed values of $A'\bar{X}\bar{X}'A$, $A'S_1A$, v , and the tabulated critical value of $F_{1, n-1, \alpha}$ are the following:

$$A'\bar{X}\bar{X}'A = \frac{1}{4}; \quad A'S_1A = \frac{80}{16} = 5;$$

$$v = 4(3)\left(\frac{1}{5 \times 4}\right) = 0.6; \quad F_{1, n-1, \alpha} = F_{1, 3, 0.05} = 10.13.$$

As $0.6 < 10.13$, H_o is not rejected.

6.3.5. Likelihood ratio test for testing $H_o : \mu_1 = \dots = \mu_p$ when Σ is unknown

In the entire parameter space Ω of a $N_p(\mu, \Sigma)$ population, μ is estimated by the sample average \bar{X} and, as previously determined, the maximum of the likelihood function is

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}} \quad (i)$$

where S is the sample sum of products matrix and n is the sample size. Under the hypothesis $H_o : \mu_1 = \dots = \mu_p = v$, where v is unknown, this v is estimated by $\hat{v} = \frac{1}{np} \sum_{i,j} x_{ij} = \frac{1}{p} J' \bar{X}$, $J' = (1, \dots, 1)$, the $p \times 1$ sample vectors $X'_j = (x_{1j}, \dots, x_{pj})$, $j = 1, \dots, n$, being independently distributed. Thus, under the null hypothesis H_o , the population covariance matrix is estimated by $\frac{1}{n}(S + n(\bar{X} - \hat{\mu})(\bar{X} - \hat{\mu})')$, and, proceeding as was done to obtain Eq. (6.3.3), the λ -criterion reduces to

$$\lambda = \frac{|S|^{\frac{n}{2}}}{|S + n(\bar{X} - \hat{\mu})(\bar{X} - \hat{\mu})'|^{\frac{n}{2}}} \quad (6.3.12)$$

$$= \frac{1}{(1+u)^{\frac{n}{2}}}, \quad u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu}). \quad (6.3.13)$$

Given the structure of u in (6.3.13), we can take the Gaussian population covariance matrix Σ to be the identity matrix I , as was explained in Sect. 6.3.1. Observe that

$$(\bar{X} - \hat{\mu})' = (\bar{X} - \frac{1}{p} J J' \bar{X})' = \bar{X}' [I - \frac{1}{p} J J'] \quad (ii)$$

where $I - \frac{1}{p} J J'$ is idempotent of rank $p - 1$; hence there exists an orthonormal matrix P , $P P' = I$, $P' P = I$, such that

$$I - \frac{1}{p} J J' = P \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} P' \Rightarrow$$

$$\sqrt{n}(\bar{X} - \hat{\mu})' = \sqrt{n} \bar{X}' (I - \frac{1}{p} J J') = \sqrt{n} \bar{X}' P \begin{bmatrix} I_{p-1} & O \\ O' & 0 \end{bmatrix} = [V_1', 0], \quad V = \sqrt{n} \bar{X}' P,$$

where V_1 is the subvector of the first $p - 1$ components of V . Then the quadratic form u , which is our test statistic, reduces to the following:

$$u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu}) = [V_1', 0] S^{-1} \begin{bmatrix} V_1 \\ 0 \end{bmatrix} = V_1' S^{11} V_1,$$

$$S^{-1} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}.$$

We note that the test statistic u has the same structure that of u in Theorem 6.3.1 with p replaced by $p - 1$. Accordingly, $u = n(\bar{X} - \hat{\mu})' S^{-1} (\bar{X} - \hat{\mu})$ is distributed as a real scalar type-2 beta variable with the parameters $\frac{p-1}{2}$ and $\frac{n-(p-1)}{2}$, so that $\frac{n-p+1}{p-1} u \sim F_{p-1, n-p+1}$. Thus, the test criterion consists of

$$\text{rejecting } H_o \text{ if the observed value of } \frac{n-p+1}{p-1} u \geq F_{p-1, n-p+1, \alpha},$$

$$\text{with } Pr\{F_{p-1, n-p+1} \geq F_{p-1, n-p+1, \alpha}\} = \alpha. \quad (6.3.14)$$

Example 6.3.5. Let the population be $N_2(\mu, \Sigma)$, $\Sigma > O$, $\mu' = (\mu_1, \mu_2)$ and the null hypothesis be $H_o : \mu_1 = \mu_2 = \nu$ where ν and Σ are unknown. The sample values, as specified in Example 6.3.3, are

$$X_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \Rightarrow \bar{X} = \frac{1}{4} \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

The maximum likelihood estimate of μ under H_o , is

$$\hat{\mu} = \frac{1}{p} J J' \bar{X}, \quad J = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

and

$$(\bar{X} - \hat{\mu})' = \bar{X}' [I - \frac{1}{p} J J'] = \bar{X}' [I - \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}] = \frac{1}{4} [-1 \quad 1].$$

As previously calculated, the sample sum of products matrix is

$$S_1 = \frac{1}{4^2} \begin{bmatrix} 64 & 32 \\ 32 & 80 \end{bmatrix} \Rightarrow S_1^{-1} = \frac{16}{4096} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix}; \quad n = 4, \quad p = 2.$$

The test statistic v and its observed value are

$$\begin{aligned}
 v &= \frac{(n-p+1)}{(p-1)} n(\bar{X} - \hat{\mu})' S_1^{-1} (\bar{X} - \hat{\mu}) \sim F_{p-1, n-p+1} = F_{1, 3} \\
 &= \frac{(4-2+1)}{(2-1)} (4) \frac{1}{4^2} \frac{1}{256} \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 80 & -32 \\ -32 & 64 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
 &= \frac{(3)(4)}{(4^2)(256)} [(-1)^2(80) + (1)^2(64) - 2(32)(-1)(1)] \\
 &= 0.61.
 \end{aligned}$$

At significance level $\alpha = 0.05$, the tabulated critical value $F_{1, 3, 0.05}$ is 10.13, and since the observed value 0.61 is less than 10.13, H_0 is not rejected.

6.4. Testing Hypotheses on the Population Covariance Matrix

Let the $p \times 1$ independent vectors X_j , $j = 1, \dots, n$, have a p -variate real nonsingular $N_p(\mu, \Sigma)$ distribution and the $p \times n$ matrix $\mathbf{X} = (X_1, \dots, X_n)$ be the sample matrix. Denoting the sample average by $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ and letting $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$, each column of $\bar{\mathbf{X}}$ being equal to \bar{X} , the sample sum of products matrix is $S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$. We have already established that S is Wishart distributed with $m = n - 1$ degrees of freedom, that is, $S \sim W_p(m, \Sigma)$, $\Sigma > O$. Letting $S_\mu = (\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'$ where $\mathbf{M} = (\mu, \dots, \mu)$, each of its column being the $p \times 1$ vector μ , $S_\mu \sim W_p(n, \Sigma)$, $\Sigma > O$, where the number of degrees of freedom is n itself whereas the number of degrees of freedom associated with S is $m = n - 1$. Let us consider the hypothesis $H_0 : \Sigma = \Sigma_0$ where Σ_0 is a given known matrix and μ is unspecified. Then, the MLE's of μ and Σ in the entire parameter space are $\hat{\mu} = \bar{X}$ and $\hat{\Sigma} = \frac{1}{n}S$, and the joint density of the sample values X_1, \dots, X_n , denoted by L , is given by

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \frac{n}{2}\text{tr}(\bar{X} - \mu)(\bar{X} - \mu)'}. \quad (6.4.1)$$

Thus, as previously determined, the maximum of L in the parameter space $\Omega = \{(\mu, \Sigma) | \Sigma > O\}$ is

$$\max_{\Omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}, \quad (i)$$

the maximum of L under the null hypothesis $H_0 : \Sigma = \Sigma_0$ being given by

$$\max_{H_0} L = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma_0^{-1}S)}}{(2\pi)^{\frac{np}{2}} |\Sigma_0|^{\frac{n}{2}}}. \quad (ii)$$

Then, the λ -criterion is the following:

$$\lambda = \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}}} |\Sigma_o^{-1} S|^{\frac{n}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma_o^{-1} S)}. \quad (6.4.2)$$

Letting $u = \lambda^{\frac{2}{n}}$,

$$u = \frac{e^p}{n^p} |\Sigma_o^{-1} S| e^{-\frac{1}{n}\text{tr}(\Sigma_o^{-1} S)}, \quad (6.4.3)$$

and we would reject H_o for small values of u since it is a monotonically increasing function of λ , which means that the null hypothesis ought to be rejected for large values of $\text{tr}(\Sigma_o^{-1} S)$ as the exponential function dominates the polynomial function for large values of the argument. Let us determine the distribution of $w = \text{tr}(\Sigma_o^{-1} S)$ whose Laplace transform with parameter s is

$$L_w(s) = E[e^{-sw}] = E[e^{-s \text{tr}(\Sigma_o^{-1} S)}]. \quad (iii)$$

This can be evaluated by integrating out over the density of S which has a Wishart distribution with $m = n - 1$ degrees of freedom when μ is estimated:

$$L_w(s) = \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma|^{\frac{m}{2}}} \int_{S>O} |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1} S) - s \text{tr}(\Sigma_o^{-1} S)} dS. \quad (iv)$$

The exponential part is $-\frac{1}{2}\text{tr}(\Sigma^{-1} S) - s \text{tr}(\Sigma_o^{-1} S) = -\frac{1}{2}\text{tr}[(\Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}})(I + 2s \Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}})]$ and hence,

$$L_w(s) = |I + 2s \Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}}|^{-\frac{m}{2}}. \quad (6.4.4)$$

The null case, $\Sigma = \Sigma_o$

In this case, $\Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}} = I$, so that

$$L_w(s) = |I + 2sI|^{-\frac{m}{2}} = (1 + 2s)^{-\frac{mp}{2}} \Rightarrow w \sim \chi_{mp}^2. \quad (6.4.5)$$

Thus, the test criterion is the following:

$$\text{Reject } H_o \text{ if } w \geq \chi_{mp, \alpha}^2, \text{ with } Pr\{\chi_{mp}^2 \geq \chi_{mp, \alpha}^2\} = \alpha. \quad (6.4.6)$$

When μ is known, it is used instead of its MLE to determine S_μ , and the resulting criterion consists of rejecting H_o whenever the observed $w_\mu = \text{tr}(\Sigma_o^{-1} S_\mu) \geq \chi_{np, \alpha}^2$ where n is the sample size. These results are summarized in the following theorem.

Theorem 6.4.1. *Let the null hypothesis be $H_o : \Sigma = \Sigma_o$ (given) and $w = \text{tr}(\Sigma_o^{-1} S)$ where S is the sample sum of products matrix. Then, the null distribution of $w = \text{tr}(\Sigma_o^{-1} S)$*

has a real scalar chisquare distribution with $(n-1)p$ degrees of freedom when the estimate of μ , namely $\hat{\mu} = \bar{X}$, is utilized to compute S ; when μ is specified, w has a chisquare distribution having np degrees of freedom where n is the sample size.

The non-null density of w

The non-null density of w is available from (6.4.4). Let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of $\Sigma^{\frac{1}{2}} \Sigma_o^{-1} \Sigma^{\frac{1}{2}}$. Then $L_w(s)$ in (6.4.4) can be re-expressed as follows:

$$L_w(s) = \prod_{j=1}^p (1 + 2\lambda_j s)^{-\frac{m}{2}}. \quad (6.4.7)$$

This is the Laplace transform of a variable of the form $w = \lambda_1 w_1 + \dots + \lambda_p w_p$ where w_1, \dots, w_p are independently distributed real scalar chisquare random variables, each having $m = n - 1$ degrees of freedom, where $\lambda_j > 0$, $j = 1, \dots, p$. The distribution of linear combinations of chisquare random variables corresponds to the distribution of quadratic forms; the reader may refer to Mathai and Provost (1992) for explicit representations of their density functions.

Note 6.4.1. If the population mean value μ is known, then one can proceed by making use of μ instead of the sample mean to determine S_μ , in which case n , the sample size, ought to be used instead of $m = n - 1$ in the above discussion.

6.4.1. Arbitrary moments of λ

From (6.4.2), the h -th moment of the λ -criterion for testing $H_o : \Sigma = \Sigma_o$ (given) in a real nonsingular $N_p(\mu, \Sigma)$ population, is obtained as follows:

$$\begin{aligned} \lambda^h &= \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}}} |\Sigma_o^{-1}|^{\frac{nh}{2}} |S|^{\frac{nh}{2}} e^{-\frac{h}{2} \text{tr}(\Sigma_o^{-1} S)} \Rightarrow \\ E[\lambda^h] &= \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}} 2^{\frac{(n-1)p}{2}} |\Sigma_o|^{\frac{nh}{2}} |\Sigma|^{\frac{n-1}{2}} \Gamma_p(\frac{n-1}{2})} \int_{S>O} |S|^{\frac{nh}{2} + \frac{n-1}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - \frac{h}{2} \text{tr}(\Sigma_o^{-1} S)} dS \\ &= \frac{e^{\frac{np}{2}} 2^{p(\frac{nh}{2} + \frac{n-1}{2})} \Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{n^{\frac{np}{2}} 2^{\frac{(n-1)p}{2}} |\Sigma_o|^{\frac{nh}{2}} |\Sigma|^{\frac{n-1}{2}} \Gamma_p(\frac{n-1}{2})} |\Sigma^{-1} + h \Sigma_o^{-1}|^{-\frac{(nh}{2} + \frac{n-1}{2})} \\ &= e^{\frac{np}{2}} \left(\frac{2}{n}\right)^{\frac{np}{2}} \frac{|\Sigma|^{\frac{nh}{2}} \Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{|\Sigma_o|^{\frac{nh}{2}} \Gamma_p(\frac{n-1}{2})} |I + h \Sigma \Sigma_o^{-1}|^{-\frac{(nh}{2} + \frac{n-1}{2})} \end{aligned} \quad (6.4.8)$$

for $\Re(\frac{nh}{2} + \frac{n-1}{2}) > \frac{p-1}{2}$, $I + h\Sigma\Sigma_o^{-1} > O$. Under $H_o : \Sigma = \Sigma_o$, we have $|I + h\Sigma\Sigma_o^{-1}|^{-(\frac{nh}{2} + \frac{n-1}{2})} = (1+h)^{-p(\frac{nh}{2} + \frac{n-1}{2})}$. Thus, the h -th null moment is given by

$$E[\lambda^h | H_o] = e^{\frac{np}{2}h} \left(\frac{2}{n}\right)^{\frac{np}{2}} \frac{\Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{\Gamma_p(\frac{n-1}{2})} (1+h)^{-p(\frac{nh}{2} + \frac{n-1}{2})} \quad (6.4.9)$$

for $\Re(\frac{nh}{2} + \frac{n-1}{2}) > \frac{p-1}{2}$.

6.4.2. The asymptotic distribution of $-2 \ln \lambda$ when testing $H_o : \Sigma = \Sigma_o$

Let us determine the asymptotic distribution of $-2 \ln \lambda$ where λ is the likelihood ratio statistic for testing $H_o : \Sigma = \Sigma_o$ (specified) in a real nonsingular $N_p(\mu, \Sigma)$ population, as $n \rightarrow \infty$, n being the sample size. This distribution can be determined by expanding both real matrix-variate gamma functions appearing in (6.4.9) and applying Stirling's approximation formula as given in (6.5.14) by letting $\frac{n}{2}(1+h) \rightarrow \infty$ in the numerator gamma functions and $\frac{n}{2} \rightarrow \infty$ in the denominator gamma functions. Then, we have

$$\begin{aligned} \frac{\Gamma_p(\frac{nh}{2} + \frac{n-1}{2})}{\Gamma_p(\frac{n-1}{2})} &= \prod_{j=1}^p \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n}{2} - \frac{1}{2} - \frac{j-1}{2})} \\ &\rightarrow \prod_{j=1}^p \frac{(2\pi)^{\frac{1}{2}} [\frac{n}{2}(1+h)]^{\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j}{2}} e^{-\frac{n}{2}(1+h)}}{(2\pi)^{\frac{1}{2}} [\frac{n}{2}]^{\frac{n}{2} - \frac{1}{2} - \frac{j}{2}} e^{-\frac{n}{2}}} \\ &= \left(\frac{n}{2}\right)^{\frac{np}{2}} e^{-\frac{np}{2}h} (1+h)^{\frac{np}{2}(1+h) - \frac{p}{2} - \frac{p(p+1)}{4}}. \end{aligned}$$

Hence, from (6.4.9)

$$E[\lambda^h | H_o] \rightarrow (1+h)^{-\frac{p(p+1)}{4}} \text{ as } n \rightarrow \infty, \quad (6.4.10)$$

where $(1+h)^{-\frac{p(p+1)}{4}}$ is the h -th moment of the distribution of $e^{-y/2}$ when $y \sim \chi_{\frac{p(p+1)}{2}}^2$. Thus, under H_o , $-2 \ln \lambda \rightarrow \chi_{\frac{p(p+1)}{2}}^2$ as $n \rightarrow \infty$. For general procedures leading to asymptotic normality, see Mathai (1982).

Theorem 6.4.2. *Letting λ be the likelihood ratio statistic for testing $H_o : \Sigma = \Sigma_o$ (given) on the covariance matrix of a real nonsingular $N_p(\mu, \Sigma)$ distribution, the null distribution of $-2 \ln \lambda$ is asymptotically (as then sample size tends to ∞) that of a real scalar chisquare random variable having $\frac{p(p+1)}{2}$ degrees of freedom, where n denotes the sample size. This number of degrees of freedom is also equal to the number of parameters restricted by the null hypothesis.*

Note 6.4.2. Sugiura and Nagao (1968) have shown that the test based on the statistic λ as specified in (6.4.2) is biased whereas it becomes unbiased upon replacing n , the sample size, by the degrees of freedom $n - 1$ in (6.4.2). Accordingly, percentage points are then computed for $-2 \ln \lambda_1$, where λ_1 is the statistic λ given in (6.4.2) wherein $n - 1$ is substituted to n . Korin (1968), Davis (1971), and Nagarsenker and Pillai (1973) computed 5% and 1% percentage points for this test statistic. Davis and Field (1971) evaluated the percentage points for $p = 2(1)10$ and $n = 6(1)30(5)50, 60, 120$ and Korin (1968), for $p = 2(1)10$.

Example 6.4.1. Let us take the same 3-variate real Gaussian population $N_3(\mu, \Sigma)$, $\Sigma > O$ and the same data as in Example 6.3.1, so that intermediate calculations could be utilized. The sample size is 5 and the sample values are the following:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix},$$

the sample average and the sample sum of products matrix being

$$\bar{X} = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, S = \frac{1}{5^2} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix}.$$

Let us consider the hypothesis $\Sigma = \Sigma_o$ where

$$\Sigma_o = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} \Rightarrow |\Sigma_o| = 10, \text{Cof}(\Sigma_o) = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix};$$

$$\Sigma_o^{-1} = \frac{\text{Cof}(\Sigma_o)}{|\Sigma_o|} = \frac{1}{10} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix};$$

$$\begin{aligned} \text{tr}(\Sigma_o^{-1}S) &= \frac{1}{(10)(5^2)} \text{tr} \left\{ \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} \begin{bmatrix} 270 & -110 & -170 \\ -110 & 80 & 85 \\ -170 & 85 & 170 \end{bmatrix} \right\} \\ &= \frac{1}{(10)(5^2)} [5(270) + (4(80) + 2(85)) + (2(85) + 6(170))] \\ &= 12.12; n = 5, p = 3. \end{aligned}$$

Let us test the null hypothesis at the significance level $\alpha = 0.05$. The distribution of the test statistic w and the tabulated critical value are as follows:

$$w = \text{tr}(\Sigma_o^{-1}S) \sim \chi_{(n-1)p}^2 \simeq \chi_{12}^2; \chi_{12, 0.05}^2 = 21.03.$$

As the observed value $12.12 < 21.03$, H_0 is not rejected. The asymptotic distribution of $-2 \ln \lambda$, as $n \rightarrow \infty$, is $\chi_{p(p+1)/2}^2 \simeq \chi_6^2$ where λ is the likelihood ratio criterion statistic. Since $\chi_{6, 0.05}^2 = 12.59$ and $12.59 > 12.12$, we still do not reject H_0 as $n \rightarrow \infty$.

6.4.3. Tests on Wilks' concept of generalized variance

The concept of *generalized variance* was explained in Chap. 5. The sample generalized variance is simply the determinant of S , the sample sum of products matrix. When the population is p -variate Gaussian, it has already been shown in Chap. 5 that S is Wishart distributed with $m = n - 1$ degrees of freedom, n being the sample size, and parameter matrix $\Sigma > O$, which is the population covariance matrix. When the population is multivariate normal, several types of tests of hypotheses involve the sample generalized variance. The first author has given the exact distributions of such tests, see Mathai (1972a,b) and Mathai and Rathie (1971).

6.5. The Sphericity Test or Testing if $H_0 : \Sigma = \sigma^2 I$, Given a $N_p(\mu, \Sigma)$ Sample

When the covariance matrix $\Sigma = \sigma^2 I$, where $\sigma^2 > 0$ is a real scalar quantity, the ellipsoid $(X - \mu)' \Sigma^{-1} (X - \mu) = c > 0$, which represents a specific contour of constant density for a nonsingular $N_p(\mu, \Sigma)$ distribution, becomes the sphere defined by the equation $\frac{1}{\sigma^2} (X - \mu)' (X - \mu) = c$ or $\frac{1}{\sigma^2} ((x_1 - \mu_1)^2 + \dots + (x_p - \mu_p)^2) = c > 0$, whose center is located at the point μ ; hence the test's name, the sphericity test. Given a $N_p(\mu, \Sigma)$ sample of size n , the maximum of the likelihood function in the entire parameter space is

$$\sup_{\Omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}},$$

as was previously established. However, under the null hypothesis $H_0 : \Sigma = \sigma^2 I$, $\text{tr}(\Sigma^{-1} S) = (\sigma^2)^{-1} (\text{tr}(S))$ and $|\Sigma| = (\sigma^2)^p$. Thus, if we let $\theta = \sigma^2$ and substitute $\hat{\mu} = \bar{X}$ in L , under H_0 the loglikelihood function will be $\ln L_{\omega} = -\frac{np}{2} \ln(2\pi) - \frac{np}{2} \ln \theta - \frac{1}{2\theta} \text{tr}(S)$. Differentiating this function with respect to θ and equating the result to zero produces the following estimator for θ :

$$\hat{\theta} = \hat{\sigma}^2 = \frac{\text{tr}(S)}{np}. \quad (6.5.1)$$

Accordingly, the maximum of the likelihood function under H_0 is the following:

$$\max_{\omega} L = \frac{n^{\frac{np}{2}} e^{-\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} \left[\frac{\text{tr}(S)}{p} \right]^{\frac{n}{2}}}.$$

Thus, the λ -criterion for testing

$$H_o : \Sigma = \sigma^2 I, \sigma^2 > 0 \text{ (unknown)}$$

is

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{[\frac{\text{tr}(S)}{p}]^{\frac{n}{2}}}. \quad (6.5.2)$$

In the complex Gaussian case when $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma = \Sigma^* > O$ where an asterisk indicates the conjugate transpose, $\tilde{X}_j = X_{j1} + iX_{j2}$ where X_{j1} and X_{j2} are real $p \times 1$ vectors and $i = \sqrt{-1}$. The covariance matrix associated with \tilde{X}_j is then defined as

$$\begin{aligned} \text{Cov}(\tilde{X}_j) &= E[(\tilde{X}_j - \tilde{\mu})(\tilde{X}_j - \tilde{\mu})^*] \\ &= E[(X_{j1} - \mu_{(1)}) + i(X_{j2} - \mu_{(2)})][(X_{j1} - \mu_{(1)})' - i(X_{j2} - \mu_{(2)})'] \\ &= \Sigma_{11} + \Sigma_{22} + i(\Sigma_{21} - \Sigma_{12}) \equiv \Sigma, \text{ with } \mu = \mu_{(1)} + i\mu_{(2)}, \end{aligned}$$

where Σ is assumed to be Hermitian positive definite, with $\Sigma_{11} = \text{Cov}(X_{j1})$, $\Sigma_{22} = \text{Cov}(X_{j2})$, $\Sigma_{12} = \text{Cov}(X_{j1}, X_{j2})$ and $\Sigma_{21} = \text{Cov}(X_{j2}, X_{j1})$. Thus, the hypothesis of sphericity in the complex Gaussian case is $\Sigma = \sigma^2 I$ where σ is real and positive. Then, under the null hypothesis $\tilde{H}_o : \Sigma = \sigma^2 I$, the Hermitian form $\tilde{Y}^* \Sigma \tilde{Y} = c > 0$ where c is real and positive, becomes $\sigma^2 \tilde{Y}^* \tilde{Y} = c \Rightarrow |\tilde{y}_1|^2 + \dots + |\tilde{y}_p|^2 = \frac{c}{\sigma^2} > 0$, which defines a sphere in the complex space, where $|\tilde{y}_j|$ denotes the absolute value or modulus of \tilde{y}_j . If $\tilde{y}_j = y_{j1} + iy_{j2}$ with $i = \sqrt{-1}$, y_{j1}, y_{j2} being real, then $|\tilde{y}_j|^2 = y_{j1}^2 + y_{j2}^2$.

The joint density of the sample values in the real Gaussian case is the following:

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1} (X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} S) - n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)}}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}}$$

where $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, $X_j, j = 1, \dots, n$ are iid $N_p(\mu, \Sigma)$, $\Sigma > O$. We have already derived the maximum of L in the entire parameter space Ω , which, in the real case, is

$$\sup_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}}, \quad (6.5.3)$$

where S is the sample sum of products matrix. Under H_o , $|\Sigma|^{\frac{n}{2}} = (\sigma^2)^{\frac{np}{2}}$ and $\text{tr}(\Sigma^{-1} S) = \frac{1}{\sigma^2}(s_{11} + \dots + s_{pp}) = \frac{1}{\sigma^2} \text{tr}(S)$. Thus, the maximum likelihood estimator of σ^2 is $\frac{1}{np} \text{tr}(S)$. Accordingly, the λ -criterion is

$$\lambda = |S|^{\frac{n}{2}} / \left(\frac{\text{tr}(S)}{p} \right)^{\frac{np}{2}} \Rightarrow u_1 = \lambda^{\frac{2}{n}} = \frac{p^p |S|}{[\text{tr}(S)]^p}, \quad (6.5.4)$$

in the real case. Interestingly, $(u_1)^{1/p}$ is the ratio of the geometric mean of the eigenvalues of S to their arithmetic mean. The structure remains the same in the complex domain, in which case $\det(\tilde{S})$ is replaced by the absolute value $|\det(\tilde{S})|$ so that

$$\tilde{u}_1 = \frac{p^p |\det(\tilde{S})|}{[\text{tr}(\tilde{S})]^p}. \quad (6.5a.1)$$

For arbitrary h , the h -th moment of u_1 in the real case can be obtained by integrating out over the density of S , which, as explained in Sect. 5.5, 5.5a, is a Wishart density with $n - 1 = m$ degrees of freedom and parameter matrix $\Sigma > O$. However, when the null hypothesis H_o holds, $\Sigma = \sigma^2 I_p$, so that the h -th moment in the real case is

$$E[u_1^h | H_o] = \frac{p^{ph}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})(\sigma^2)^{\frac{mp}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2\sigma^2}\text{tr}(S)} (\text{tr}(S))^{-ph} dS. \quad (i)$$

In order to evaluate this integral, we replace $[\text{tr}(S)]^{-ph}$ by an equivalent integral:

$$[\text{tr}(S)]^{-ph} = \frac{1}{\Gamma(ph)} \int_{x=0}^{\infty} x^{ph-1} e^{-x(\text{tr}(S))} dx, \quad \Re(h) > 0. \quad (ii)$$

Then, substituting (ii) in (i), the exponent becomes $-\frac{1}{2\sigma^2}(1 + 2\sigma^2 x)(\text{tr}(S))$. Now, letting $S_1 = \frac{1}{2\sigma^2}(1 + 2\sigma^2 x)S \Rightarrow dS = (2\sigma^2)^{\frac{p(p+1)}{2}} (1 + 2\sigma^2 x)^{-\frac{p(p+1)}{2}} dS_1$, and we have

$$\begin{aligned} E[u_1^h | H_o] &= \frac{(2\sigma^2)^{ph}}{\Gamma_p(\frac{m}{2})} \frac{p^{ph}}{\Gamma(ph)} \int_0^{\infty} x^{ph-1} (1 + 2\sigma^2 x)^{-\frac{(m}{2}+h)p} dx \\ &\quad \times \int_{S_1>O} |S_1|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\text{tr}(S_1)} dS_1 \\ &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \frac{p^{ph}}{\Gamma(ph)} \int_0^{\infty} y^{ph-1} (1 + y)^{-\frac{(m}{2}+h)p} dy, \quad y = 2\sigma^2 x \\ &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} p^{ph} \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \quad \Re(h) > 0, \quad m = n - 1. \end{aligned} \quad (6.5.5)$$

The corresponding h -th moment in the complex case is the following:

$$E[\tilde{u}_1^h | H_o] = \frac{\tilde{\Gamma}_p(m + h)}{\tilde{\Gamma}_p(m)} p^{ph} \frac{\tilde{\Gamma}(mp)}{\tilde{\Gamma}(mp + ph)}, \quad \Re(h) > 0, \quad m = n - 1. \quad (6.5a.2)$$

By making use of the multiplication formula for gamma functions, one can expand the real gamma function $\Gamma(mz)$ as follows:

$$\Gamma(mz) = (2\pi)^{\frac{1-m}{2}} m^{mz-\frac{1}{2}} \Gamma(z) \Gamma(z + \frac{1}{m}) \cdots \Gamma(z + \frac{m-1}{m}), \quad m = 1, 2, \dots, \quad (6.5.6)$$

and for $m = 2$, we have the duplication formula

$$\Gamma(2z) = \pi^{-\frac{1}{2}} 2^{2z-1} \Gamma(z) \Gamma\left(z + \frac{1}{2}\right). \quad (6.5.7)$$

Then on applying (6.5.6),

$$\frac{p^{ph} \Gamma\left(\frac{mp}{2}\right)}{\Gamma\left(\frac{mp}{2} + ph\right)} = \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m}{2} + h\right)} \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} + h + \frac{j}{p}\right)}. \quad (iii)$$

Moreover, it follows from the definition of the real matrix-variate gamma functions that

$$\frac{\Gamma_p\left(\frac{m}{2} + h\right)}{\Gamma_p\left(\frac{m}{2}\right)} = \prod_{j=1}^p \frac{\Gamma\left(\frac{m}{2} - \frac{j-1}{2} + h\right)}{\Gamma\left(\frac{m}{2} - \frac{j-1}{2}\right)}. \quad (iv)$$

On canceling $\Gamma\left(\frac{m}{2} + h\right)/\Gamma\left(\frac{m}{2}\right)$ when multiplying (iii) by (iv), we are left with

$$E[u_1^h | H_o] = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} - \frac{j}{2}\right)} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} - \frac{j}{2} + h\right)}{\Gamma\left(\frac{m}{2} + \frac{j}{p} + h\right)} \right\}, \quad m = n - 1. \quad (6.5.8)$$

The corresponding h -th moment in the complex case is the following:

$$E[\tilde{u}_1^h | H_o] = \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}\left(m + \frac{j}{p}\right)}{\tilde{\Gamma}(m - j)} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}(m - j + h)}{\tilde{\Gamma}\left(m + \frac{j}{p} + h\right)} \right\}, \quad m = n - 1. \quad (6.5a.3)$$

For $h = s - 1$, one can treat $E[u_1^{s-1} | H_o]$ as the Mellin transform of the density of u_1 in the real case. Letting this density be denoted by $f_{u_1}(u_1)$, it can be expressed in terms of a G-function as follows:

$$f_{u_1}(u_1 | H_o) = c_1 G_{p-1, p-1}^{p-1, 0} \left[u_1 \left| \begin{matrix} \frac{m}{2} + \frac{j}{p} - 1, j=1, \dots, p-1 \\ \frac{m}{2} - \frac{j}{2} - 1, j=1, \dots, p-1 \end{matrix} \right. \right], \quad 0 \leq u_1 \leq 1, \quad (6.5.9)$$

and $f_{u_1}(u_1 | H_o) = 0$ elsewhere, where

$$c_1 = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma\left(\frac{m}{2} + \frac{j}{p}\right)}{\Gamma\left(\frac{m}{2} - \frac{j}{2}\right)} \right\},$$

the corresponding density in the complex case being the following:

$$\tilde{f}_{\tilde{u}_1|H_o}(\tilde{u}_1) = \tilde{c}_1 \tilde{G}_{p-1,p-1}^{p-1,0} \left[\tilde{u}_1 \left| \begin{matrix} m+\frac{j}{p}-1, & j=1,\dots,p-1 \\ m-j-1, & j=1,\dots,p-1 \end{matrix} \right. \right], \quad 0 \leq |\tilde{u}_2| \leq 1, \quad (6.5a.4)$$

and $\tilde{f}_{\tilde{u}_1}(\tilde{u}_1) = 0$ elsewhere, where \tilde{G} is a real G-function whose parameters are different from those appearing in (6.5.9), and

$$\tilde{c}_1 = \left\{ \prod_{j=1}^{p-1} \frac{\tilde{\Gamma}(m + \frac{j}{p})}{\tilde{\Gamma}(m - j)} \right\}.$$

For computable series representation of a G-function with general parameters, the reader may refer to Mathai (1970a, 1993). Observe that u_1 in the real case is structurally a product of $p - 1$ mutually independently distributed real scalar type-1 beta random variables with the parameters $(\alpha_j = \frac{m}{2} - \frac{j}{2}, \beta_j = \frac{j}{2} + \frac{j}{p})$, $j = 1, \dots, p - 1$. In the complex case, \tilde{u}_1 is structurally a product of $p - 1$ mutually independently distributed real scalar type-1 beta random variables with the parameters $(\alpha_j = m - j, \beta_j = j + \frac{j}{p})$, $j = 1, \dots, p - 1$. This observation is stated as a result.

Theorem 6.5.1. *Consider the sphericity test statistic for testing the hypothesis $H_o : \Sigma = \sigma^2 I$ where $\sigma^2 > 0$ is an unknown real scalar. Let u_1 and the corresponding complex quantity \tilde{u}_1 be as defined in (6.5.4) and (6.5a.1) respectively. Then, in the real case, u_1 is structurally a product of $p - 1$ independently distributed real scalar type-1 beta random variables with the parameters $(\alpha_j = \frac{m}{2} - \frac{j}{2}, \beta_j = \frac{j}{2} + \frac{j}{p})$, $j = 1, \dots, p - 1$, and, in the complex case, \tilde{u}_1 is structurally a product of $p - 1$ independently distributed real scalar type-1 beta random variables with the parameters $(\alpha_j = m - j, \beta_j = j + \frac{j}{p})$, $j = 1, \dots, p - 1$, where $m = n - 1$, $n =$ the sample size.*

For certain special cases, one can represent (6.5.9) and (6.5a.4) in terms of known elementary functions. Some such cases are now being considered.

Real case: $p = 2$

In the real case, for $p = 2$

$$E[u_1^h | H_o] = \frac{\Gamma(\frac{m}{2} + \frac{1}{2}) \Gamma(\frac{m}{2} - \frac{1}{2} + h)}{\Gamma(\frac{m}{2} - \frac{1}{2}) \Gamma(\frac{m}{2} + \frac{1}{2} + h)} = \frac{\frac{m}{2} - \frac{1}{2}}{\frac{m}{2} - \frac{1}{2} + h}.$$

This means u_1 is a real type-1 beta variable with the parameters $(\alpha = \frac{m}{2} - \frac{1}{2}, \beta = 1)$. The corresponding result in the complex case is that \tilde{u}_1 is a real type-1 beta variable with the parameters $(\alpha = m - 1, \beta = 1)$.

Real case: $p = 3$

In the real case

$$E[u_1^h | H_0] = \frac{\Gamma(\frac{m}{2} + \frac{1}{3})\Gamma(\frac{m}{2} + \frac{2}{3})}{\Gamma(\frac{m}{2} - \frac{1}{2})\Gamma(\frac{m}{2} - 1)} \frac{\Gamma(\frac{m}{2} - \frac{1}{2} + h)\Gamma(\frac{m}{2} - 1 + h)}{\Gamma(\frac{m}{2} + \frac{1}{3} + h)\Gamma(\frac{m}{2} + \frac{2}{3} + h)},$$

so that u_1 is equivalent to the product of two independently distributed real type-1 beta random variables with the parameters $(\alpha_j, \beta_j) = (\frac{m}{2} - \frac{j}{2}, \frac{j}{2} + \frac{j}{3})$, $j = 1, 2$. This density can be obtained by treating $E[u_1^h | H_0]$ for $h = s - 1$ as the Mellin transform of the density of u_1 . The density is then available by taking the inverse Mellin transform. Thus, again denoting it by $f_{u_1}(u_1)$, we have

$$\begin{aligned} f_{u_1}(u_1 | H_0) &= c_3 \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \phi_3(s) ds, \quad c > \frac{1}{2} \\ &= c_3 \left[\sum_{\nu=0}^{\infty} R_{\nu} + \sum_{\nu=0}^{\infty} R'_{\nu} \right], \\ c_3 &= \frac{\Gamma(\frac{m}{2} + \frac{1}{3})\Gamma(\frac{m}{2} + \frac{2}{3})}{\Gamma(\frac{m}{2} - \frac{1}{2})\Gamma(\frac{m}{2} - 1)}, \\ \phi_3(s) &= \frac{\Gamma(\frac{m}{2} - \frac{1}{2} - 1 + s)\Gamma(\frac{m}{2} - 1 - 1 + s)}{\Gamma(\frac{m}{2} + \frac{1}{3} - 1 + s)\Gamma(\frac{m}{2} + \frac{2}{3} - 1 + s)} u_1^{-s}, \end{aligned}$$

where R_{ν} is the residue of the integrand $\phi_3(s)$ at the poles of $\Gamma(\frac{m}{2} - \frac{3}{2} + s)$ and R'_{ν} is the residue of the integrand $\phi_3(s)$ at the pole of $\Gamma(\frac{m}{2} - 2 + s)$. Letting $s_1 = \frac{m}{2} - \frac{3}{2} + s$,

$$\begin{aligned} R_{\nu} &= \lim_{s \rightarrow -\nu + \frac{3}{2} - \frac{m}{2}} \phi_3(s) = \lim_{s_1 \rightarrow -\nu} [(s_1 + \nu) u_1^{\frac{m}{2} - \frac{3}{2}} \frac{\Gamma(s_1)\Gamma(-\frac{1}{2} + s_1)u_1^{-s_1}}{\Gamma(\frac{1}{3} + \frac{1}{2} + s_1)\Gamma(\frac{2}{3} + \frac{1}{2} + s_1)}] \\ &= u_1^{\frac{m}{2} - \frac{3}{2}} \frac{(-1)^{\nu}}{\nu!} \frac{\Gamma(-\frac{1}{2} - \nu)}{\Gamma(\frac{1}{3} + \frac{1}{2} - \nu)\Gamma(\frac{2}{3} + \frac{1}{2} - \nu)} u_1^{\nu}. \end{aligned}$$

We can replace negative ν in the arguments of the gamma functions with positive ν by making use of the following formula:

$$\Gamma(a - \nu) = \frac{(-1)^{\nu} \Gamma(a)}{(-a + 1)_{\nu}}, \quad a \neq 1, 2, \dots, \quad \nu = 0, 1, \dots, \quad (6.5.10)$$

where for example, $(b)_{\nu}$ is the Pochhammer symbol

$$(b)_{\nu} = b(b + 1) \cdots (b + \nu - 1), \quad b \neq 0, \quad (b)_0 = 1, \quad (6.5.11)$$

so that

$$\begin{aligned}\Gamma\left(-\frac{1}{2}-\nu\right) &= \frac{(-1)^\nu \Gamma(-\frac{1}{2})}{(\frac{3}{2})_\nu}, \quad \Gamma\left(\frac{1}{3}+\frac{1}{2}-\nu\right) = \frac{(-1)^\nu \Gamma(\frac{1}{3}+\frac{1}{2})}{(-\frac{1}{3}+\frac{1}{2})_\nu}, \\ \Gamma\left(\frac{2}{3}+\frac{1}{2}-\nu\right) &= \frac{(-1)^\nu \Gamma(\frac{2}{3}+\frac{1}{2})}{(-\frac{2}{3}+\frac{1}{2})_\nu}.\end{aligned}$$

The sum of the residues then becomes

$$\sum_{\nu=0}^{\infty} R_\nu = \frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{1}{3}+\frac{1}{2})\Gamma(\frac{2}{3}+\frac{1}{2})} u_1^{\frac{m}{2}-\frac{3}{2}} {}_2F_1\left(-\frac{1}{3}+\frac{1}{2}; -\frac{2}{3}+\frac{1}{2}; \frac{3}{2}; u_1\right), \quad 0 \leq u_1 \leq 1.$$

It can be similarly shown that

$$\sum_{\nu=0}^{\infty} R'_\nu = \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{3}+1)\Gamma(\frac{2}{3}+1)} u_1^{\frac{m}{2}-2} {}_2F_1\left(-\frac{1}{3}, -\frac{2}{3}; \frac{1}{2}; u_1\right), \quad 0 \leq u_1 \leq 1.$$

Accordingly, the density of u_1 for $p = 3$ is the following:

$$\begin{aligned}f_1(u_1|H_o) &= c_3 \left\{ \frac{\Gamma(-\frac{1}{2})}{\Gamma(\frac{5}{6})\Gamma(\frac{7}{6})} u_1^{\frac{m}{2}-\frac{3}{2}} {}_2F_1\left(\frac{1}{6}, -\frac{1}{6}; \frac{3}{2}; u_1\right) \right. \\ &\quad \left. + \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{4}{3})\Gamma(\frac{5}{3})} u_1^{\frac{m}{2}-2} {}_2F_1\left(-\frac{1}{3}, -\frac{2}{3}; \frac{1}{2}; u_1\right) \right\}, \quad 0 \leq u_1 \leq 1 \quad (6.5.12)\end{aligned}$$

and $f_{u_1}(u_1|H_o) = 0$ elsewhere.

Real case: $p = 4$

In this case,

$$E[u_1^h|H_o] = c_4 \frac{\Gamma(\frac{m}{2}-\frac{3}{2}+s)\Gamma(\frac{m}{2}-2+s)\Gamma(\frac{m}{2}-\frac{5}{2}+s)}{\Gamma(\frac{m}{2}-\frac{3}{4}+s)\Gamma(\frac{m}{2}-\frac{2}{4}+s)\Gamma(\frac{m}{2}-\frac{1}{4}+s)},$$

where c_4 is the normalizing constant. However, noting that

$$\frac{\Gamma(\frac{m}{2}-\frac{3}{2}+s)}{\Gamma(\frac{m}{2}-\frac{1}{2}+s)} = \frac{1}{\frac{m}{2}-\frac{3}{2}+s},$$

there is one pole at $s = -\frac{m}{2} + \frac{3}{2}$. The poles of $\Gamma(\frac{m}{2} - \frac{5}{2} + s)$ occur at $s = -\frac{m}{2} + \frac{5}{2} - \nu$, $\nu = 0, 1, \dots$, and hence at $\nu = 1$, the pole coincides with the earlier pole and there is a pole of order 2 at $s = -\frac{m}{2} + \frac{3}{2}$. Each one of other poles of the integrand is simple, that is, of order 1. The second order pole will bring in a logarithmic function. As all the cases for which $p \geq 4$ will bring in poles of higher orders, they will not be herein discussed. The general expansion of a G-function of the type $G_{m,m}^{m,0}(\cdot)$ is provided in Mathai (1970a, 1993). In the complex case, starting from $p \geq 3$, poles of higher orders are coming in, so that the densities can only be written in terms of logarithms, psi and zeta functions; hence, these will not be considered. Observe that \tilde{u}_1 corresponds to product of independently distributed real type-1 beta random variables, even though the densities are available only in terms of logarithms, psi and zeta functions for $p \geq 3$. The null and non-null densities of the λ -criterion in the general case, were derived by the first author and some results obtained under the null distribution can also be found in Mathai and Saxena (1973). Several researchers have contributed to various aspects of the sphericity and multi-sample sphericity tests; for some of the first author's contributions, the reader may refer to Mathai and Rathie (1970) and Mathai (1977, 1984, 1986).

Gamma products such as those appearing in (6.5.8) and (6.5a.3) are frequently encountered when considering various types of tests on the parameters of a real or complex Gaussian or certain other types of distributions. Structural representations in the form of product of independently distributed real scalar type-1 beta random variables occur in numerous situations. Thus, a general asymptotic result on the h -th moment of such products of type-1 beta random variables will be derived. This is now stated as a result.

Theorem 6.5.2. *Let u be a real scalar random variable whose h -th moment is of the form*

$$E[u^h] = \frac{\Gamma_p(\alpha + \alpha h + \gamma)}{\Gamma_p(\alpha + \gamma)} \frac{\Gamma_p(\alpha + \gamma + \delta)}{\Gamma_p(\alpha + \alpha h + \gamma + \delta)} \quad (6.5.13)$$

where $\Gamma_p(\cdot)$ is a real matrix-variate gamma function on $p \times p$ real positive definite matrices, α is real, γ is bounded, δ is real, $0 < \delta < \infty$ and h is arbitrary. Then, as $\alpha \rightarrow \infty$, $-2 \ln u \rightarrow \chi_{2p\delta}^2$, a real chisquare random variable having $2p\delta$ degrees of freedom, that is, a real gamma random variable with the parameters ($\alpha = p\delta$, $\beta = 2$).

Proof: On expanding the real matrix-variate gamma functions, we have the following:

$$\frac{\Gamma_p(\alpha + \gamma + \delta)}{\Gamma_p(\alpha + \gamma)} = \prod_{j=1}^p \frac{\Gamma(\alpha + \gamma + \delta - \frac{j-1}{2})}{\Gamma(\alpha + \gamma - \frac{j-1}{2})}. \quad (i)$$

$$\frac{\Gamma_p(\alpha(1+h) + \gamma)}{\Gamma_p(\alpha(1+h) + \gamma + \delta)} = \prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma - \frac{j-1}{2})}{\Gamma(\alpha(1+h) + \gamma + \delta - \frac{j-1}{2})}. \quad (ii)$$

Consider the following form of Stirling's asymptotic approximation formula for gamma functions, namely,

$$\Gamma(z + \eta) \approx \sqrt{2\pi} z^{z+\eta-\frac{1}{2}} e^{-z} \text{ for } |z| \rightarrow \infty \text{ and } \eta \text{ bounded.} \quad (6.5.14)$$

On applying this asymptotic formula to the gamma functions appearing in (i) and (ii) for $\alpha \rightarrow \infty$, we have

$$\prod_{j=1}^p \frac{\Gamma(\alpha + \gamma + \delta - \frac{j-1}{2})}{\Gamma(\alpha + \gamma - \frac{j-1}{2})} \rightarrow \alpha^{p\delta}$$

and

$$\prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma - \frac{j-1}{2})}{\Gamma(\alpha(1+h) + \gamma + \delta - \frac{j-1}{2})} \rightarrow [\alpha(1+h)]^{-p\delta}, \quad (iii)$$

so that

$$E[u^h] \rightarrow (1+h)^{-p\delta}. \quad (iv)$$

On noting that $E[u^h] = E[e^{h \ln u}] \rightarrow (1+h)^{-p\delta}$, it is seen that $\ln u$ has the mgf $(1+h)^{-p\delta}$ for $1+h > 0$ or $-2 \ln u$ has mgf $(1-2h)^{-p\delta}$ for $1-2h > 0$, which happens to be the mgf of a real scalar chisquare variable with $2p\delta$ degrees of freedom if $2p\delta$ is a positive integer or a real gamma variable with the parameters $(\alpha = p\delta, \beta = 2)$. Hence the following result.

Corollary 6.5.1. Consider a slightly more general case than that considered in Theorem 6.5.2. Let the h -th moment of u be of the form

$$E[u^h] = \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha(1+h) + \gamma_j)}{\Gamma(\alpha + \gamma_j)} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(\alpha + \gamma_j + \delta_j)}{\Gamma(\alpha(1+h) + \gamma_j + \delta_j)} \right\}. \quad (6.5.15)$$

Then as $\alpha \rightarrow \infty$, $E[u^h] \rightarrow (1+h)^{-(\delta_1 + \dots + \delta_p)}$, which implies that $-2 \ln u \rightarrow \chi_{2(\delta_1 + \dots + \delta_p)}^2$ whenever $2(\delta_1 + \dots + \delta_p)$ is a positive integer or, equivalently, $-2 \ln u$ tends to a real gamma variable with the parameters $(\alpha = \delta_1 + \dots + \delta_p, \beta = 2)$.

Let us examine the asymptotic distribution of the test statistic for the sphericity test in the light of Theorem 6.5.2. It is seen from (6.5.4) that $\lambda^h = u^{h \frac{n}{2}}$. Thus, by replacing h by $\frac{n}{2}h$ in (6.5.8) with $m = n - 1$, we have

$$E[\lambda^h | H_0] = \left\{ \prod_{j=1}^{p-1} \frac{\Gamma(\frac{n-1}{2} + \frac{j}{p})}{\Gamma(\frac{n-1}{2} - \frac{j}{2})} \right\} \left\{ \prod_{j=1}^{p-1} \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j}{2})}{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} + \frac{j}{p})} \right\}. \quad (6.5.16)$$

Then, it follows from Corollary 6.5.1 that $-2 \ln \lambda \rightarrow \chi_{2 \sum_{j=1}^{p-1} (\frac{j}{2} + \frac{j}{p})}^2$, a chi-square random variable having $2 \sum_{j=1}^{p-1} (\frac{j}{2} + \frac{j}{p}) = \frac{p(p-1)}{2} + (p-1) = \frac{(p-1)(p+2)}{2}$ degrees of freedom. Hence the following result:

Theorem 6.5.3. Consider the λ -criterion for testing the hypothesis of sphericity. Then, under the null hypothesis, $-2 \ln \lambda \rightarrow \chi_{\frac{(p-1)(p+2)}{2}}^2$, as the sample size $n \rightarrow \infty$. In the complex case, as $n \rightarrow \infty$, $-2 \ln \lambda \rightarrow \chi_{(p-1)(p+1)}^2$, a real scalar chisquare variable with $2[\frac{p(p-1)}{2} + \frac{(p-1)}{2}] = p(p-1) + (p-1) = (p-1)(p+1)$ degrees of freedom.

Note 6.5.1. We observe that the degrees of freedom of the real chisquare variable in the real scalar case is $\frac{(p-1)(p+2)}{2}$, which is also equal to the number of parameters restricted by the null hypothesis. Indeed, when $\Sigma = \sigma^2 I$, we have $\sigma_{ij} = 0, i \neq j$, which produces $\frac{p(p-1)}{2}$ restrictions and, since σ^2 is unknown, requiring that the diagonal elements are such that $\sigma_{11} = \dots = \sigma_{pp}$ produces $p-1$ additional restrictions for a total of $\frac{(p-1)(p+2)}{2}$ restrictions being imposed. Thus, the degrees of freedom of the asymptotic chisquare variable corresponds to the number of restrictions imposed by H_o , which, actually, is a general result.

6.6. Testing the Hypothesis that the Covariance Matrix is Diagonal

Consider the null hypothesis that Σ , the nonsingular covariance matrix of a p -variate real normal distribution, is diagonal, that is,

$$H_o : \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp}).$$

Since the population is assumed to be normally distributed, this implies that the components of the p -variate Gaussian vector are mutually independently distributed as univariate normal random variables whose respective variances are $\sigma_{jj}, j = 1, \dots, p$. Consider a simple random sample of size n from a nonsingular $N_p(\mu, \Sigma)$ population or, equivalently, let X_1, \dots, X_n be independently distributed as $N_p(\mu, \Sigma)$ vectors, $\Sigma > 0$. Under H_o, σ_{jj} is estimated by its MLE which is $\hat{\sigma}_{jj} = \frac{1}{n} s_{jj}$ where s_{jj} is the j -th diagonal element of $S = (s_{ij})$, the sample sum of products matrix. The maximum of the likelihood function under the null hypothesis is then

$$\max_{H_o} L = \prod_{j=1}^p \max_{H_o} L_j = \frac{1}{(2\pi)^{\frac{np}{2}} \prod_{j=1}^p [s_{jj}]^{\frac{n}{2}}} n^{\frac{np}{2}} e^{-\frac{1}{2}(np)},$$

the likelihood function being the joint density evaluated at an observed value of the sample. Observe that the overall maximum or the maximum in the entire parameter space remains the same as that given in (6.1.1). Thus, the λ -criterion is given by

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{\prod_{j=1}^p s_{jj}^{\frac{n}{2}}} \Rightarrow u_2 = \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{j=1}^p s_{jj}} \tag{6.6.1}$$

where $S \sim W_p(m, \Sigma)$, $\Sigma > O$, and $m = n - 1$, n being the sample size. Under H_0 , $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp})$. Then for an arbitrary h , the h -th moment of u_2 is available by taking the expected value of $\lambda^{\frac{2}{n}}$ with respect to the density of S , that is,

$$E[u_2^h | H_0] = \int_{S>O} \frac{|S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} (\prod_{j=1}^p s_{jj})^{-h}}{2^{\frac{mp}{2}} |\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} dS \tag{6.6.2}$$

where, under H_0 , $|\Sigma| = \sigma_{11} \cdots \sigma_{pp}$. As was done in Sect. 6.1.1, we may replace s_{jj}^{-h} by the equivalent integral,

$$s_{jj}^{-h} = \frac{1}{\Gamma(h)} \int_0^\infty x_j^{h-1} e^{-x_j(s_{jj})} dx_j, \Re(h) > 0.$$

Thus,

$$\prod_{j=1}^p s_{jj}^{-h} = \frac{1}{[\Gamma(h)]^p} \int_0^\infty \cdots \int_0^\infty x_1^{h-1} \cdots x_p^{h-1} e^{-\text{tr}(YS)} dx_1 \wedge \cdots \wedge dx_p \tag{i}$$

where $Y = \text{diag}(x_1, \dots, x_p)$, so that $\text{tr}(YS) = x_1 s_{11} + \cdots + x_p s_{pp}$. Then, (6.6.2) can be reexpressed as follow:

$$\begin{aligned} E[u_2^h | H_0] &= \frac{\int_0^\infty \cdots \int_0^\infty x_1^{h-1} \cdots x_p^{h-1}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) (\prod_{j=1}^p \sigma_{jj})^{\frac{mp}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}((\Sigma^{-1}+2Y)S)} dS \\ &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \frac{\int_0^\infty \cdots \int_0^\infty x_1^{h-1} \cdots x_p^{h-1}}{2^{\frac{mp}{2}} (\prod_{j=1}^p \sigma_{jj})^{\frac{mp}{2}}} \\ &\quad \times \left| \frac{(\Sigma^{-1} + 2Y)}{2} \right|^{-\left(\frac{m}{2}+h\right)} dx_1, \wedge \cdots \wedge dx_p, \end{aligned}$$

and observing that, under H_0 ,

$$\begin{aligned} \left| \frac{(\Sigma^{-1} + 2Y)}{2} \right| &= \left| \frac{\Sigma^{-1}}{2} \right| |I + 2\Sigma Y| \text{ with} \\ |I + 2\Sigma Y| &= (1 + 2\sigma_{11}y_1) \cdots (1 + 2\sigma_{pp}y_p), \end{aligned}$$

$$\begin{aligned}
E[u_2^h | H_0] &= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \prod_{j=1}^p \frac{1}{\Gamma(h)} \int_0^\infty y_j^{h-1} (1 + y_j)^{-(\frac{m}{2} + h)} dy_j, \quad y_j = 2x_j \sigma_{jj} \\
&= \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \left[\frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} \right]^p, \quad \Re(\frac{m}{2} + h) > \frac{p-1}{2}.
\end{aligned} \tag{6.6.3}$$

Thus,

$$\begin{aligned}
E[u_2^h | H_0] &= \left[\frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} \right]^p \prod_{j=1}^p \frac{\Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{m}{2} - \frac{j-1}{2})} \\
&= \frac{[\Gamma(\frac{m}{2})]^{p-1}}{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2})\}} \frac{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2} + h)\}}{[\Gamma(\frac{m}{2} + h)]^{p-1}}.
\end{aligned}$$

Denoting the density of u_2 as $f_{u_2}(u_2 | H_0)$, we can express it as an inverse Mellin transform by taking $h = s - 1$. Then,

$$f_{u_2}(u_2 | H_0) = c_{2,p-1} G_{p-1,p-1}^{p-1,0} \left[u_2 \left| \begin{matrix} \frac{m}{2}-1, \dots, \frac{m}{2}-1 \\ \frac{m}{2}-\frac{j}{2}-1, \quad j=1, \dots, p-1 \end{matrix} \right. \right], \quad 0 \leq u_2 \leq 1, \tag{6.6.4}$$

and zero elsewhere, where

$$c_{2,p-1} = \frac{[\Gamma(\frac{m}{2})]^{p-1}}{\{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} - \frac{j}{2})\}}.$$

Some special cases of this density are expounded below.

Real and complex cases: $p = 2$

When $p = 2$, u_2 has a real type-1 beta density with the parameters $(\alpha = \frac{m}{2} - \frac{1}{2}, \beta = \frac{1}{2})$ in the real case. In the complex case, it has a real type-1 beta density with the parameters $(\alpha = m - 1, \beta = 1)$.

Real and complex cases: $p = 3$

In this case, $f_{u_2}(u_2 | H_0)$ is given by

$$f_{u_2}(u_2 | H_0) = c_{2,2} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s) \Gamma(\frac{m}{2} - 2 + s)}{[\Gamma(\frac{m}{2} - 1 + s)]^2} u_2^{-s} ds.$$

The poles of the integrand are simple. Those coming from $\Gamma(\frac{m}{2} - \frac{3}{2} + s)$ occur at $s = -\frac{m}{2} + \frac{3}{2} - \nu$, $\nu = 0, 1, \dots$. The residue R_ν is the following:

$$R_\nu = u_2^{\frac{m}{2} - \frac{3}{2} + \nu} \frac{(-1)^\nu \Gamma(-\frac{1}{2} - \nu)}{\nu! [\Gamma(\frac{1}{2} - \nu)]^2} = \frac{\Gamma(-\frac{1}{2}) (\frac{1}{2})_\nu (\frac{1}{2})_\nu (-1)^\nu u_2^{\frac{m}{2} - \frac{3}{2} + \nu}}{[\Gamma(\frac{1}{2})]^2 (\frac{3}{2})_\nu \nu!}.$$

Summing the residues, we have

$$\sum_{\nu=0}^{\infty} R_\nu = \frac{\Gamma(-\frac{1}{2})}{[\Gamma(\frac{1}{2})]^2} u_2^{\frac{m}{2} - \frac{3}{2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; u_2\right), \quad 0 \leq u_2 \leq 1.$$

Now, consider the sum of the residues at the poles of $\Gamma(\frac{m}{2} - 2 + s)$. Observing that $\Gamma(\frac{m}{2} - 2 + s)$ cancels out one of the gamma functions in the denominator, namely $\Gamma(\frac{m}{2} - 1 + s) = (\frac{m}{2} - 2 - s)\Gamma(\frac{m}{2} - 2 + s)$, the integrand becomes

$$\frac{\Gamma(\frac{m}{2} - \frac{3}{2} + s) u_2^{-s}}{(\frac{m}{2} - 2 + s)\Gamma(\frac{m}{2} - 1 + s)},$$

the residue at the pole $s = -\frac{m}{2} + 2$ being $\frac{\Gamma(\frac{1}{2}) u_2^{\frac{m}{2} - 2}}{\Gamma(1)}$. Then, noting that $\Gamma(-\frac{1}{2}) = -2\Gamma(\frac{1}{2}) = -2\sqrt{\pi}$, the density is the following:

$$f_{u_2}(u_2|H_0) = c_{2,2} \left\{ \sqrt{\pi} u_2^{\frac{m}{2} - 2} - \frac{2}{\sqrt{\pi}} u_2^{\frac{m}{2} - \frac{3}{2}} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; u_2\right) \right\}, \quad 0 \leq u_2 \leq 1, \quad (6.6.5)$$

and zero elsewhere.

In the complex case, the integrand is

$$\frac{\Gamma(m - 2 + s)\Gamma(m - 3 + s)}{[\Gamma(m - 1 + s)]^2} u_2^{-s} = \frac{1}{(m - 2 + s)^2(m - 3 + s)} u_2^{-s},$$

and hence there is a pole of order 1 at $s = -m + 3$ and a pole of order 2 at $s = -m + 2$.

The residue at $s = -m + 3$ is $\frac{u_2^{m-3}}{(1)^2} = u_2^{m-3}$ and the residue at $s = -m + 2$ is given by

$$\lim_{s \rightarrow -m+2} \frac{\partial}{\partial s} (m - 2 + s)^2 \left[\frac{1}{(m - 2 + s)^2(m - 3 + s)} u_2^{-s} \right] = \lim_{s \rightarrow -m+2} \frac{\partial}{\partial s} \left[\frac{u_2^{-s}}{(m - 3 + s)} \right],$$

which gives the residue as $u_2^{m-2} \ln u_2 - u_2^{m-2}$. Thus, the sum of the residues is $u_2^{m-3} + u_2^{m-2} \ln u_2 - u_2^{m-2}$ and the constant part is

$$\frac{[\Gamma(m)]^2}{\Gamma(m-1)\Gamma(m-2)} = (m-1)^2(m-2), \quad m > 2,$$

so that the density is

$$f_{u_2}(u_2) = (m-1)^2(m-2)[u_2^{m-3} + u_2^{m-2} \ln u_2 - u_2^{m-2}], \quad 0 < u_2 \leq 1, \quad m \geq 3,$$

and zero elsewhere. Note that as $u_2 \rightarrow 0$, the limit of $\frac{u_2^{m-1}}{m-1} \ln u_2$ is zero. By integrating out over $0 < u_2 \leq 1$ while $m \geq 3$, it can be verified that $f_{u_2}(\cdot)$ is indeed a density function.

Real and complex cases: $p \geq 4$

As poles of higher orders are present when $p \geq 4$, both in the real and complex cases, the exact density function of the test statistic will not be herein explicitly given for those cases. Actually, the resulting densities would involve G-functions for which general expansions are for instance provided in Mathai (1993). The exact null and non-null densities of $u = \lambda^{\frac{2}{n}}$ have been previously derived by the first author. Percentage points accurate to the 11th decimal place are available from Mathai and Katiyar (1979a, 1980) for the null case; as well, various aspects of the distribution of the test statistic are discussed in Mathai and Rathie (1971) and Mathai (1973, 1984, 1985)

Let us now consider the asymptotic distribution of the λ -criterion under the null hypothesis,

$$H_o : \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{pp}).$$

Given the representation of the h -th moment of u_2 provided in (6.6.3) and referring to Corollary 6.5.1, it is seen that the sum of the δ_j 's is $\sum_{j=1}^{p-1} \delta_j = \sum_{j=1}^{p-1} \frac{j}{2} = \frac{p(p-1)}{4}$, so that the number of degrees of freedom of the asymptotic chisquare distribution is $2[\frac{p(p-1)}{4}] = \frac{p(p-1)}{2}$ which, as it should be, is the number of restrictions imposed by H_o , noting that when Σ is diagonal, $\sigma_{ij} = 0$, $i \neq j$, which produces $\frac{p(p-1)}{2}$ restrictions. Hence, the following result:

Theorem 6.6.1. *Let λ be the likelihood ratio criterion for testing the hypothesis that the covariance matrix Σ of a nonsingular $N_p(\mu, \Sigma)$ distribution is diagonal. Then, as $n \rightarrow \infty$, $-2 \ln \lambda \rightarrow \chi_{\frac{p(p-1)}{2}}^2$ in the real case. In the corresponding complex case, as $n \rightarrow \infty$, $-2 \ln \lambda \rightarrow \chi_{p(p-1)}^2$, a real scalar chisquare variable having $p(p-1)$ degrees of freedom.*

6.7. Equality of Diagonal Elements, Given that Σ is Diagonal, Real Case

In the case of a p -variate real nonsingular $N_p(\mu, \Sigma)$ population, whenever Σ is diagonal, the individual components are independently distributed as univariate normal random variables. Consider a simple random sample of size n , that is, a set of

$p \times 1$ vectors X_1, \dots, X_n , that are iid as $X_j \sim N_p(\mu, \Sigma)$ where it is assumed that $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Letting $X'_j = (x_{1j}, \dots, x_{pj})$, the joint density of the x_{rj} 's, $j = 1, \dots, n$, in the above sample, which is denoted by L_r , is given by

$$L_r = \prod_{j=1}^n \frac{e^{-\frac{1}{2\sigma_r^2}(x_{rj}-\mu_r)^2}}{(2\pi)^{\frac{1}{2}}(\sigma_r^2)^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2\sigma_r^2}\sum_{j=1}^n(x_{rj}-\mu_r)^2}}{(2\pi)^{\frac{n}{2}}(\sigma_r^2)^{\frac{n}{2}}}.$$

Then, on substituting the maximum likelihood estimators of μ_r and σ_r^2 in L_r , its maximum is

$$\max L_r = \frac{n^{\frac{n}{2}} e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} (s_{rr})^{\frac{n}{2}}}, \quad s_{rr} = \sum_{j=1}^n (x_{rj} - \bar{x}_r)^2.$$

Under the null hypothesis H_0 , $\sigma_1^2 = \dots = \sigma_p^2 \equiv \sigma^2$ and the MLE of σ^2 is a pooled estimate which is equal to $\frac{1}{np}(s_{11} + \dots + s_{pp})$. Thus, the λ -criterion is the following in this case:

$$\lambda = \frac{\sup_{\omega}}{\sup_{\Omega}} = \frac{[s_{11}s_{22}\cdots s_{pp}]^{\frac{n}{2}}}{\left(\frac{s_{11}+\cdots+s_{pp}}{p}\right)^{\frac{np}{2}}}. \quad (6.7.1)$$

If we let

$$u_3 = \lambda^{\frac{2}{n}} = \frac{p^p (\prod_{j=1}^p s_{jj})}{(\sum_{j=1}^p s_{jj})^p}, \quad (6.7.2)$$

then, for arbitrary h , the h -th moment of u_3 is the following:

$$E[u_3^h | H_0] = E\left[\frac{p^{ph} (\prod_{j=1}^p s_{jj}^h)}{(s_{11} + \dots + s_{pp})^{ph}}\right] = E\left[p^{ph} \left(\prod_{j=1}^p s_{jj}^h\right) (s_{11} + \dots + s_{pp})^{-ph}\right]. \quad (6.7.3)$$

Observe that $\frac{s_{jj}}{\sigma^2} \stackrel{iid}{\sim} \chi_{n-1}^2 = \chi_m^2$, $m = n - 1$, for $j = 1, \dots, p$, the density of s_{jj} being of the form

$$f_{s_{jj}}(s_{jj}) = \frac{s_{jj}^{\frac{m}{2}-1} e^{-\frac{s_{jj}}{2\sigma^2}}}{(2\sigma^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}, \quad 0 \leq s_{jj} < \infty, \quad m = n - 1 = 1, 2, \dots, \quad (i)$$

under H_0 . Note that $(s_{11} + \dots + s_{pp})^{-ph}$ can be replaced by an equivalent integral as

$$(s_{11} + \dots + s_{pp})^{-ph} = \frac{1}{\Gamma(ph)} \int_0^\infty x^{ph-1} e^{-x(s_{11}+\cdots+s_{pp})} dx, \quad \Re(h) > 0. \quad (ii)$$

Due to independence of the s_{jj} 's, the joint density of s_{11}, \dots, s_{pp} , is the product of the densities appearing in (i), and on integrating out s_{11}, \dots, s_{pp} , we end up with the following:

$$\frac{1}{(2\sigma^2)^{\frac{mp}{2}}} \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{m}{2} + h)}{\Gamma(\frac{m}{2})} \right\} \left\{ \prod_{j=1}^n \left[\frac{(1 + 2\sigma^2 x)}{2\sigma^2} \right]^{-\frac{m}{2} + h} \right\} = \frac{[\Gamma(\frac{m}{2} + h)]^p}{[\Gamma(\frac{m}{2})]^p} \left[\frac{(1 + 2\sigma^2 x)^{-p(\frac{m}{2} + h)}}{(2\sigma^2)^{-ph}} \right].$$

Now, the integral over x can be evaluated as follows:

$$\frac{(2\sigma^2)^{ph}}{\Gamma(ph)} \int_0^\infty x^{ph-1} (1 + 2\sigma^2 x)^{-p(\frac{m}{2} + h)} dx = \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \Re(h) > 0.$$

Thus,

$$E[u_3^h | H_0] = p^{ph} \frac{\Gamma^p(\frac{m}{2} + h)}{\Gamma^p(\frac{m}{2})} \frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)}, \Re(h) > 0. \quad (6.7.4)$$

The density of u_3 can be written in terms of an H-function. Since p is a positive integer, we can expand one gamma ratio using Gauss' multiplication formula:

$$\frac{\Gamma(\frac{mp}{2})}{\Gamma(\frac{mp}{2} + ph)} = \frac{(2\pi)^{\frac{1-p}{2}} p^{\frac{pm}{2} - \frac{1}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{m}{2} + \frac{1}{p}) \cdots \Gamma(\frac{m}{2} + \frac{p-1}{p})}{(2\pi)^{\frac{1-p}{2}} p^{\frac{mp}{2} - \frac{1}{2} + ph} \Gamma(\frac{m}{2} + h) \cdots \Gamma(\frac{m}{2} + \frac{p-1}{p} + h)}$$

for $p = 1, 2, \dots, m \geq p$. Accordingly,

$$\begin{aligned} E[u_3^h | H_0] &= \frac{[\Gamma(\frac{m}{2} + h)]^p}{[\Gamma(\frac{m}{2})]^p} \prod_{j=0}^{p-1} \frac{\Gamma(\frac{m}{2} + \frac{j}{p})}{\Gamma(\frac{m}{2} + \frac{j}{p} + h)} \\ &= \frac{[\Gamma(\frac{m}{2} + h)]^{p-1}}{[\Gamma(\frac{m}{2})]^{p-1}} \prod_{j=1}^{p-1} \frac{\Gamma(\frac{m}{2} + \frac{j}{p})}{\Gamma(\frac{m}{2} + \frac{j}{p} + h)} = c_{3,p-1} \frac{[\Gamma(\frac{m}{2} + h)]^{p-1}}{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} + \frac{j}{p} + h)}, \end{aligned} \quad (6.7.5)$$

$$c_{3,p-1} = \frac{\prod_{j=1}^{p-1} \Gamma(\frac{m}{2} + \frac{j}{p})}{[\Gamma(\frac{m}{2})]^{p-1}}, \Re(\frac{m}{2} + h) > 0. \quad (6.7.6)$$

Hence, for $h = s - 1$, (6.7.5) is the Mellin transform of the density of u_3 . Thus, denoting the density by $f_{u_3}(u_3)$, we have

$$f_{u_3}(u_3 | H_0) = c_{3,p-1} G_{p-1,p-1}^{p-1,0} \left[u_3 \left| \begin{matrix} \frac{m}{2} - 1 + \frac{j}{p}, j=1, \dots, p-1 \\ \frac{m}{2} - 1, \dots, \frac{m}{2} - 1 \end{matrix} \right. \right], 0 \leq u_3 \leq 1, \quad (6.7.7)$$

and zero elsewhere.

In the complex case, the h -th moment is the following:

$$E[\tilde{u}_3^h | H_o] = \tilde{c}_{3,p-1} \frac{[\tilde{\Gamma}(m+h)]^{p-1}}{\prod_{j=1}^{p-1} \tilde{\Gamma}(m + \frac{j}{p} + h)}, \quad (6.7a.1)$$

$$\tilde{c}_{3,p-1} = \frac{\prod_{j=1}^{p-1} \tilde{\Gamma}(m + \frac{j}{p})}{[\tilde{\Gamma}(m)]^{p-1}}. \quad (6.7a.2)$$

and the corresponding density is given by

$$\tilde{f}_{\tilde{u}_3}(\tilde{u}_3 | H_o) = \tilde{c}_{3,p-1} G_{p-1,p-1}^{p-1,0} \left[\tilde{u}_3 \left| \begin{matrix} m-1+\frac{j}{p}, j=1,\dots,p-1 \\ m-1,\dots,m-1 \end{matrix} \right. \right], \quad 0 \leq |\tilde{u}_3| \leq 1, \quad (6.7a.3)$$

and zero elsewhere, G denoting a real G -function.

Real and complex cases: $p = 2$

It is seen from (6.7.5) that for $p = 2$, u_3 is a real type-1 beta with the parameters ($\alpha = \frac{m}{2}$, $\beta = \frac{1}{p}$) in the real case. Whenever $p \geq 3$, poles of order 2 or more are occurring, and the resulting density functions which are expressible in terms generalized hypergeometric functions, will not be explicitly provided. For a general series expansion of the G -function, the reader may refer to Mathai (1970a, 1993).

In the complex case, when $p = 2$, \tilde{u}_3 has a real type-1 beta density with the parameters ($\alpha = m$, $\beta = \frac{1}{p}$). In this instance as well, poles of higher orders will be present when $p \geq 3$, and hence explicit forms of the corresponding densities will not be herein provided. The exact null and non-null distributions of the test statistic are derived for the general case in Mathai and Saxena (1973), and highly accurate percentage points are provided in Mathai (1979a,b).

An asymptotic result can also be obtained as $n \rightarrow \infty$. Consider the h -th moment of λ , which is available from (6.7.5) in the real case and from (6.7a.1) in the complex case. Then, referring to Corollary 6.5.2, $\delta_j = \frac{j}{p}$ whether in the real or in the complex situations. Hence, $2[\sum_{j=1}^{p-1} \delta_j] = 2 \sum_{j=1}^{p-1} \frac{j}{p} = (p-1)$ in both the real and the complex cases. As well, observe that in the complex case, the diagonal elements are real since $\tilde{\Sigma}$ is Hermitian positive definite. Accordingly, the number of restrictions imposed by H_o in either the real or complex cases is $p-1$. Thus, the following result:

Theorem 6.7.1. *Consider the λ -criterion for testing the equality of the diagonal elements, given that the covariance matrix is already diagonal. Then, as $n \rightarrow \infty$, the null distribution of $-2 \ln \lambda \rightarrow \chi_{p-1}^2$ in both the real and the complex cases.*

6.8. Hypothesis that the Covariance Matrix is Block Diagonal, Real Case

We will discuss a generalization of the problem examined in Sect. 6.6, considering again the case of real Gaussian vectors. Let X_1, \dots, X_n be iid as $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, and

$$X_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{pj} \end{bmatrix} = \begin{bmatrix} X_{(1j)} \\ \vdots \\ X_{(kj)} \end{bmatrix}, \quad X_{(1j)} = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{p_1j} \end{bmatrix}, \quad X_{(2j)} = \begin{bmatrix} x_{p_1+1,j} \\ \vdots \\ x_{p_1+p_2,j} \end{bmatrix}, \dots,$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & O & \cdots & O \\ O & \Sigma_{22} & \ddots & O \\ \vdots & \vdots & \cdots & \vdots \\ O & O & \cdots & \Sigma_{kk} \end{bmatrix}, \quad \Sigma_{jj} \text{ being } p_j \times p_j, \quad j = 1, \dots, k.$$

In this case, the $p \times 1$ real Gaussian vector is subdivided into subvectors of orders p_1, \dots, p_k , so that $p_1 + \dots + p_k = p$, and, under the null hypothesis H_0 , Σ is assumed to be a block diagonal matrix, which means that the subvectors are mutually independently distributed p_j -variate real Gaussian vectors with corresponding mean value vector $\mu_{(j)}$ and covariance matrix Σ_{jj} , $j = 1, \dots, k$. Then, the joint density of the sample values under the null hypothesis can be written as $L = \prod_{r=1}^k L_r$ where L_r is the joint density of the sample values corresponding to the subvector $X_{(rj)}$, $j = 1, \dots, n$, $r = 1, \dots, k$. Letting the $p \times n$ general sample matrix be $\mathbf{X} = (X_1, \dots, X_n)$, we note that the sample representing the first p_1 rows of \mathbf{X} corresponds to the sample from the first subvector $X_{(1j)} \stackrel{iid}{\sim} N_p(\mu_{(1)}, \Sigma_{11})$, $\Sigma_{11} > O$, $j = 1, \dots, n$. The MLE's of $\mu_{(r)}$ and Σ_{rr} are the corresponding sample mean and sample covariance matrix. Thus, the maximum of L_r is available as

$$\max L_r = \frac{e^{-\frac{np_r}{2}} n^{\frac{np_r}{2}}}{(2\pi)^{\frac{np_r}{2}} |S_{rr}|^{\frac{n}{2}}} \xrightarrow{ind} \prod_{r=1}^k \max L_r = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} \prod_{r=1}^k |S_{rr}|^{\frac{n}{2}}}.$$

Hence,

$$\lambda = \frac{\sup_{\omega} L}{\sup_{\Omega} L} = \frac{|S|^{\frac{n}{2}}}{\prod_{r=1}^k |S_{rr}|^{\frac{n}{2}}}, \quad (6.8.1)$$

and

$$u_4 \equiv \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{r=1}^k |S_{rr}|}. \quad (6.8.2)$$

Observe that the covariance matrix $\Sigma = (\sigma_{ij})$ can be written in terms of the matrix of population correlations. If we let $D = \text{diag}(\sigma_1, \dots, \sigma_p)$ where $\sigma_t^2 = \sigma_{tt}$ denotes the

variance associated the component x_{ij} in $X'_j = (x_{1j}, \dots, x_{pj})$ where $\text{Cov}(X) = \Sigma$, and $R = (\rho_{rs})$ be the population correlation matrix, where ρ_{rs} is the population correlation between the components x_{rj} and x_{sj} , then, $\Sigma = DRD$. Consider a partitioning of Σ into $k \times k$ blocks as well as the corresponding partitioning of D and R :

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \cdots & \Sigma_{kk} \end{bmatrix}, D = \begin{bmatrix} D_1 & O & \cdots & O \\ O & D_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & D_k \end{bmatrix}, R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ R_{21} & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ R_{k1} & R_{k2} & \cdots & R_{kk} \end{bmatrix},$$

where, for example, Σ_{jj} is $p_j \times p_j$, $p_1 + \dots + p_k = p$, and the corresponding partitioning of D and R . Consider a corresponding partitioning of the sample sum of products matrix $S = (S_{ij})$, $D^{(s)}$ and $R^{(s)}$ where $R^{(s)}$ is the sample correlation matrix and $D^{(s)} = \text{diag}(\sqrt{s_{11}}, \dots, \sqrt{s_{pp}})$, where S_{jj} , $D_j^{(s)}$, $R_{jj}^{(s)}$ are $p_j \times p_j$, $p_1 + \dots + p_k = p$. Then,

$$\frac{|\Sigma|}{\prod_{j=1}^k |\Sigma_{jj}|} = \frac{|R|}{\prod_{j=1}^k |R_{jj}|} \quad (6.8.3)$$

and

$$u_4 \equiv \lambda^{\frac{2}{n}} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = \frac{|R^{(s)}|}{\prod_{j=1}^k |R_{jj}^{(s)}|}. \quad (6.8.4)$$

An additional interesting property is now pointed out. Consider a linear function of the original $p \times 1$ vector $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, in the form CX_j where C is the diagonal matrix, $\text{diag}(c_1, \dots, c_p)$. In this case, the product CX_j is such that the r -th component of X_j is weighted or multiplied by c_r . Let C be a block diagonal matrix that is partitioned similarly to D so that its j -th diagonal block matrix be the $p_j \times p_j$ diagonal submatrix C_j . Then,

$$u_c = \frac{|CSC'|}{\prod_{j=1}^k |C_j S_{jj} C'_j|} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = u_4. \quad (6.8.5)$$

In other words, u_4 is invariant under linear transformations on $X_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$. That is, if $Y_j = CX_j + d$ where d is a constant column vector, then the $p \times n$ sample matrix on Y_j , namely, $\bar{Y} = (Y_1, \dots, Y_n) = (CX_1 + d, \dots, CX_n + d)$,

$$\mathbf{Y} - \bar{\mathbf{Y}} = C(\mathbf{X} - \bar{\mathbf{X}}) \Rightarrow S_y = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})' = C(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'C' = CSC'.$$

Letting S_y be partitioned as S into $k \times k$ blocks and $S_y = (S_{ijy})$, we have

$$u_y \equiv \frac{|S_y|}{\prod_{j=1}^k |S_{jyy}|} = \frac{|CSC'|}{\prod_{j=1}^k |C_j S_{jj} C'_j|} = \frac{|S|}{\prod_{j=1}^k |S_{jj}|} = u_4. \quad (6.8.6)$$

Arbitrary moments of u_4 can be derived by proceeding as in Sect. 6.6. The h -th null moment, that is, the h -th moment under the null hypothesis H_o , is then

$$E[u_4^h | H_o] = \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{m}{2}}} \int_{S>O} |S|^{\frac{m}{2}+h-\frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma_o^{-1}S)} \left\{ \prod_{r=1}^k |S_{rr}|^{-h} \right\} dS \quad (i)$$

where $m = n - 1$, n being the sample size, and

$$\Sigma_o = \begin{bmatrix} \Sigma_{11} & O & \cdots & O \\ O & \Sigma_{22} & \cdots & O \\ \vdots & \vdots & \ddots & O \\ O & O & \cdots & \Sigma_{kk} \end{bmatrix}, \quad \text{tr}(\Sigma_o^{-1}S) = \text{tr}(\Sigma_{11}^{-1}S_{11}) + \cdots + \text{tr}(\Sigma_{kk}^{-1}S_{kk}), \quad (ii)$$

where S_{rr} is the r -th diagonal block of S , corresponding to Σ_{rr} of Σ whose order $p_r \times p_r$, $r = 1, \dots, k$, $p_1 + \cdots + p_k = p$. On noting that

$$|S_{rr}|^{-h} = \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r>O} |Y_r|^{h-\frac{p_r+1}{2}} e^{-\text{tr}(Y_r S_{rr})} dY_r, \quad r = 1, \dots, k, \quad (iii)$$

where $Y_r > O$ is a $p_r \times p_r$ real positive definite matrix, and replacing each $|S_{rr}|^{-h}$ by its integral representation as given in (iii), the exponent of e in (i) becomes

$$-\frac{1}{2}[\text{tr}(\Sigma_o^{-1}S) + 2\text{tr}(YS)], \quad Y = \begin{bmatrix} Y_1 & O & \cdots & O \\ O & Y_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & Y_k \end{bmatrix}, \quad \text{tr}(YS) = \text{tr}(Y_1 S_{11}) + \cdots + \text{tr}(Y_k S_{kk}).$$

The right-hand side of equation (i) then becomes

$$E[u_4^h | H_o] = \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{m}{2}}} 2^{ph} \left\{ \prod_{r=1}^k \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r>O} |Y_r|^{h-\frac{p_r+1}{2}} \right\} \\ \times |\Sigma_o^{-1} + 2Y|^{-\frac{m}{2}+h} dY_1 \wedge \dots \wedge dY_k, \quad \Re(h) > -\frac{m}{2} + \frac{p-1}{2}. \quad (iv)$$

It should be pointed out that the non-null moments of u_4 can be obtained by substituting a general Σ to Σ_o in (iv). Note that if we replace $2Y$ by Y , the factor containing 2, namely 2^{ph} , will disappear. Further, under H_o ,

$$|\Sigma_o^{-1} + 2Y|^{-\frac{m}{2}+h} = \left\{ \prod_{r=1}^k |\Sigma_{rr}|^{\frac{m}{2}+h} \right\} \left\{ \prod_{r=1}^k |I + 2\Sigma_{rr}Y_r|^{-\frac{m}{2}+h} \right\}. \quad (v)$$

Then, each Y_r -integral can be evaluated as follows:

$$\begin{aligned} & \frac{1}{\Gamma_{p_r}(h)} \int_{Y_r > 0} |Y_r|^{h - \frac{p_r + 1}{2}} |I + 2 \Sigma_{rr} Y_r|^{-(\frac{m}{2} + h)} dY_r \\ &= 2^{-p_r h} |\Sigma_{rr}|^{-h} \frac{1}{\Gamma_{p_r}(h)} \int_{Z_r > 0} |Z_r|^{h - \frac{p_r + 1}{2}} |I + Z_r|^{-(\frac{m}{2} + h)} dZ_r, \quad Z_r = 2 \Sigma_{rr}^{\frac{1}{2}} Y_r \Sigma_{rr}^{\frac{1}{2}} \\ &= 2^{-p_r h} |\Sigma_{rr}|^{-h} \frac{\Gamma_{p_r}(h)}{\Gamma_{p_r}(h)} \frac{\Gamma_{p_r}(\frac{m}{2})}{\Gamma_{p_r}(\frac{m}{2} + h)}. \end{aligned} \tag{vi}$$

On combining equations (i) to (vi), we have

$$E[u_4^h | H_0] = \frac{\Gamma_p(\frac{m}{2} + h)}{\Gamma_p(\frac{m}{2})} \prod_{r=1}^k \frac{\Gamma_{p_r}(\frac{m}{2})}{\Gamma_{p_r}(\frac{m}{2} + h)}, \quad \Re(\frac{m}{2} + h) > \frac{p-1}{2}, \tag{6.8.7}$$

$$= c_{4,p} \frac{\Gamma_p(\frac{m}{2} + h)}{\prod_{r=1}^k \Gamma_{p_r}(\frac{m}{2} + h)} = c_{4,p} c^* \frac{\prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{r=1}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2} + h)]}, \tag{6.8.8}$$

$$c_{4,p} = \frac{\prod_{r=1}^k \Gamma_{p_r}(\frac{m}{2})}{\Gamma_p(\frac{m}{2})} = \frac{[\prod_{r=1}^k \pi^{\frac{p_r(p_r-1)}{4}}] \prod_{r=1}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2})]}{\pi^{\frac{p(p-1)}{4}} \prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2})}, \tag{6.8.9}$$

$$c^* = \frac{\pi^{\frac{p(p-1)}{4}}}{\prod_{r=1}^k \pi^{\frac{p_r(p_r-1)}{4}}}$$

so that when $h = 0$, $E[u_4^h | H_0] = 1$. Observe that one set of gamma products can be canceled in (6.8.8) and (6.8.9). When that set is the product of the first p_1 gamma functions, the h -th moment of u_4 is given by

$$E[u_4^h | H_0] = c_{4,p-p_1} \frac{\prod_{j=p_1+1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{r=2}^k [\prod_{i=1}^{p_r} \Gamma(\frac{m}{2} - \frac{i-1}{2} + h)]}, \tag{6.8.10}$$

where $c_{4,p-p_1}$ is such that $E[u_4^h | H_0] = 1$ when $h = 0$. Since the structure of the expression given in (6.8.10) is that of the h -th moment of a product of $p - p_1$ independently distributed real scalar type-1 beta random variables, it can be inferred that the distribution of $u_4 | H_0$ is also that of a product of $p - p_1$ independently distributed real scalar type-1 beta random variables whose parameters can be determined from the arguments of the gamma functions appearing in (6.8.10).

Some of the gamma functions appearing in (6.8.10) will cancel out for certain values of p_1, \dots, p_k , thereby simplifying the representation of the moments and enabling one to

express the density of u_4 in terms of elementary functions in such instances. The exact null density in the general case was derived by the first author. For interesting representations of the exact density, the reader is referred to Mathai and Rathie (1971) and Mathai and Saxena (1973), some exact percentage points of the null distribution being included in Mathai and Katiyar (1979a). As it turns out, explicit forms are available in terms of elementary functions for the following special cases, see also Anderson (2003): $p_1 = p_2 = p_3 = 1$; $p_1 = p_2 = p_3 = 2$; $p_1 = p_2 = 1, p_3 = p - 2$; $p_1 = 1, p_2 = p_3 = 2$; $p_1 = 1, p_2 = 2, p_3 = 3$; $p_1 = 2, p_2 = 2, p_3 = 4$; $p_1 = p_2 = 2, p_3 = 3$; $p_1 = 2, p_2 = 3, p$ is even.

6.8.1. Special case: $k = 2$

Let us consider a certain 2×2 partitioning of S , which corresponds to the special case $k = 2$. When $p_1 = 1$ and $p_2 = p - 1$ so that $p_1 + p_2 = p$, the test statistic is

$$\begin{aligned} u_4 &= \frac{|S|}{|S_{11}| |S_{22}|} = \frac{|S_{11} - S_{12} S_{22}^{-1} S_{21}|}{|S_{11}|} \\ &= \frac{s_{11} - S_{12} S_{22}^{-1} S_{21}}{s_{11}} = 1 - r_{1.(2\dots p)}^2 \end{aligned} \quad (6.8.11)$$

where $r_{1.(2\dots p)}$ is the multiple correlation between x_1 and (x_2, \dots, x_p) . As stated in Theorem 5.6.3, $1 - r_{1.(2\dots p)}^2$ is distributed as a real scalar type-1 beta variable with the parameters $(\frac{n-1}{2} - \frac{p-1}{2}, \frac{p-1}{2})$. The simplifications in (6.8.11) are achieved by making use of the properties of determinants of partitioned matrices, which are discussed in Sect. 1.3. Since s_{11} is 1×1 in this case, the numerator determinant is a real scalar quantity. Thus, this yields a type-2 beta distribution for $w = \frac{u_4}{1-u_4}$ and thereby $\frac{n-p}{p-1} w$ has an F -distribution, so that the test can be based on an F statistic having $(n-1) - (p-1) = n-p$ and $p-1$ degrees of freedom.

6.8.2. General case: $k = 2$

If in a 2×2 partitioning of S , S_{11} is of order $p_1 \times p_1$ and S_{22} is of order $p_2 \times p_2$ with $p_2 = p - p_1$. Then u_4 can be expressed as

$$\begin{aligned} u_4 &= \frac{|S|}{|S_{11}| |S_{22}|} = \frac{|S_{11} - S_{12} S_{22}^{-1} S_{21}|}{|S_{11}|} \\ &= |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}| = |I - U|, \quad U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} \end{aligned} \quad (6.8.12)$$

where U is called the *multiple correlation matrix*. It will be shown that U has a real matrix-variate type-1 beta distribution when S_{11} is of general order rather than being a scalar.

Theorem 6.8.1. Consider u_4 for $k = 2$. Let S_{11} be $p_1 \times p_1$ and S_{22} be $p_2 \times p_2$ so that $p_1 + p_2 = p$. Without any loss of generality, let us assume that $p_1 \leq p_2$. Then, under $H_o : \Sigma_{12} = O$, the multiple correlation matrix U has a real matrix-variate type-1 beta distribution with the parameters $(\frac{p_2}{2}, \frac{m}{2} - \frac{p_2}{2})$, with $m = n - 1$, n being sample size, and thereby $(I - U) \sim$ type-1 beta $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$, the determinant of $I - U$ being u_4 under the null hypothesis when $k = 2$.

Proof: Since Σ under H_o can readily be eliminated from a structure such as u_4 , we will take a Wishart matrix S having $m = n - 1$ degrees of freedom, n denoting the sample size, and parameter matrix I , the identity matrix. At first, assume that Σ is a block diagonal matrix and make the transformation $S_1 = \Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}}$. As a result, u_4 will be free of Σ_{11} and Σ_{22} , and so, we may take $S \sim W_p(m, I)$. Now, consider the submatrices S_{11} , S_{22} , S_{12} so that $dS = dS_{11} \wedge dS_{22} \wedge dS_{12}$. Let $f(S)$ denote the $W_p(m, I)$ density. Then,

$$f(S)dS = \frac{|S|^{\frac{m-p+1}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} e^{-\frac{1}{2}\text{tr}(S)} dS, \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{21} = S'_{12}.$$

However, appealing to a result stated in Sect. 1.3, we have

$$\begin{aligned} |S| &= |S_{22}| |S_{11} - S_{12} S_{22}^{-1} S_{21}| \\ &= |S_{22}| |S_{11}| |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}|. \end{aligned}$$

The joint density of S_{11} , S_{22} , S_{12} denoted by $f_1(S_{11}, S_{22}, S_{12})$ is then

$$\begin{aligned} f_1(S_{11}, S_{22}, S_{12}) &= |S_{11}|^{\frac{m-p+1}{2}} |S_{22}|^{\frac{m-p+1}{2}} |I - U|^{\frac{m-p+1}{2}} \\ &\quad \times \frac{e^{-\frac{1}{2}\text{tr}(S_{11}) - \frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})}, \quad U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}. \end{aligned}$$

Letting $Y = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}}$, it follows from a result on Jacobian of matrix transformation, previously established in Chap. 1, that $dY = |S_{11}|^{-\frac{p_2}{2}} |S_{22}|^{-\frac{p_1}{2}} dS_{12}$. Thus, the joint density of S_{11} , S_{22} , Y , denoted by $f_2(S_{11}, S_{22}, Y)$, is given by

$$\begin{aligned} f_2(S_{11}, S_{22}, Y) &= |S_{11}|^{\frac{m+p_2-p+1}{2}} |S_{22}|^{\frac{m+p_1-p+1}{2}} |I - YY'|^{\frac{m-p+1}{2}} \\ &\quad \times \frac{e^{-\frac{1}{2}\text{tr}(S_{11}) - \frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{pm}{2}} \Gamma_p(\frac{m}{2})}, \end{aligned}$$

Note that S_{11} , S_{22} , Y are independently distributed as $f_2(\cdot)$ can be factorized into functions of S_{11} , S_{22} , Y . Now, letting $U = YY'$, it follows from Theorem 4.2.3 that

$$dY = \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_1}(\frac{p_2}{2})} |U|^{\frac{p_2}{2} - \frac{p_1 + 1}{2}} dU,$$

and the density of U , denoted by $f_3(U)$, can then be expressed as follows:

$$f_3(U) = c |U|^{\frac{p_2}{2} - \frac{p_1 + 1}{2}} |I - U|^{\frac{m}{2} - \frac{p_2}{2} - \frac{p_1 + 1}{2}}, \quad 0 < U < I, \quad (6.8.13)$$

which is a real matrix-variate type-1 beta density with the parameters $(\frac{p_2}{2}, \frac{m}{2} - \frac{p_2}{2})$, where c is the normalizing constant. As a result, $I - U$ has a real matrix-variate type-1 beta distribution with the parameters $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$. Finally, observe that u_4 is the determinant of $I - U$.

Corollary 6.8.1. Consider u_4 as given in (6.8.12) and the determinant $|I - U|$ where U and $I - U$ are defined in Theorem 6.8.1. Then for $k = 2$ and an arbitrary h , $E[u_4^h | H_o] = |I - U|^h$.

Proof: On letting $k = 2$ in (6.8.8), we obtain the h -th moment of $u_4 | H_o$ as

$$E[u_4^h | H_o] = c_{4,p} \frac{\prod_{j=1}^p \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\{\prod_{j=1}^{p_1} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)\} \{\prod_{j=1}^{p_2} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)\}}. \quad (i)$$

After canceling p_2 of the gamma functions, the remaining gamma product in the numerator of (i) is

$$\Gamma\left(\alpha - \frac{p_2}{2}\right) \Gamma\left(\alpha - \frac{p_2 + 1}{2}\right) \cdots \Gamma\left(\alpha - \frac{p-1}{2}\right) = \Gamma_{p_1}\left(\alpha - \frac{p_2}{2}\right), \quad \alpha = \frac{m}{2} + h,$$

excluding $\pi^{\frac{p_1(p_1-1)}{4}}$. The remainder of the gamma product present in the denominator is comprised of the gamma functions coming from $\Gamma_{p_1}(\frac{m}{2} + h)$, excluding $\pi^{\frac{p_1(p_1-1)}{4}}$. The normalizing constant will automatically take care of the factors containing π . Now, the resulting part containing h is $\Gamma_{p_1}(\frac{m}{2} - \frac{p_2}{2} + h) / \Gamma_{p_1}(\frac{m}{2} + h)$, which is the gamma ratio in the h -th moment of a $p_1 \times p_1$ real matrix-variate type-1 beta distribution with the parameters $(\frac{m}{2} - \frac{p_2}{2}, \frac{p_2}{2})$.

Since this happens to be $E[|I - U|^h]$ for $I - U$ distributed as is specified in Theorem 6.8.1, the Corollary is established.

An asymptotic result can be established from Corollary 6.5.1 and the λ -criterion for testing block-diagonality or equivalently the independence of subvectors in a p -variate

Gaussian population. The resulting chisquare variable will have $2 \sum_j \delta_j$ degrees of freedom where δ_j is as defined in Corollary 6.5.1 for the second parameter of the real scalar type-1 beta distribution. Referring to (6.8.10), we have

$$\begin{aligned} \sum_j \delta_j &= \sum_{j=p_1+1}^p \frac{j-1}{2} - \sum_{j=2}^k \sum_{i=1}^{p_j} \frac{i-1}{2} = \sum_{j=p_1+1}^p \frac{j-1}{2} - \sum_{j=2}^k \frac{p_j(p_j-1)}{4} \\ &= \sum_{j=1}^p \frac{j-1}{2} - \sum_{j=1}^k \frac{p_j(p_j-1)}{4} = \frac{p(p-1)}{4} - \sum_{j=1}^k \frac{p_j(p_j-1)}{4} = \sum_{j=1}^k \frac{p_j(p-p_j)}{4}. \end{aligned}$$

Accordingly, the degrees of freedom of the resulting chisquare is $2[\sum_{j=1}^k \frac{p_j(p-p_j)}{4}] = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$ in the real case. It can also be observed that the number of restrictions imposed by the null hypothesis H_o is obtained by first letting all the off-diagonal elements of $\Sigma = \Sigma'$ equal to zero and subtracting the off-diagonal elements of the k diagonal blocks which produces $\frac{p(p-1)}{2} - \sum_{j=1}^k \frac{p_j(p_j-1)}{2} = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$. In the complex case, the number of degrees of freedom will be twice that obtained for the real case, the chisquare variable remaining a real scalar chisquare random variable. This is now stated as a theorem.

Theorem 6.8.2. Consider the λ -criterion given in (6.8.1) in the real case and let the corresponding λ in the complex case be $\tilde{\lambda}$. Then $-2 \ln \lambda \rightarrow \chi_{\delta}^2$ as $n \rightarrow \infty$ where n is the sample size and $\delta = \sum_{j=1}^k \frac{p_j(p-p_j)}{2}$, which is also the number of restrictions imposed by H_o . Analogously, in the complex case, $-2 \ln \tilde{\lambda} \rightarrow \chi_{\tilde{\delta}}^2$ as $n \rightarrow \infty$, where the chisquare variable remains a real scalar chisquare random variable, $\tilde{\delta} = \sum_{j=1}^k p_j(p-p_j)$ and n denotes the sample size.

6.9. Hypothesis that the Mean Value and Covariance Matrix are Given

Consider a real p -variate Gaussian population $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, and a simple random sample, X_1, \dots, X_n , from this population, the X_i 's being iid as X_j . Let the sample mean and the sample sum of products matrix be denoted by \bar{X} and S , respectively. Consider the hypothesis $H_o : \mu = \mu_o, \Sigma = \Sigma_o$ where μ_o and Σ_o are specified. Let us examine the likelihood ratio test for testing H_o and obtain the resulting λ -criterion. Let the parameter space be $\Omega = \{(\mu, \Sigma) | \Sigma > O, -\infty < \mu_j < \infty, j = 1, \dots, p, \mu' = (\mu_1, \dots, \mu_p)\}$. Let the joint density of X_1, \dots, X_n be denoted by L . Then, as previously obtained, the maximum value of L is

$$\max_{\Omega} L = \frac{e^{-\frac{np}{2}} n^{\frac{np}{2}}}{(2\pi)^{\frac{np}{2}} |S|^{\frac{n}{2}}} \quad (6.9.1)$$

and the maximum under H_o is

$$\max_{H_o} L = \frac{e^{-\frac{1}{2}\text{tr}(\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o))}}{(2\pi)^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}}. \quad (6.9.2)$$

Thus,

$$\lambda = \frac{\max_{H_o} L}{\max_{\Omega} L} = \frac{e^{\frac{np}{2}} |S|^{\frac{n}{2}}}{n^{\frac{np}{2}} |\Sigma_o|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o))}. \quad (6.9.3)$$

We reject H_o for small values of λ . Since the exponential part dominates the polynomial part for large values, we reject for large values of the exponent, excluding (-1) , which means for large values of $\sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \sim \chi_{np}^2$ since $(X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \stackrel{iid}{\sim} \chi_p^2$ for each j . Hence the criterion consists of

$$\text{rejecting } H_o \text{ if the observed values of } \sum_{j=1}^n (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) \geq \chi_{np, \alpha}^2$$

with

$$Pr\{\chi_{np}^2 \geq \chi_{np, \alpha}^2\} = \alpha. \quad (6.9.4)$$

Let us determine the h -th moment of λ for an arbitrary h . Note that

$$\lambda^h = \frac{e^{\frac{np}{2}}}{n^{\frac{np}{2}} |\Sigma_o|^{\frac{nh}{2}}} |S|^{\frac{nh}{2}} e^{-\frac{h}{2}\text{tr}(\Sigma_o^{-1} S) - \frac{hn}{2}(\bar{X} - \mu_o)' \Sigma_o^{-1} (\bar{X} - \mu_o)}. \quad (6.9.5)$$

Since λ contains S and \bar{X} and these quantities are independently distributed, we can integrate out the part containing S over a Wishart density having $m = n - 1$ degrees of freedom and the part containing \bar{X} over the density of \bar{X} . Thus, for $m = n - 1$,

$$\begin{aligned} E \left[(|S|^{\frac{nh}{2}} / |\Sigma_o|^{\frac{nh}{2}}) e^{-\frac{h}{2}\text{tr}(\Sigma_o^{-1} S)} | H_o \right] &= \frac{\int_{S>O} |S|^{\frac{m}{2} + \frac{nh}{2} - \frac{p+1}{2}} e^{-\frac{(1+h)}{2}\text{tr}(\Sigma_o^{-1} S)} dS}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2}) |\Sigma_o|^{\frac{n}{2}(1+h) - \frac{1}{2}}} \\ &= 2^{\frac{np}{2}} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n}{2} - \frac{1}{2})} (1+h)^{-[\frac{n}{2}(1+h) - \frac{1}{2}]p}. \quad (i) \end{aligned}$$

Under H_o , the integral over \bar{X} gives

$$\int_{\bar{X}} \frac{\sqrt{n}}{(2\pi)^{\frac{p}{2}} |\Sigma_o|^{\frac{1}{2}}} e^{-(1+h)\frac{n}{2}(\bar{X} - \mu_o)' \Sigma_o^{-1} (\bar{X} - \mu_o)} d\bar{X} = (1+h)^{-\frac{p}{2}}. \quad (ii)$$

From (i) and (ii), we have

$$E[\lambda^h | H_o] = \frac{e^{\frac{np}{2}} 2^{\frac{np}{2}}}{n^{\frac{np}{2}}} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n-1}{2})} (1+h)^{-[\frac{n}{2}(1+h)]p}. \quad (6.9.6)$$

The inversion of this expression is quite involved due to branch points. Let us examine the asymptotic case as $n \rightarrow \infty$. On expanding the gamma functions by making use of the version of Stirling's asymptotic approximation formula for gamma functions given in (6.5.14), namely $\Gamma(z + \eta) \approx \sqrt{2\pi} z^{z+\eta-\frac{1}{2}} e^{-z}$ for $|z| \rightarrow \infty$ and η bounded, we have

$$\begin{aligned} \frac{\Gamma_p(\frac{n}{2}(1+h) - \frac{1}{2})}{\Gamma_p(\frac{n-1}{2})} &= \prod_{j=1}^p \frac{\Gamma(\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n-1}{2} - \frac{j-1}{2})} \\ &= \prod_{j=1}^p \frac{\sqrt{2\pi} [\frac{n}{2}(1+h)]^{\frac{n}{2}(1+h) - \frac{1}{2} - \frac{j-1}{2}} e^{-\frac{n}{2}(1+h)}}{\sqrt{2\pi} [\frac{n}{2}]^{\frac{n}{2} - \frac{1}{2} - \frac{j-1}{2}} e^{-\frac{n}{2}}} \\ &= \left[\frac{n}{2}\right]^{\frac{np}{2}} e^{-\frac{np}{2}} (1+h)^{\frac{n}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}}. \end{aligned} \quad (iii)$$

Thus, as $n \rightarrow \infty$, it follows from (6.9.6) and (iii) that

$$E[\lambda^h | H_o] = (1+h)^{-\frac{1}{2}(p + \frac{p(p+1)}{2})}, \quad (6.9.7)$$

which implies that, asymptotically, $-2 \ln \lambda$ has a real scalar chisquare distribution with $p + \frac{p(p+1)}{2}$ degrees of freedom in the real Gaussian case. Hence the following result:

Theorem 6.9.1. *Given a $N_p(\mu, \Sigma)$, $\Sigma > O$, population, consider the hypothesis $H_o : \mu = \mu_o, \Sigma = \Sigma_o$ where μ_o and Σ_o are specified. Let λ denote the λ -criterion for testing this hypothesis. Then, in the real case, $-2 \ln \lambda \rightarrow \chi_{\delta}^2$ as $n \rightarrow \infty$ where $\delta = p + \frac{p(p+1)}{2}$ and, in the corresponding complex case, $-2 \ln \lambda \rightarrow \chi_{\delta_1}^2$ as $n \rightarrow \infty$ where $\delta_1 = 2p + p(p+1)$, the chisquare variable remaining a real scalar chisquare random variable.*

Note 6.9.1. In the real case, observe that the hypothesis $H_o : \mu = \mu_o, \Sigma = \Sigma_o$ imposes p restrictions on the μ parameters and $\frac{p(p+1)}{2}$ restrictions on the Σ parameters, for a total of $p + \frac{p(p+1)}{2}$ restrictions, which corresponds to the degrees of freedom for the asymptotic chisquare distribution in the real case. In the complex case, there are twice as many restrictions.

Example 6.9.1. Consider the real trivariate Gaussian distribution $N_3(\mu, \Sigma)$, $\Sigma > O$ and the hypothesis $H_o : \mu = \mu_o, \Sigma = \Sigma_o$ where μ_o, Σ_o and an observed sample of size 5 are as follows:

$$\mu_o = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \Sigma_o = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 3 \end{bmatrix} \Rightarrow |\Sigma_o| = 24, \text{Cof}(\Sigma_o) = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 12 \end{bmatrix},$$

$$\Sigma_o^{-1} = \frac{\text{Cof}(\Sigma_o)}{|\Sigma_o|} = \frac{1}{24} \begin{bmatrix} 8 & 0 & 0 \\ 0 & 9 & 6 \\ 0 & 6 & 12 \end{bmatrix};$$

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

Now,

$$\begin{aligned} (X_1 - \mu_o)' \Sigma_o^{-1} (X_1 - \mu_o) &= \frac{36}{24}, & (X_2 - \mu_o)' \Sigma_o^{-1} (X_2 - \mu_o) &= \frac{33}{24}, \\ (X_3 - \mu_o)' \Sigma_o^{-1} (X_3 - \mu_o) &= \frac{104}{24}, & (X_4 - \mu_o)' \Sigma_o^{-1} (X_4 - \mu_o) &= \frac{144}{24}, \\ (X_5 - \mu_o)' \Sigma_o^{-1} (X_5 - \mu_o) &= \frac{20}{24}, \end{aligned}$$

and

$$\sum_{j=1}^5 (X_j - \mu_o)' \Sigma_o^{-1} (X_j - \mu_o) = \frac{1}{24} [36 + 33 + 104 + 144 + 20] = \frac{337}{24} = 14.04.$$

Note that, in this example, $n = 5$, $p = 3$ and $np = 15$. Letting the significance level of the test be $\alpha = 0.05$, H_o is not rejected since $14.04 < \chi_{15, 0.05}^2 = 25$.

6.10. Testing Hypotheses on Linear Regression Models or Linear Hypotheses

Let the $p \times 1$ real vector X_j have an expected value μ and a covariance matrix $\Sigma > O$ for $j = 1, \dots, n$, and the X_j 's be independently distributed. Let X_j, μ, Σ be partitioned as follows where x_{1j}, μ_1 and σ_{11} are 1×1 , $\mu_{(2)}, \Sigma_{21}$ are $(p-1) \times 1$, $\Sigma_{12} = \Sigma'_{21}$ and Σ_{22} is $(p-1) \times (p-1)$:

$$X_j = \begin{bmatrix} x_{1j} \\ X_{(2)j} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_{(2)} \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \quad (i)$$

If the conditional expectation of x_{1j} , given $X_{(2)j}$ is linear in $X_{(2)j}$, then omitting the subscript j since the X_j 's are iid, it was established in Eq. (3.3.5) that

$$E[x_1|X_{(2)}] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_{(2)} - \mu_{(2)}). \quad (6.10.1)$$

When the regression is linear, the best linear predictor of x_1 in terms of $X_{(2)}$ will be of the form

$$E[x_1|X_{(2)}] - E(x_1) = \beta'(X_{(2)} - E(X_{(2)})), \quad \beta' = (\beta_2, \dots, \beta_p). \quad (6.10.2)$$

Then, by appealing to properties of the conditional expectation and conditional variance, it was shown in Chap. 3 that $\beta' = \Sigma_{12}\Sigma_{22}^{-1}$. Hypothesizing that $X_{(2)}$ is not random, or equivalently that the predictor function is a function of the preassigned values of $X_{(2)}$, amounts to testing whether $\Sigma_{12}\Sigma_{22}^{-1} = O$. Noting that $\Sigma_{22} > O$ since $\Sigma > O$, the null hypothesis thus reduces to $H_o : \Sigma_{12} = O$. If the original population X is p -variate real Gaussian, this hypothesis is then equivalent to testing the independence of x_1 and $X_{(2)}$. Actually, this has already been discussed in Sect. 6.8.2 for the case of $k = 2$, and is also tantamount to testing whether the population multiple correlation $\rho_{1.(2..k)} = 0$. Assuming that the population is Gaussian and letting of $u = \lambda^{\frac{2}{n}}$ where λ is the lambda criterion, $u \sim$ type-1 beta($\frac{n-p}{2}$, $\frac{p-1}{2}$) under the null hypothesis; this was established in Theorem 6.8.1 for $p_1 = 1$ and $p_2 = p - 1$. Then, $v = \frac{u}{1-u} \sim$ type-2 beta($\frac{n-p}{2}$, $\frac{p-1}{2}$), that is, $v \sim \frac{n-p}{p-1} F_{n-p, p-1}$ or $\frac{(p-1)}{(n-p)} \frac{u}{1-u} \sim F_{n-p, p-1}$. Hence, in order to test $H_o : \beta = O$,

$$\text{reject } H_o \text{ if } F_{n-p, p-1} \geq F_{n-p, p-1, \alpha}, \text{ with } Pr\{F_{n-p, p-1} \geq F_{n-p, p-1, \alpha}\} = \alpha \quad (6.10.3)$$

The test statistic u is of the form

$$u = \frac{|S|}{s_{11}|S_{22}|}, \quad S \sim W_p(n-1, \Sigma), \quad \Sigma > O, \quad \frac{(p-1)}{(n-p)} \frac{u}{1-u} \sim F_{n-p, p-1},$$

where the submatrices of S , s_{11} is 1×1 and S_{22} is $(p-1) \times (p-1)$. Observe that the number of parameters being restricted by the hypothesis $\Sigma_{12} = O$ is $p_1 p_2 = 1(p-1) = p-1$. Hence as $n \rightarrow \infty$, the null distribution of $-2 \ln \lambda$ is a real scalar chisquare having $p-1$ degrees of freedom. Thus, the following result:

Theorem 6.10.1. *Let the $p \times 1$ vector X_j be partitioned into the subvectors x_{1j} of order 1 and $X_{(2)j}$ of order $p-1$. Let the regression of x_{1j} on $X_{(2)j}$ be linear in $X_{(2)j}$, that is, $E[x_{1j}|X_{(2)j}] - E(x_{1j}) = \beta'(X_{(2)j} - E(X_{(2)j}))$. Consider the hypothesis $H_o : \beta = O$. Let $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, for $j = 1, \dots, n$, the X_j 's being independently distributed, and let λ be the λ -criterion for testing this hypothesis. Then, as $n \rightarrow \infty$, $-2 \ln \lambda \rightarrow \chi_{p-1}^2$.*

Example 6.10.1. Let the population be $N_3(\mu, \Sigma)$, $\Sigma > O$, and the observed sample of size $n = 5$ be

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, X_4 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, X_5 = \begin{bmatrix} -1 \\ -1 \\ -2 \end{bmatrix}.$$

The resulting sample average \bar{X} and deviation vectors are then

$$\begin{aligned} \bar{X} &= \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, X_1 - \bar{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 3 \end{bmatrix}, \\ X_2 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix}, X_3 - \bar{X} = \frac{1}{5} \begin{bmatrix} -6 \\ 3 \\ -2 \end{bmatrix}, \\ X_4 - \bar{X} &= \frac{1}{5} \begin{bmatrix} 4 \\ 3 \\ 8 \end{bmatrix}, X_5 - \bar{X} = \frac{1}{5} \begin{bmatrix} -6 \\ -7 \\ -12 \end{bmatrix}. \end{aligned}$$

Letting

$$\mathbf{X} = [X_1, \dots, X_5], \bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}] \text{ and } S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})',$$

$$\mathbf{X} - \bar{\mathbf{X}} = \frac{1}{5} \begin{bmatrix} 4 & 4 & -6 & 4 & -6 \\ 3 & -2 & 3 & 3 & -7 \\ 3 & 3 & -2 & 8 & -12 \end{bmatrix},$$

$$S = \frac{1}{5^2} \begin{bmatrix} 120 & 40 & 140 \\ 40 & 80 & 105 \\ 140 & 105 & 230 \end{bmatrix} = \begin{bmatrix} s_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

$$s_{11} = \frac{120}{25}, S_{12} = \frac{1}{25}[40, 140],$$

$$S_{22} = \frac{1}{25} \begin{bmatrix} 80 & 105 \\ 105 & 230 \end{bmatrix}, S_{22}^{-1} = \frac{25}{7375} \begin{bmatrix} 230 & -105 \\ -105 & 80 \end{bmatrix},$$

$$S_{12}S_{22}^{-1}S_{21} = \left[\frac{1}{(25)^2} \right] \left[\frac{25}{7375} \right] [40 \quad 140] \begin{bmatrix} 230 & -105 \\ -105 & 80 \end{bmatrix} \begin{bmatrix} 40 \\ 140 \end{bmatrix} = \frac{760000}{(25)(7375)},$$

so that the test statistic is

$$\begin{aligned} u &= \frac{s_{11} - S_{12}S_{22}^{-1}S_{21}}{s_{11}} = 1 - \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}} = 1 - r_{1,(2,3)}^2 \\ &= 1 - \frac{760000}{(120)(7375)} = 1 - 0.859 = 0.141 \Rightarrow \\ v &= \frac{(p-1)u}{(n-p)(1-u)} = \left(\frac{2}{2}\right)\left(\frac{0.141}{0.859}\right) = 0.164, \quad v \sim F_{n-p, p-1}. \end{aligned}$$

Let us test H_o at the significance level $\alpha = 0.05$. In that case, the critical value which is available from F tables is $F_{n-p, p-1, \alpha} = F_{2, 2, 0.05} = 19$. Since the observed value of v is $0.164 < 19$, H_o is not rejected.

Note 6.10.1. Observe that

$$u = 1 - r_{1,(2,3)}^2, \quad r_{1,(2,3)}^2 = \frac{S_{12}S_{22}^{-1}S_{21}}{s_{11}}$$

where $r_{1,(2,3)}$ is the sample multiple correlation between the first component of $X_j \sim N_3(\mu, \Sigma)$, $\Sigma > O$, and the other two components of X_j . If the population covariance matrix Σ is similarly partitioned, that is,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \text{where } \sigma_{11} \text{ is } 1 \times 1, \quad \Sigma_{22} \text{ is } (p-1) \times (p-1),$$

then, the population multiple correlation coefficient is $\rho_{1,(2,3)}$ where

$$\rho_{1,(2,3)}^2 = \frac{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}{\sigma_{11}}.$$

Thus, if $\Sigma_{12} = \Sigma_{21}' = O$, $\rho_{1,(2,3)} = 0$ and conversely since $\sigma_{11} > 0$ and $\Sigma_{22} > O \Rightarrow \Sigma_{22}^{-1} > O$. The regression coefficient β being equal to the transpose of $\Sigma_{12}\Sigma_{22}^{-1}$, $\Sigma_{12} = O$ also implies that the regression coefficient $\beta = O$ and conversely. Accordingly, the hypothesis that the regression coefficient vector $\beta = O$ is equivalent to hypothesizing that the population multiple correlation $\rho_{1,(2,\dots,p)} = 0$, which also implies the hypothesis that the two subvectors are independently distributed in the multivariate normal case, or that the covariance matrix $\Sigma_{12} = O$, the only difference being that the test on regression coefficients is in the conditional space whereas testing the independence of the subvectors or whether the population multiple correlation equals zero is carried out in the entire space. The numerical example included in this section also illustrates the main result presented in Sect. 6.8.1 in connection with testing whether a population multiple correlation coefficient is equal to zero.

6.10.1. A simple linear model

Consider a linear model of the following form where a real scalar variable y is estimated by a linear function of pre-assigned real scalar variables z_1, \dots, z_q :

$$y_j = \beta_o + \beta_1 z_{1j} + \dots + \beta_q z_{qj} + e_j, \quad j = 1, \dots, n \quad (i)$$

where y_1, \dots, y_n are n observations on y , $z_{i1}, z_{i2}, \dots, z_{in}$, $i = 1, \dots, q$, are preassigned values on z_1, \dots, z_q , and $\beta_o, \beta_1, \dots, \beta_q$ are unknown parameters. The random components e_j , $j = 1, \dots, n$, are the corresponding sum total contributions coming from all unknown factors. There are two possibilities with respect to this model: $\beta_o = 0$ or $\beta_o \neq 0$. If $\beta_o = 0$, β_o is omitted in model (i) and we let $y_j = x_j$. If $\beta_o \neq 0$, the model is modified by taking $x_j = y_j - \bar{y}$ where $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n)$, then becoming

$$x_j = y_j - \bar{y} = \beta_1(z_{1j} - \bar{z}_1) + \dots + \beta_q(z_{qj} - \bar{z}_q) + \epsilon_j \quad (ii)$$

for some error term ϵ_j , where $\bar{z}_i = \frac{1}{n}(z_{i1} + \dots + z_{in})$, $i = 1, \dots, q$. Letting $Z'_j = (z_{1j}, \dots, z_{qj})$ if $\beta_o = 0$ and $Z'_j = (z_{1j} - \bar{z}_1, \dots, z_{qj} - \bar{z}_q)$, otherwise, equation (ii) can be written in vector/matrix notation as follows:

$$\begin{aligned} \epsilon_j &= (x_j - \beta_1 z_{1j} - \dots - \beta_q z_{qj}) = (x_j - \beta' Z_j), \quad \beta' = (\beta_1, \dots, \beta_q), \\ \sum_{j=1}^n \epsilon_j^2 &= \sum_{j=1}^n (x_j - \beta_1 z_{1j} - \dots - \beta_q z_{qj})^2 = \sum_{j=1}^n (x_j - \beta' Z_j)^2. \end{aligned}$$

Letting

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad \text{and} \quad Z = \begin{bmatrix} z_{11} & \dots & z_{1n} \\ z_{21} & \dots & z_{2n} \\ \vdots & \ddots & \vdots \\ z_{q1} & \dots & z_{qn} \end{bmatrix}, \quad (iii)$$

$$\epsilon' = (X' - \beta' Z) \Rightarrow \sum_{j=1}^n \epsilon_j^2 = \epsilon' \epsilon = (X' - \beta' Z)(X - Z' \beta). \quad (iv)$$

The least squares minimum is thus available by differentiating $\epsilon' \epsilon$ with respect to β , equating the resulting expression to a null vector and solving, which will produce a single critical point that corresponds to the minimum as the maximum occurs at $+\infty$:

$$\begin{aligned} \frac{\partial}{\partial \beta} \left(\sum_{j=1}^n \epsilon_j^2 \right) = 0 &\Rightarrow \sum_{j=1}^n x_j Z'_j - \hat{\beta}' \sum_{j=1}^n Z_j Z'_j \\ &\Rightarrow \hat{\beta} = \left(\sum_{j=1}^n Z_j Z'_j \right)^{-1} \left(\sum_{j=1}^n x_j Z'_j \right); \end{aligned} \quad (v)$$

$$\frac{\partial}{\partial \beta} (\epsilon' \epsilon) = 0 \Rightarrow \hat{\beta} = (ZZ')^{-1} ZX, \quad (vi)$$

that is,

$$\hat{\beta} = \left(\sum_{j=1}^n Z_j Z'_j \right)^{-1} \left(\sum_{j=1}^n x_j Z'_j \right). \quad (6.10.4)$$

Since the z_{ij} 's are preassigned quantities, it can be assumed without any loss of generality that (ZZ') is nonsingular, and thereby that $(ZZ')^{-1}$ exists, so that the least squares minimum, usually denoted by s^2 , is available by substituting $\hat{\beta}$ for β in $\epsilon' \epsilon$. Then, at $\beta = \hat{\beta}$,

$$\begin{aligned} \epsilon' |_{\beta=\hat{\beta}} &= X' - X' Z' (ZZ')^{-1} Z = X' [I - Z' (ZZ')^{-1} Z] \text{ so that} \\ s^2 = \epsilon' \epsilon |_{\beta=\hat{\beta}} &= X' [I - Z' (ZZ')^{-1} Z] X \end{aligned} \quad (6.10.5)$$

where $I - Z' (ZZ')^{-1} Z$ is idempotent and of rank $(n - 1) - q$. Observe that if $\beta_o \neq 0$ in (i) and we had proceeded without eliminating β_o , then β would have been of order $(k + 1) \times 1$ and $I - Z' (ZZ')^{-1} Z$, of rank $n - (q + 1) = n - 1 - q$, whereas if $\beta_o \neq 0$ and we had eliminated β_o from the model, then the rank of $I - Z' (ZZ')^{-1} Z$ would have been $(n - 1) - q$, that is, unchanged, since $\sum_{j=1}^n (x_j - \bar{x})^2 = X' [I - \frac{1}{n} J J'] X$, $J' = (1, \dots, 1)$ and the rank of $I - \frac{1}{n} J J' = n - 1$.

Some distributional assumptions on ϵ_j are required in order to test hypotheses on β . Let $\epsilon_j \sim N_1(0, \sigma^2)$, $\sigma^2 > 0$, $j = 1, \dots, n$, be independently distributed. Then $x_j \sim N_1(\beta' Z_j, \sigma^2)$, $j = 1, \dots, n$ are independently distributed but not identically distributed as the mean value depends on j . Under the normality assumption for the ϵ_j 's, it can readily be seen that the least squares estimators of β and σ^2 coincide with the maximum likelihood estimators. It can also be observed that σ^2 is estimated by $\frac{s^2}{n}$ where n is the sample size. In this simple linear regression context, the parameter space $\Omega = \{(\beta, \sigma^2) | \sigma^2 > 0\}$. Thus, under the normality assumption, the maximum of the likelihood function L is given by

$$\max_{\Omega} L = \frac{e^{-\frac{n}{2} \frac{s^2}{\sigma^2}}}{(2\pi)^{\frac{n}{2}} [s^2]^{\frac{n}{2}}}. \quad (vi)$$

Under the hypothesis $H_o : \beta = O$ or $\beta_1 = 0 = \dots = \beta_q$, the least squares minimum, usually denoted as s_o^2 , is $X'X$ and, assuming normality, the maximum of the likelihood function under H_o is the following:

$$\max_{H_o} L = \frac{e^{-\frac{n}{2}} n^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} [s_o^2]^{\frac{n}{2}}}. \quad (\text{vii})$$

Thus, the λ -criterion is

$$\begin{aligned} \lambda &= \frac{\max_{H_o} L}{\max_{\Omega} L} = \left[\frac{s^2}{s_o^2} \right]^{\frac{n}{2}} \\ \Rightarrow u &= \lambda^{\frac{2}{n}} = \frac{X'[I - Z'(ZZ')^{-1}Z]X}{X'X} \\ &= \frac{X'[I - Z'(ZZ')^{-1}Z]X}{X'[I - Z'(ZZ')^{-1}Z]X + X'Z'(ZZ')^{-1}ZX} \\ &= \frac{1}{1 + u_1} \end{aligned} \quad (6.10.6)$$

where

$$u_1 = \frac{X'Z'(ZZ')^{-1}ZX}{X'[I - Z'(ZZ')^{-1}Z]X} = \frac{s_o^2 - s^2}{s^2} \quad (\text{viii})$$

with the matrices $Z'(ZZ')^{-1}Z$ and $I - Z'(ZZ')^{-1}Z$ being idempotent, mutually orthogonal, and of ranks q and $(n - 1) - q$, respectively. We can interpret $s_o^2 - s^2$ as the sum of squares due to the hypothesis and s^2 as the residual part. Under the normality assumption, $s_o^2 - s^2$ and s^2 are independently distributed in light of independence of quadratic forms that was discussed in Sect. 3.4.1; moreover, their representations as quadratic forms in idempotent matrices of ranks q and $(n - 1) - q$ implies that $\frac{s_o^2 - s^2}{s^2} \sim \chi_q^2$ and $\frac{s^2}{s^2} \sim \chi_{(n-1)-q}^2$. Accordingly, under the null hypothesis,

$$u_2 = \frac{(s_o^2 - s^2)/q}{s^2/[(n - 1) - q]} \sim F_{q, n-1-q}, \quad (6.10.7)$$

that is, an F -statistic having q and $n - 1 - q$ degrees of freedom. Thus, we reject H_o for small values of λ or equivalently for large values of u_2 or large values of $F_{q, n-1-q}$. Hence, the following criterion:

$$\text{Reject } H_o \text{ if the observed value of } u_2 \geq F_{q, n-1-q, \alpha}, \quad Pr\{F_{q, n-1-q} \geq F_{q, n-1-q, \alpha}\} = \alpha. \quad (6.10.8)$$

A detailed discussion of the real scalar variable case is provided in Mathai and Haubold (2017).

6.10.2. Hypotheses on individual parameters

Denoting the expected value of (\cdot) by $E[(\cdot)]$, it follows from (6.10.4) that

$$\begin{aligned} E[\hat{\beta}] &= (ZZ')^{-1}ZE(X) = (ZZ')^{-1}ZZ'\beta = \beta, \quad E[X] = Z'\beta, \quad \text{and} \\ \text{Cov}(\hat{\beta}) &= (ZZ')^{-1}Z[\text{Cov}(X)]Z'(ZZ')^{-1} = (ZZ')^{-1}Z(\sigma^2I)Z'(ZZ')^{-1} \\ &= \sigma^2(ZZ')^{-1}. \end{aligned}$$

Under the normality assumption on x_j , we have $\hat{\beta} \sim N_q(\beta, \sigma^2(ZZ')^{-1})$. Letting the (r, r) -th diagonal element of $(ZZ')^{-1}$ be b_{rr} , then $\hat{\beta}_r$, the estimator of the r -th component of the parameter vector β , is distributed as $\hat{\beta}_r \sim N_1(\beta_r, \sigma^2b_{rr})$, so that

$$\frac{\hat{\beta}_r - \beta_r}{\hat{\sigma} \sqrt{b_{rr}}} \sim t_{n-1-q} \quad (6.10.9)$$

where t_{n-1-q} denotes a Student- t distribution having $n - 1 - q$ degrees of freedom and $\hat{\sigma}^2 = \frac{s^2}{n-1-q}$ is an unbiased estimator for σ^2 . On writing s^2 in terms of ϵ , it is easily seen that $E[s^2] = (n - 1 - q)\sigma^2$ where s^2 is the least squares minimum in the entire parameter space Ω . Thus, one can test hypotheses on β_r and construct confidence intervals for that parameter by means of the Student- t statistic specified in (6.10.9) or its square t_{n-1-q}^2 which has an F distribution having 1 and $n - 1 - q$ degrees of freedom, that is, $t_{n-1-q}^2 \sim F_{1, n-1-q}$.

Example 6.10.2. Let us consider a linear model of the following form:

$$y_j = \beta_o + \beta_1 z_{1j} + \cdots + \beta_q z_{qj} + e_j, \quad j = 1, \dots, n,$$

where the z_{ij} 's are preassigned numbers of the variable z_i . Let us take $n = 5$ and $q = 2$, so that the sample is of size 5 and, excluding β_o , the model has two parameters. Let the observations on y and the preassigned values on the z_i 's be the following:

$$\begin{aligned} 1 &= \beta_o + \beta_1(0) + \beta_2(1) + e_1 \\ 2 &= \beta_o + \beta_1(1) + \beta_2(-1) + e_2 \\ 4 &= \beta_o + \beta_1(-1) + \beta_2(2) + e_3 \\ 6 &= \beta_o + \beta_1(2) + \beta_2(-2) + e_4 \\ 7 &= \beta_o + \beta_1(-2) + \beta_2(5) + e_5. \end{aligned}$$

The averages on y , z_1 and z_2 are then

$$\bar{y} = \frac{1}{5}[1 + 2 + 4 + 6 + 7] = 4, \quad \bar{z}_1 = \frac{1}{5}[0 + 1 + (-1) + 2 + (-2)] = 0 \quad \text{and}$$

$$\bar{z}_2 = \frac{1}{5}[1 + (-1) + 2 + (-2) + 5] = 1,$$

and, in terms of deviations, the model becomes

$$x_j = y_j - \bar{y} = \beta_1(z_{1j} - \bar{z}_1) + \beta_2(z_{2j} - \bar{z}_2) + \epsilon_j, \quad \epsilon_j = e_j - \bar{e}.$$

That is,

$$\begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ -1 & 1 \\ 2 & -3 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \Rightarrow X = Z'\beta + \epsilon.$$

When minimizing $\epsilon'\epsilon = (X - Z'\beta)'(X - Z'\beta)$, we determined that $\hat{\beta}$, the least squares estimate of β , the least squares minimum s^2 and $s_o^2 - s^2 =$ corresponding to the sum of squares due to β , could be express as

$$\hat{\beta} = (ZZ')^{-1}ZX, \quad s_o^2 - s^2 = X'Z'(ZZ')^{-1}ZX,$$

$$s^2 = X'[I - Z'(ZZ')^{-1}Z]X \quad \text{and that}$$

$$Z'(ZZ')^{-1}Z = [Z'(ZZ')^{-1}Z]^2, \quad I - Z'(ZZ')^{-1}Z = [I - Z'(ZZ')^{-1}Z]^2,$$

$$[I - Z'(ZZ')^{-1}Z][Z'(ZZ')^{-1}Z] = O.$$

Let us evaluate those quantities:

$$Z = \begin{bmatrix} 0 & 1 & -1 & 2 & -2 \\ 0 & -2 & 1 & -3 & 4 \end{bmatrix}, \quad ZZ' = \begin{bmatrix} 10 & -17 \\ -17 & 30 \end{bmatrix},$$

$$|ZZ'| = 11, \quad \text{Cof}(ZZ') = \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix}, \quad (ZZ')^{-1} = \frac{1}{11} \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix},$$

$$(ZZ')^{-1}Z = \frac{1}{11} \begin{bmatrix} 30 & 17 \\ 17 & 10 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 & 2 & -2 \\ 0 & -2 & 1 & -3 & 4 \end{bmatrix}$$

$$= \frac{1}{11} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix};$$

$$\begin{aligned}\hat{\beta} &= (ZZ')^{-1}ZX = \frac{1}{11} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} 50 \\ 32 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix};\end{aligned}$$

$$\begin{aligned}Z'(ZZ')^{-1}Z &= \frac{1}{11} \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ -1 & 1 \\ 2 & -3 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 0 & -4 & -13 & 9 & 8 \\ 0 & -3 & -7 & 4 & 6 \end{bmatrix} \\ &= \frac{1}{11} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & -4 \\ 0 & 1 & 6 & -5 & -2 \\ 0 & 1 & -5 & 6 & -2 \\ 0 & -4 & -2 & -2 & 8 \end{bmatrix}.\end{aligned}$$

Then,

$$\begin{aligned}X'Z'(ZZ')^{-1}ZX &= \frac{1}{11} \begin{bmatrix} -3 & -2 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & -4 \\ 0 & 1 & 6 & -5 & -2 \\ 0 & 1 & -5 & 6 & -2 \\ 0 & -4 & -2 & -2 & 8 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 0 \\ 2 \\ 3 \end{bmatrix} \\ &= \frac{120}{11}, \quad X'X = 26,\end{aligned}$$

$$X'[I - Z'(ZZ')^{-1}Z]X = 26 - \frac{120}{11} = \frac{166}{11}, \quad \text{with } n = 5 \text{ and } q = 2.$$

The test statistics u_1 and u_2 and their observed values are the following:

$$\begin{aligned} u_1 &= \frac{X'Z'(ZZ')^{-1}ZX}{X'[I - Z'(ZZ')^{-1}Z]X} = \frac{s_o^2 - s^2}{s^2} \\ &= \frac{120}{166} = 0.72; \\ u_2 &= \frac{(s_o^2 - s^2)/q}{s^2/(n - 1 - q)} \sim F_{q, n-1-q} \\ &= \frac{120/2}{166/2} = 0.72. \end{aligned}$$

Letting the significance level be $\alpha = 0.05$, the required tabulated critical value is $F_{q, n-1-q, \alpha} = F_{2, 2, 0.05} = 19$. Since $0.72 < 19$, the hypothesis $H_o : \beta = O$ is not rejected. Thus, we will not proceed to test individual hypotheses on the regression coefficients β_1 and β_2 . For tests on general linear models, refer for instance to Mathai (1971).

6.11. Problem Involving Two or More Independent Gaussian Populations

Consider k independent p -variate real normal populations $X_j \sim N_p(\mu_{(j)}, \Sigma)$, $\Sigma > O$, $j = 1, \dots, k$, having the same nonsingular covariance matrix Σ but possibly different mean values. We consider the problem of testing hypotheses on linear functions of the mean values. Let $b = a_1\mu_{(1)} + \dots + a_k\mu_{(k)}$ where a_1, \dots, a_k , are real scalar constants, and the null hypothesis be $H_o : b = b_o$ (given), which means that the a_j 's and $\mu_{(j)}$'s, $j = 1, \dots, k$, are all specified. It is also assumed that Σ is known. Suppose that simple random samples of sizes n_1, \dots, n_k from these k independent normal populations can be secured, and let the sample values be X_{jq} , $q = 1, \dots, n_j$, where X_{j1}, \dots, X_{jn_j} are iid as $N_p(\mu_{(j)}, \Sigma)$, $\Sigma > O$. Let the sample averages be denoted by $\bar{X}_j = \frac{1}{n_j} \sum_{q=1}^{n_j} X_{jq}$, $j = 1, \dots, k$. Consider the test statistic $U_k = a_1\bar{X}_1 + \dots + a_k\bar{X}_k$. Since the populations are independent and U_k is a linear function of independent vector normal variables, U_k is normally distributed with the mean value $b = a_1\mu_{(1)} + \dots + a_k\mu_{(k)}$ and covariance matrix $\frac{1}{n}\Sigma$, where $\frac{1}{n} = (\frac{a_1^2}{n_1} + \dots + \frac{a_k^2}{n_k})$ and so, $\sqrt{n}\Sigma^{-\frac{1}{2}}(U_k - b) \sim N_p(O, I)$. Then, under the hypothesis $H_o : b = b_o$ (given), which is being tested against the alternative $H_1 : b \neq b_o$, the test criterion is obtained by proceeding as was done in the single population case. Thus, the test statistic is $z = n(U_k - b_o)' \Sigma^{-1}(U_k - b_o) \sim \chi_p^2$ and the criterion will be to reject the null hypothesis for large values of the z . Accordingly, the criterion is

$$\begin{aligned} \text{Reject } H_o : b = b_o \text{ if the observed value of } n(U_k - b_o)' \Sigma^{-1}(U_k - b_o) &\geq \chi_{p, \alpha}^2, \\ \text{with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} &= \alpha. \end{aligned} \quad (6.11.1)$$

In particular, suppose that we wish to test the hypothesis $H_o : \delta = \mu_{(1)} - \mu_{(2)} = \delta_o$, such as $\delta_o = 0$ as is often the case, against the natural alternative. In this case, when $\delta_o = 0$, the null hypothesis is that the mean value vectors are equal, that is, $\mu_{(1)} = \mu_{(2)}$, and the test statistic is $z = n(\bar{X}_1 - \bar{X}_2)' \Sigma^{-1} (\bar{X}_1 - \bar{X}_2) \sim \chi_p^2$ with $\frac{1}{n} = \frac{1}{n_1} + \frac{1}{n_2}$, the test criterion being

$$\begin{aligned} \text{Reject } H_o : \mu_{(1)} - \mu_{(2)} = 0 \text{ if the observed value of } z \geq \chi_{p, \alpha}^2, \\ \text{with } Pr\{\chi_p^2 \geq \chi_{p, \alpha}^2\} = \alpha. \end{aligned} \quad (6.11.2)$$

For a numerical example, the reader is referred to Example 6.2.3. One can also determine the power of the test or the probability of rejecting $H_o : \delta = \mu_{(1)} - \mu_{(2)} = \delta_o$, under an alternative hypothesis, in which case the distribution of z is a noncentral chisquare variable with p degrees of freedom and non-centrality parameter $\lambda = \frac{1}{2} \frac{n_1 n_2}{n_1 + n_2} (\delta - \delta_o)' \Sigma^{-1} (\delta - \delta_o)$, $\delta = \mu_{(1)} - \mu_{(2)}$, where n_1 and n_2 are the sample sizes. Under the null hypothesis, the non-centrality parameter λ is equal to zero. The power is given by

$$\text{Power} = Pr\{\text{reject } H_o | H_1\} = Pr\{\chi_p^2(\lambda) \geq \chi_{p, \alpha}^2(\lambda)\}. \quad (6.11.3)$$

When the population covariance matrices are identical and the common covariance matrix is unknown, one can also construct a statistic for testing hypotheses on linear functions of the mean value vectors by making use of steps parallel to those employed in the single population case, with the resulting criterion being based on Hotelling's T^2 statistic for testing $H_o : \mu_{(1)} = \mu_{(2)}$.

6.11.1. Equal but unknown covariance matrices

Let us consider the same procedure as in Sect. 6.11 to test a hypothesis on a linear function $b = a_1 \mu_{(1)} + \cdots + a_k \mu_{(k)}$ where a_1, \dots, a_k are known real scalar constants and $\mu_{(j)}$, $j = 1, \dots, k$, are the population mean values. We wish to test the hypothesis $H_o : b = b_o$ (given) in the sense all the mean values $\mu_{(j)}$, $j = 1, \dots, k$ and a_1, \dots, a_k , are specified. Let $U_k = a_1 \bar{X}_1 + \cdots + a_k \bar{X}_k$ as previously defined. Then, $E[U_k] = b$ and $\text{Cov}(U_k) = (\frac{a_1^2}{n_1} + \cdots + \frac{a_k^2}{n_k}) \Sigma$, where $(\frac{a_1^2}{n_1} + \cdots + \frac{a_k^2}{n_k}) \equiv \frac{1}{n}$ for some symbol n . The common covariance matrix Σ has the MLE $\frac{1}{n_1 + \cdots + n_k} (S_1 + \cdots + S_k)$ where S_j is the sample sum of products matrix for the j -th Gaussian population. It has been established that $S = S_1 + \cdots + S_k$ has a Wishart distribution with $(n_1 - 1) + \cdots + (n_k - 1) = N - k$, $N = n_1 + \cdots + n_k$, degrees of freedom, that is,

$$S \sim W_p(N - k, \Sigma), \quad \Sigma > O, \quad N = n_1 + \cdots + n_k. \quad (6.11.4)$$

Then, when Σ is unknown, it follows from a derivation parallel to that provided in Sect. 6.3 for the single population case that

$$w \equiv n(U_k - b)'S^{-1}(U_k - b) \sim \text{type-2 beta} \left(\frac{p}{2}, \frac{N - k - p}{2} \right), \quad (6.11.5)$$

or, w has a real scalar type-2 beta distribution with the parameters $(\frac{p}{2}, \frac{N-k-p}{2})$. Letting $w = \frac{p}{N-k-p}F$, this F is an F -statistic with p and $N - k - p$ degrees of freedom.

Theorem 6.11.1. *Let U_k , n , N , b , S be as defined above. Then $w = n(U_k - b)'S^{-1}(U_k - b)$ has a real scalar type-2 beta distribution with the parameters $(\frac{p}{2}, \frac{N-k-p}{2})$. Letting $w = \frac{p}{N-k-p}F$, this F is an F -statistic with p and $N - k - p$ degrees of freedom.*

Hence for testing the hypothesis $H_o : b = b_o$ (given), the criterion is the following:

$$\text{Reject } H_o \text{ if the observed value of } F = \frac{N - k - p}{p}w \geq F_{p, N-k-p, \alpha}, \quad (6.11.6)$$

$$\text{with } Pr\{F_{p, N-k-p} \geq F_{p, N-k-p, \alpha}\} = \alpha. \quad (6.11.7)$$

Note that by exploiting the connection between type-1 and type-2 real scalar beta random variables, one can obtain a number of properties on this F -statistic.

This situation has already been covered in Theorem 6.3.4 for the case $k = 2$.

6.12. Equality of Covariance Matrices in Independent Gaussian Populations

Let $X_j \sim N_p(\mu_{(j)}, \Sigma_j)$, $\Sigma_j > O$, $j = 1, \dots, k$, be independently distributed real p -variate Gaussian populations. Consider simple random samples of sizes n_1, \dots, n_k from these k populations, whose sample values, denoted by X_{jq} , $q = 1, \dots, n_j$, are iid as X_{j1} , $j = 1, \dots, k$. The sample sums of products matrices denoted by S_1, \dots, S_k , respectively, are independently distributed as Wishart matrix random variables with $n_j - 1$, $j = 1, \dots, k$, degrees of freedoms. The joint density of all the sample values is then given by

$$L = \prod_{j=1}^k L_j, \quad L_j = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma_j^{-1}S_j) - \frac{n_j}{2}(\bar{X}_j - \mu_{(j)})' \Sigma_j^{-1}(\bar{X}_j - \mu_{(j)})}}{(2\pi)^{\frac{n_j p}{2}} |\Sigma_j|^{\frac{n_j}{2}}}, \quad (6.12.1)$$

the MLE's of $\mu_{(j)}$ and Σ_j being $\hat{\mu}_{(j)} = \bar{X}_j$ and $\hat{\Sigma}_j = \frac{1}{n_j}S_j$. The maximum of L in the entire parameter space Ω is

$$\max_{\Omega} L = \prod_{j=1}^k \max_{\Omega} L_j = \frac{\left[\prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right] e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \prod_{j=1}^k |S_j|^{\frac{n_j}{2}}}, \quad N = n_1 + \dots + n_k. \quad (i)$$

Let us test the hypothesis of equality of covariance matrices:

$$H_o : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k = \Sigma$$

where Σ is unknown. Under this null hypothesis, the MLE of $\mu_{(j)}$ is \bar{X}_j and the MLE of the common Σ is $\frac{1}{N}(S_1 + \dots + S_k) = \frac{1}{N}S$, $N = n_1 + \dots + n_k$, $S = S_1 + \dots + S_k$. Thus, the maximum of L under H_o is

$$\max_{H_o} L = \frac{N^{\frac{Np}{2}} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \prod_{j=1}^k |S|^{\frac{n_j}{2}}} = \frac{N^{\frac{Np}{2}} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} |S|^{\frac{N}{2}}}, \quad (ii)$$

and the λ -criterion is the following:

$$\lambda = \frac{N^{\frac{Np}{2}} \left\{ \prod_{j=1}^k |S_j|^{\frac{n_j}{2}} \right\}}{|S|^{\frac{N}{2}} \left\{ \prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right\}}. \quad (6.12.2)$$

Let us consider the h -th moment of λ for an arbitrary h . Letting $c = \frac{N^{\frac{Np}{2}}}{\left\{ \prod_{j=1}^k n_j^{\frac{n_j p}{2}} \right\}}$,

$$\lambda^h = c^h \frac{\left\{ \prod_{j=1}^k |S_j|^{\frac{n_j h}{2}} \right\}}{|S|^{\frac{Nh}{2}}} = c^h \left\{ \prod_{j=1}^k |S_j|^{\frac{n_j h}{2}} \right\} |S|^{-\frac{Nh}{2}}. \quad (6.12.3)$$

The factor causing a difficulty, namely $|S|^{-\frac{Nh}{2}}$, will be replaced by an equivalent integral. Letting $Y > O$ be a real $p \times p$ positive definite matrix, we have the identity

$$|S|^{-\frac{Nh}{2}} = \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{Y>O} |Y|^{\frac{Nh}{2} - \frac{p+1}{2}} e^{-\text{tr}(YS)} dY, \quad \Re\left(\frac{Nh}{2}\right) > \frac{p-1}{2}, \quad S > O, \quad (6.12.4)$$

where

$$\text{tr}(YS) = \text{tr}(YS_1) + \dots + \text{tr}(YS_k). \quad (iii)$$

Thus, once (6.12.4) is substituted in (6.12.3), λ^h splits into products involving S_j , $j = 1, \dots, k$; this enables one to integrate out over the densities of S_j , which are Wishart densities with $m_j = n_j - 1$ degrees of freedom. Noting that the exponent involving S_j is $-\frac{1}{2}\text{tr}(\Sigma_j^{-1}S_j) - \text{tr}(YS_j) = -\frac{1}{2}\text{tr}[S_j(\Sigma_j^{-1} + 2Y)]$, the integral over the Wishart density of S_j gives the following:

$$\begin{aligned}
 & \prod_{j=1}^k \frac{1}{2^{\frac{m_j p}{2}} \Gamma_p(\frac{m_j}{2}) |\Sigma_j|^{\frac{m_j}{2}}} \int_{S_j > O} |S_j|^{\frac{m_j}{2} + \frac{n_j h}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}[S_j(\Sigma^{-1} + 2Y)]} dS_j \\
 &= \prod_{j=1}^k \frac{2^{\frac{n_j h p}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2}) |\Sigma_j^{-1} + 2Y|^{-\frac{(m_j + n_j h)}{2}}}{\Gamma_p(\frac{m_j}{2}) |\Sigma_j|^{\frac{m_j}{2}}} \\
 &= 2^{\frac{N p h}{2}} \prod_{j=1}^k \frac{|\Sigma_j|^{\frac{n_j h}{2}} |I + 2\Sigma_j Y|^{-\frac{(m_j + n_j h)}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})}. \tag{iv}
 \end{aligned}$$

Thus, on substituting (iv) in $E[\lambda^h]$, we have

$$\begin{aligned}
 E[\lambda^h] &= \frac{c^h 2^{\frac{N p h}{2}}}{\Gamma_p(\frac{N h}{2})} \int_{Y > O} |Y|^{\frac{N h}{2} - \frac{p+1}{2}} \\
 &\quad \times \left\{ \prod_{j=1}^k \frac{|\Sigma_j|^{\frac{n_j h}{2}} |I + 2\Sigma_j Y|^{-\frac{(m_j + n_j h)}{2}} \Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \right\} dY, \tag{6.12.5}
 \end{aligned}$$

which is the non-null h -th moment of λ . The h -th null moment is available when $\Sigma_1 = \dots = \Sigma_k = \Sigma$. In the null case,

$$\begin{aligned}
 & \prod_{j=1}^k |I + 2Y \Sigma_j|^{-\frac{(m_j + n_j h)}{2}} |\Sigma_j|^{\frac{n_j h}{2}} \frac{\Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \\
 &= |\Sigma|^{\frac{N h}{2}} |I + 2Y \Sigma|^{-\frac{(N-k + N h)}{2}} \left\{ \prod_{j=1}^k \frac{\Gamma_p(\frac{m_j}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{m_j}{2})} \right\}. \tag{v}
 \end{aligned}$$

Then, substituting (v) in (6.12.5) and integrating out over Y produces

$$E[\lambda^h | H_o] = c^h \frac{\Gamma_p(\frac{N-k}{2})}{\Gamma_p(\frac{N-k}{2} + \frac{N h}{2})} \left\{ \prod_{j=1}^k \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\}. \tag{6.12.6}$$

Observe that when $h = 0$, $E[\lambda^h | H_o] = 1$. For $h = s - 1$ where s is a complex parameter, we have the Mellin transform of the density of λ , denoted by $f(\lambda)$, which can be expressed as follows in terms of an H-function:

$$f(\lambda) = \frac{1}{c} \frac{\Gamma_p(\frac{N-k}{2})}{\left\{ \prod_{j=1}^k \Gamma_p(\frac{n_j-1}{2}) \right\}} H_{1,k}^{k,0} \left[\lambda \left| \begin{matrix} (-\frac{k}{2}, \frac{N}{2}) \\ (-\frac{1}{2}, \frac{n_j}{2}), j=1, \dots, k \end{matrix} \right. \right], \quad 0 < \lambda < 1, \tag{6.12.7}$$

and zero elsewhere, where the H-function is defined in Sect. 5.4.3, more details being available from Mathai and Saxena (1978) and Mathai et al. (2010). Since the coefficients of $\frac{h}{2}$ in the gammas, that is, n_1, \dots, n_k and N , are all positive integers, one can expand all gammas by using the multiplication formula for gamma functions, and then, $f(\lambda)$ can be expressed in terms of a G-function as well. It may be noted from (6.12.5) that for obtaining the non-null moments, and thereby the non-null density, one has to integrate out Y in (6.12.5). This has not yet been worked out for a general k . For $k = 2$, one can obtain a series form in terms of zonal polynomials for the integral in (6.12.5). The rather intricate derivations are omitted.

6.12.1. Asymptotic behavior

We now investigate the asymptotic behavior of $-2 \ln \lambda$ as $n_j \rightarrow \infty$, $j = 1, \dots, k$, $N = n_1 + \dots + n_k$. On expanding the real matrix-variate gamma functions in the the gamma ratio involving h in (6.12.6), we have

$$\frac{\prod_{j=1}^k \Gamma_p\left(\frac{n_j}{2}(1+h) - \frac{1}{2}\right)}{\Gamma_p\left(\frac{N}{2}(1+h) - \frac{k}{2}\right)} \rightarrow \frac{\prod_{j=1}^k \left\{ \prod_{i=1}^p \Gamma\left(\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i-1}{2}\right) \right\}}{\prod_{i=1}^p \Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i-1}{2}\right)}, \quad (i)$$

excluding the factor containing π . Letting $\frac{n_j}{2}(1+h) \rightarrow \infty$, $j = 1, \dots, k$, and $\frac{N}{2}(1+h) \rightarrow \infty$ as $n_j \rightarrow \infty$, $j = 1, \dots, k$, with $N \rightarrow \infty$, we now express all the gamma functions in (i) in terms of Sterling's asymptotic formula. For the numerator, we have

$$\begin{aligned} & \prod_{j=1}^k \left\{ \prod_{i=1}^p \Gamma\left(\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i-1}{2}\right) \right\} \\ & \rightarrow \prod_{j=1}^k \left\{ \prod_{i=1}^p \sqrt{(2\pi)} \left[\frac{n_j}{2}(1+h)\right]^{\frac{n_j}{2}(1+h) - \frac{1}{2} - \frac{i}{2}} e^{-\frac{n_j}{2}(1+h)} \right\} \\ & = \prod_{j=1}^k (\sqrt{2\pi})^p \left[\frac{n_j}{2}(1+h)\right]^{\frac{n_j}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}} e^{-\frac{n_j}{2}(1+h)p} \\ & = (\sqrt{2\pi})^{kp} \left[\prod_{j=1}^k \left(\frac{n_j}{2}\right)^{\frac{n_j}{2}(1+h)p - \frac{p}{2} - \frac{p(p+1)}{4}} \right] e^{-\frac{N}{2}(1+h)p} \\ & \quad \times (1+h)^{\frac{N}{2}(1+h) - \frac{kp}{2} - k\frac{p(p+1)}{4}}, \end{aligned} \quad (ii)$$

and the denominator in (i) has the following asymptotic representation:

$$\begin{aligned} \prod_{i=1}^p \Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i-1}{2}\right) &\rightarrow \prod_{i=1}^p (\sqrt{2\pi}) \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} - \frac{i}{2}} e^{-\frac{N}{2}(1+h)} \\ &= (\sqrt{2\pi})^p \left[\frac{N}{2}\right]^{\frac{N}{2}(1+h)p - \frac{kp}{2} - \frac{p(p+1)}{4}} e^{-\frac{N}{2}(1+h)p} \\ &\quad \times (1+h)^{\frac{N}{2}(1+h)p - \frac{kp}{2} - \frac{p(p+1)}{4}}. \end{aligned} \tag{iii}$$

Now, expanding the gammas in the constant part $\Gamma_p(\frac{N-k}{2}) / \prod_{j=1}^k \Gamma_p(\frac{n_j-1}{2})$ and then taking care of c^h , we see that the factors containing π , the n_j 's and N disappear leaving

$$(1+h)^{-(k-1)\frac{p(p+1)}{4}}.$$

Hence $-2 \ln \lambda \rightarrow \chi_{(k-1)\frac{p(p+1)}{2}}^2$ and hence we have the following result:

Theorem 6.12.1. Consider the λ -criterion in (6.12.2) or the null density in (6.12.7). When $n_j \rightarrow \infty, j = 1, \dots, k$, the asymptotic null density of $-2 \ln \lambda$ is a real scalar chisquare with $(k-1)\frac{p(p+1)}{2}$ degrees of freedom.

Observe that the number of parameters restricted by the null hypothesis $H_o : \Sigma_1 = \dots = \Sigma_k = \Sigma$ where Σ is unknown, is $(k-1)$ times the number of distinct parameters in Σ , which is $\frac{p(p+1)}{2}$, which coincides with the number of degrees of freedom of the asymptotic chisquare distribution under H_o .

6.13. Testing the Hypothesis that k Independent p -variate Real Gaussian Populations are Identical and Multivariate Analysis of Variance

Consider k independent p -variate real Gaussian populations $X_{ij} \sim N_p(\mu^{(i)}, \Sigma_i), \Sigma_i > O, i = 1, \dots, k$, and $j = 1, \dots, n_i$, where the $p \times 1$ vector X_{ij} is the j -th sample value belonging to the i -th population, these samples (iid variables) being of sizes n_1, \dots, n_k from these k populations. The joint density of all the sample values, denoted by L , can be expressed as follows:

$$\begin{aligned} L &= \prod_{i=1}^k L_i, \quad L_i = \prod_{j=1}^{n_i} \frac{e^{-\frac{1}{2} \sum_{j=1}^{n_i} (X_{ij} - \mu^{(i)})' \Sigma_i^{-1} (X_{ij} - \mu^{(i)})}}{(2\pi)^{\frac{n_i p}{2}} |\Sigma_i|^{\frac{n_i}{2}}} \\ &= \frac{e^{-\frac{1}{2} \text{tr}(\Sigma_i^{-1} S_i) - \frac{n_i}{2} (\bar{X}_i - \mu^{(i)})' \Sigma_i^{-1} (\bar{X}_i - \mu^{(i)})}}{(2\pi)^{\frac{n_i}{2}} |\Sigma_i|^{\frac{n_i}{2}}} \end{aligned}$$

where $\bar{X}_i = \frac{1}{n_i}(X_{i1} + \dots + X_{in_i})$, $i = 1, \dots, k$, and $E[X_{ij}] = \mu^{(i)}$, $j = 1, \dots, n_i$. Then, letting $N = n_1 + \dots + n_k$,

$$\begin{aligned} \max_{\Omega} L &= \prod_{i=1}^k \max_{\Omega} L_i = \prod_{i=1}^k \frac{n_i^{\frac{n_i p}{2}} e^{-\frac{n_i p}{2}}}{(2\pi)^{\frac{n_i p}{2}} |S_i|^{\frac{n_i}{2}}} \\ &= \frac{\{\prod_{i=1}^k n_i^{\frac{n_i p}{2}}\} e^{-\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} \{\prod_{i=1}^k |S_i|^{\frac{n_i}{2}}\}}. \end{aligned}$$

Consider the hypothesis $H_o : \mu^{(1)} = \dots = \mu^{(k)} = \mu$, $\Sigma_1 = \dots = \Sigma_k = \Sigma$, where μ and Σ are unknown. This corresponds to the hypothesis of equality of these k populations. Under H_o , the maximum likelihood estimator (MLE) of μ , denoted by $\hat{\mu}$, is given by $\hat{\mu} = \frac{1}{N}[n_1 \bar{X}_1 + \dots + n_k \bar{X}_k]$ where N and \bar{X}_i are as defined above. As for the common Σ , its MLE is

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \hat{\mu})(X_{ij} - \hat{\mu})' \\ &= \frac{1}{N} [S_1 + \dots + S_k + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'] \end{aligned}$$

where S_i is the sample sum of products matrix for the i -th sample, observing that

$$\begin{aligned} \sum_{j=1}^{n_i} (X_{ij} - \hat{\mu})(X_{ij} - \hat{\mu})' &= \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i + \bar{X}_i - \hat{\mu})(X_{ij} - \bar{X}_i + \bar{X}_i - \hat{\mu})' \\ &= \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' + \sum_{j=1}^{n_i} (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \\ &= S_i + n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'. \end{aligned}$$

Hence the maximum of the likelihood function under H_o is the following:

$$\max_{H_o} L = \frac{e^{-\frac{Np}{2}} N^{\frac{Np}{2}}}{(2\pi)^{\frac{Np}{2}} |S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}}$$

where $S = S_1 + \dots + S_k$. Therefore the λ -criterion is given by

$$\lambda = \frac{\max_{H_o}}{\max_{\Omega}} = \frac{\{\prod_{i=1}^k |S_i|^{\frac{n_i}{2}}\} N^{\frac{Np}{2}}}{\{\prod_{i=1}^k n_i^{\frac{n_i p}{2}}\} |S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}}. \quad (6.13.1)$$

6.13.1. Conditional and marginal hypotheses

For convenience, we may split λ into the product $\lambda_1\lambda_2$ where λ_1 is the λ -criterion for the conditional hypothesis $H_{o1} : \mu^{(1)} = \dots = \mu^{(k)} = \mu$ given that $\Sigma_1 = \dots = \Sigma_k = \Sigma$ and λ_2 is the λ -criterion for the marginal hypothesis $H_{o2} : \Sigma_1 = \dots = \Sigma_k = \Sigma$ where μ and Σ are unknown. The conditional hypothesis H_{o1} is actually the null hypothesis usually being made when establishing the multivariate analysis of variance (MANOVA) procedure. We will only consider H_{o1} since the marginal hypothesis H_{o2} has already been discussed in Sect. 6.12. When the Σ_i 's are assumed to be equal, the common Σ is estimated by the MLE $\frac{1}{N}(S_1 + \dots + S_k)$ where S_i is the sample sum of products matrix in the i -th population. The common μ is estimated by $\frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k)$. Accordingly, the λ -criterion for this conditional hypothesis is the following:

$$\lambda_1 = \frac{|S|^{\frac{N}{2}}}{|S + \sum_{i=1}^k n_i(\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'|^{\frac{N}{2}}} \quad (6.13.2)$$

where $S = S_1 + \dots + S_k$, $\hat{\mu} = \frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k)$, $N = n_1 + \dots + n_k$. Note that the S_i 's are independently Wishart distributed with $n_i - 1$ degrees of freedom, that is, $S_i \stackrel{ind}{\sim} W_p(n_i - 1, \Sigma)$, $i = 1, \dots, k$, and hence $S \sim W_p(N - k, \Sigma)$. Let

$$Q = \sum_{i=1}^k n_i(\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})'.$$

Since Q only contains sample averages and the sample averages and the sample sum of products matrices are independently distributed, Q and S are independently distributed. Moreover, since we can write $\bar{X}_i - \hat{\mu}$ as $(\bar{X}_i - \mu) - (\hat{\mu} - \mu)$, where μ is the common true mean value vector, without any loss of generality we can deem the \bar{X}_i 's to be independently $N_p(O, \frac{1}{n_i}\Sigma)$ distributed, $i = 1, \dots, k$, and letting $Y_i = \sqrt{n_i}\bar{X}_i$, one has $Y_i \stackrel{iid}{\sim} N_p(O, \Sigma)$ under the hypothesis H_{o1} . Now, observe that

$$\begin{aligned} \bar{X}_i - \hat{\mu} &= \bar{X}_i - \frac{1}{N}(n_1\bar{X}_1 + \dots + n_k\bar{X}_k) \\ &= -\frac{n_1}{N}\bar{X}_1 - \dots - \frac{n_{i-1}}{N}\bar{X}_{i-1} + \left(1 - \frac{n_i}{N}\right)\bar{X}_i - \frac{n_{i+1}}{N}\bar{X}_{i+1} - \dots - \frac{n_k}{N}\bar{X}_k, \quad (i) \end{aligned}$$

$$\begin{aligned}
Q &= \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \\
&= \sum_{i=1}^k n_i \bar{X}_i \bar{X}_i' - N \hat{\mu} \hat{\mu}' \\
&= \sum_{i=1}^k Y_i Y_i' - \frac{1}{N} (\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k) (\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k)',
\end{aligned} \tag{ii}$$

where $\sqrt{n_1} Y_1 + \cdots + \sqrt{n_k} Y_k = (Y_1, \dots, Y_k) D J$ with J being the $k \times 1$ vector of unities, $J' = (1, \dots, 1)$, and $D = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_k})$. Thus, we can express Q as follows:

$$Q = (Y_1, \dots, Y_k) \left[I - \frac{1}{N} D J J' D \right] (Y_1, \dots, Y_k)'. \tag{iii}$$

Let $B = \frac{1}{N} D J J' D$ and $A = I - B$. Then, observing that $J' D^2 J = N$, both B and A are idempotent matrices, where B is of rank 1 since the trace of B or equivalently the trace of $\frac{1}{N} J' D^2 J$ is equal to one, so that the trace of A which is also its rank, is $k - 1$. Then, there exists an orthonormal matrix P , $P P' = I_k$, $P' P = I_k$, such that $A = P' \begin{bmatrix} I_{k-1} & O \\ O' & 0 \end{bmatrix} P$ where O is a $(k - 1) \times 1$ null vector, O' being its transpose. Letting $(U_1, \dots, U_k) = (Y_1, \dots, Y_k) P'$, the U_i 's are still independently $N_p(O, \Sigma)$ distributed under H_{01} , so that

$$\begin{aligned}
Q &= (U_1, \dots, U_k) \begin{bmatrix} I_{k-1} & O \\ O' & 0 \end{bmatrix} (U_1, \dots, U_k)' = (U_1, \dots, U_{k-1}, O) (U_1, \dots, U_{k-1}, O)' \\
&= \sum_{i=1}^{k-1} U_i U_i' \sim W_p(k - 1, \Sigma).
\end{aligned} \tag{iv}$$

Thus, $S + \sum_{i=1}^k n_i (\bar{X}_i - \hat{\mu})(\bar{X}_i - \hat{\mu})' \sim W_p(N - 1, \Sigma)$, which clearly is not independently distributed of S , referring to the ratio in (6.13.2).

6.13.2. Arbitrary moments of λ_1

Given (6.13.2), we have

$$\lambda_1 = \frac{|S|^{\frac{N}{2}}}{|S + Q|^{\frac{N}{2}}} \Rightarrow \lambda_1^h = |S|^{\frac{Nh}{2}} |S + Q|^{-\frac{Nh}{2}} \tag{v}$$

where $|S + Q|^{-\frac{Nh}{2}}$ will be replaced by the equivalent integral

$$|S + Q|^{-\frac{Nh}{2}} = \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T>O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} e^{-\text{tr}((S+Q)T)} dT \quad (vi)$$

with the $p \times p$ matrix $T > O$. Hence, the h -th moment of λ_1 , for arbitrary h , is the following expected value:

$$E[\lambda_1^h | H_{o1}] = E\{|S|^{\frac{Nh}{2}} |S + Q|^{-\frac{Nh}{2}}\}. \quad (vii)$$

We now evaluate (vii) by integrating out over the Wishart density of S and over the joint multinormal density for U_1, \dots, U_{k-1} :

$$\begin{aligned} E[\lambda_1^h | H_{o1}] &= \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T>O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} \\ &\quad \times \frac{\int_{S>O} |S|^{\frac{Nh}{2} + \frac{N-k}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S) - \text{tr}(ST)} dS}{2^{\frac{(N-k)p}{2}} \Gamma_p(\frac{N-k}{2}) |\Sigma|^{\frac{N-k}{2}}} \\ &\quad \times \int_{U_1, \dots, U_{k-1}} \frac{e^{-\frac{1}{2} \sum_{i=1}^{k-1} U_i' \Sigma^{-1} U_i - \sum_{i=1}^{k-1} \text{tr}(T U_i U_i')}}{(2\pi)^{\frac{(k-1)p}{2}} |\Sigma|^{\frac{k-1}{2}}} dU_1 \wedge \dots \wedge dU_{k-1} \wedge dT. \end{aligned}$$

The integral over S is evaluated as follows:

$$\begin{aligned} &\frac{\int_{S>O} |S|^{\frac{Nh}{2} + \frac{N-k}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}(S+2TS))} dS}{2^{\frac{(N-k)p}{2}} \Gamma_p(\frac{N-k}{2}) |\Sigma|^{\frac{N-k}{2}}} \\ &= 2^{\frac{Nhp}{2}} |\Sigma|^{\frac{Nh}{2}} \frac{\Gamma_p(\frac{N-k}{2} + \frac{Nh}{2})}{\Gamma_p(\frac{N-k}{2})} |I + 2\Sigma T|^{-\frac{Nh}{2} - \frac{N-k}{2}} \end{aligned} \quad (viii)$$

for $I + 2\Sigma T > O$, $\Re(\frac{N-k}{2} + \frac{Nh}{2}) > \frac{p-1}{2}$. The integral over U_1, \dots, U_{k-1} is the following, denoted by δ :

$$\delta = \int_{U_1, \dots, U_{k-1}} \frac{e^{-\frac{1}{2} \sum_{i=1}^{k-1} U_i' \Sigma^{-1} U_i - \text{tr}(\sum_{i=1}^{k-1} T U_i U_i')}}{(2\pi)^{\frac{(k-1)p}{2}} |\Sigma|^{\frac{k-1}{2}}} dU_1 \wedge \dots \wedge dU_{k-1}$$

where

$$\text{tr}\left(\sum_{i=1}^{k-1} T U_i U_i'\right) = \text{tr}\left(\sum_{i=1}^{k-1} U_i' T U_i\right) = \sum_{i=1}^{k-1} U_i' T U_i$$

since $U_i' T U_i$ is scalar; thus, the exponent becomes $-\frac{1}{2} \sum_{i=1}^{k-1} U_i' [\Sigma^{-1} + 2T] U_i$ and the integral simplifies to

$$\delta = |I + 2\Sigma T|^{-\frac{k-1}{2}}, \quad I + 2\Sigma T > O. \quad (ix)$$

Now the integral over T is the following:

$$\begin{aligned} & \frac{1}{\Gamma_p(\frac{Nh}{2})} \int_{T>O} |T|^{\frac{Nh}{2} - \frac{p+1}{2}} |I + 2\Sigma T|^{-\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} \\ &= \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{k-1}{2}\right)}{\Gamma_p\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} 2^{-\frac{Nhp}{2}} |\Sigma|^{-\frac{Nh}{2}}, \quad \Re\left(\frac{N-k}{2} + \frac{Nh}{2}\right) > \frac{p-1}{2}. \end{aligned} \quad (x)$$

Therefore,

$$E[\lambda_1^h | H_{o1}] = \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{Nh}{2}\right)}{\Gamma_p\left(\frac{N-k}{2}\right)} \frac{\Gamma_p\left(\frac{N-k}{2} + \frac{k-1}{2}\right)}{\Gamma_p\left(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2}\right)} \quad (6.13.3)$$

for $\Re\left(\frac{N-k}{2} + \frac{Nh}{2}\right) > \frac{p-1}{2}$.

6.13.3. The asymptotic distribution of $-2 \ln \lambda_1$

An asymptotic distribution of $-2 \ln \lambda_1$ as $N \rightarrow \infty$ can be derived from (6.13.3). First, on expanding the real matrix-variate gamma functions in (6.13.3), we obtain the following representation of the h -th null moment of λ_1 :

$$\begin{aligned} E[\lambda_1^h | H_{o1}] &= \left\{ \prod_{j=1}^p \frac{\Gamma\left(\frac{N-k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N-k}{2} - \frac{j-1}{2}\right)} \right\} \\ &\times \left\{ \prod_{j=1}^p \frac{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)} \right\}. \end{aligned} \quad (xi)$$

Let us now express all the gamma functions in terms of Sterling's asymptotic formula by taking $\frac{N}{2} \rightarrow \infty$ in the constant part and $\frac{N}{2}(1+h) \rightarrow \infty$ in the part containing h . Then,

$$\begin{aligned} \frac{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)} &\rightarrow \frac{(2\pi)^{\frac{1}{2}} \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} - \frac{j}{2}} e^{\frac{N}{2}(1+h)}}{(2\pi)^{\frac{1}{2}} \left[\frac{N}{2}(1+h)\right]^{\frac{N}{2}(1+h) - \frac{k}{2} + \frac{k-1}{2} - \frac{j}{2}} e^{\frac{N}{2}(1+h)}} \\ &= \frac{1}{\left[\frac{N}{2}(1+h)\right]^{\frac{k-1}{2}}}, \end{aligned} \quad (xii)$$

$$\frac{\Gamma\left(\frac{N}{2} - \frac{k}{2} + \frac{k-1}{2} - \frac{j-1}{2}\right)}{\Gamma\left(\frac{N}{2} - \frac{k}{2} - \frac{j-1}{2}\right)} \rightarrow \left(\frac{N}{2}\right)^{\frac{(k-1)}{2}}. \quad (xiii)$$

Hence,

$$E[\lambda_1^h | H_{o1}] \rightarrow (1+h)^{-\frac{(k-1)p}{2}} \text{ as } N \rightarrow \infty. \quad (6.13.4)$$

Thus, the following result:

Theorem 6.13.1. For the test statistic λ_1 given in (6.13.2), $-2 \ln \lambda_1 \rightarrow \chi_{(k-1)p}^2$, that is, $-2 \ln \lambda_1$ tends to a real scalar chisquare variable having $(k-1)p$ degrees of freedom as $N \rightarrow \infty$ with $N = n_1 + \dots + n_k$, n_j being the sample size of the j -th p -variate real Gaussian population.

Under the marginal hypothesis $H_{o2} : \Sigma_1 = \dots = \Sigma_k = \Sigma$ where Σ is unknown, the λ -criterion is denoted by λ_2 , and its h -th moment, which is available from (6.12.6) of Sect. 6.12, is given by

$$E[\lambda_2^h | H_{o2}] = c^h \frac{\Gamma_p(\frac{N-k}{2})}{\Gamma_p(\frac{N-k}{2} + \frac{Nh}{2})} \left\{ \prod_{j=1}^p \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\} \quad (6.13.5)$$

for $\Re(\frac{n_j-1}{2} + \frac{n_j h}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$. Hence the h -th null moment of the λ criterion for testing the hypothesis H_o of equality of the k independent p -variate real Gaussian populations is the following:

$$\begin{aligned} E[\lambda^h | H_o] &= E[\lambda_1^h | H_{o1}] E[\lambda_2^h | H_{o2}] \\ &= \frac{\Gamma_p(\frac{N-k}{2} + \frac{k-1}{2})}{\Gamma_p(\frac{Nh}{2} + \frac{N-k}{2} + \frac{k-1}{2})} c^h \left\{ \prod_{j=1}^p \frac{\Gamma_p(\frac{n_j-1}{2} + \frac{n_j h}{2})}{\Gamma_p(\frac{n_j-1}{2})} \right\} \end{aligned} \quad (6.13.6)$$

for $\Re(\frac{n_j-1}{2} + \frac{n_j h}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, $N = n_1 + \dots + n_k$, where c is the constant associated with the h -th moment of λ_2 . Combining Theorems 6.13.1 and Theorem 6.12.1, the asymptotic distribution of $-2 \ln \lambda$ of (6.13.6) is a real scalar chisquare with $(k-1)p + (k-1)\frac{p(p+1)}{2}$ degrees of freedom. Thus, the following result:

Theorem 6.13.2. For the λ -criterion for testing the hypothesis of equality of k independent p -variate real Gaussian populations, $-2 \ln \lambda \rightarrow \chi_v^2$, $v = (k-1)p + (k-1)\frac{p(p+1)}{2}$ as $n_j \rightarrow \infty$, $j = 1, \dots, k$.

Note 6.13.1. Observe that for the conditional hypothesis H_{o1} in (6.13.2), the degrees of freedom of the asymptotic chisquare distribution of $-2 \ln \lambda_1$ is $(k-1)p$, which is also the number of parameters restricted by the hypothesis H_{o1} . For the hypothesis H_{o2} , the corresponding degrees of freedom of the asymptotic chisquare distribution of $-2 \ln \lambda_2$ is

$(k-1)\frac{p(p+1)}{2}$, which as well is the number of parameters restricted by the hypothesis H_{o2} . The asymptotic chisquare distribution of $-2\ln\lambda$ for the hypothesis H_o of equality of k independent p -variate Gaussian populations the degrees of freedom is the sum of these two quantities, that is, $(k-1)p + (k-1)\frac{p(p+1)}{2} = p(k-1)\frac{(p+3)}{2}$, which also coincides with the number of parameters restricted under the hypothesis H_o .

Exercises

6.1. Derive the λ -criteria for the following tests in a real univariate Gaussian population $N_1(\mu, \sigma^2)$, assuming that a simple random sample of size n , namely x_1, \dots, x_n , which are iid as $N_1(\mu, \sigma^2)$, is available: (1): $\mu = \mu_o$ (given), σ^2 is known; (2): $\mu = \mu_o$, σ^2 unknown; (3): $\sigma^2 = \sigma_o^2$ (given), also you may refer to Mathai and Haubold (2017).

In all the following problems, it is assumed that a simple random sample of size n is available. The alternative hypotheses are the natural alternatives.

6.2. Repeat Exercise 6.1 for the corresponding complex Gaussian.

6.3. Construct the λ -criteria in the complex case for the tests discussed in Sects. 6.2–6.4.

6.4. In the real p -variate Gaussian case, consider the hypotheses (1): Σ is diagonal or the individual components are independently distributed; (2): The diagonal elements are equal, given that Σ is diagonal (which is a conditional test). Construct the λ -criterion in each case.

6.5. Repeat Exercise 6.4 for the complex Gaussian case.

6.6. Let the population be real p -variate Gaussian $N_p(\mu, \Sigma)$, $\Sigma = (\sigma_{ij}) > O$, $\mu' = (\mu_1, \dots, \mu_p)$. Consider the following tests and compute the λ -criteria: (1): $\sigma_{11} = \dots = \sigma_{pp} = \sigma^2$, $\sigma_{ij} = \nu$ for all i and j , $i \neq j$. That is, all the variances are equal and all the covariances are equal; (2): In addition to (1), $\mu_1 = \mu_2 = \dots = \mu$ or all the mean values are equal. Construct the λ -criterion in each case. The first one is known as L_{vc} criterion and the second one is known L_{mvc} criterion. Repeat the same exercise for the complex case. Some distributional aspects are examined in Mathai (1970b) and numerical tables are available in Mathai and Katiyar (1979b).

6.7. Let the population be real p -variate Gaussian $N_p(\mu, \Sigma)$, $\Sigma > O$. Consider the hypothesis (1):

$$\Sigma = \begin{bmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{bmatrix}, \quad a \neq 0, \quad b \neq 0, \quad a \neq b.$$

(2):

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} a_1 & b_1 & \cdots & b_1 \\ b_1 & a_1 & \cdots & b_1 \\ \vdots & \vdots & \ddots & \vdots \\ b_1 & b_1 & \cdots & a_1 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} a_2 & b_2 & \cdots & b_2 \\ b_2 & a_2 & \cdots & b_2 \\ \vdots & \vdots & \ddots & \vdots \\ b_2 & b_2 & \cdots & a_2 \end{bmatrix}$$

where $a_1 \neq 0$, $a_2 \neq 0$, $b_1 \neq 0$, $b_2 \neq 0$, $a_1 \neq b_1$, $a_2 \neq b_2$, $\Sigma_{12} = \Sigma'_{21}$ and all the elements in Σ_{12} and Σ_{21} are each equal to $c \neq 0$. Construct the λ -criterion in each case. (These are hypotheses on patterned matrices).

6.8. Repeat Exercise 6.7 for the complex case.

6.9. Consider k independent real p -variate Gaussian populations with different parameters, distributed as $N_p(M_j, \Sigma_j)$, $\Sigma_j > O$, $M'_j = (\mu_{1j}, \dots, \mu_{pj})$, $j = 1, \dots, k$. Construct the λ -criterion for testing the hypothesis $\Sigma_1 = \dots = \Sigma_k$ or the covariance matrices are equal. Assume that simple random samples of sizes n_1, \dots, n_k are available from these k populations.

6.10. Repeat Exercise 6.9 for the complex case.

6.11. For the second part of Exercise 6.7, which is also known as Wilks' L_{mvc} criterion, show that if $u = \lambda^{\frac{2}{n}}$ where λ is the likelihood ratio criterion and n is the sample size, then

$$u = \frac{|S|}{[s + (p-1)s_1][s - s_1 + \frac{n}{p-1} \sum_{j=1}^p (\bar{x}_j - \bar{x})^2]} \quad (i)$$

where $S = (s_{ij})$ is the sample sum of products matrix, $s = \frac{1}{p} \sum_{i=1}^p s_{ii}$, $s_1 = \frac{1}{p(p-1)} \sum_{i \neq j=1}^p s_{ij}$, $\bar{x} = \frac{1}{p} \sum_{i=1}^p \bar{x}_i$, $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik}$. For the statistic u in (i), show that the h -th null moment or the h -th moment when the null hypothesis is true, is given by the following:

$$E[u^h | H_0] = \prod_{j=0}^{p-2} \frac{\Gamma(\frac{n-1}{2} + h - \frac{j}{2})}{\Gamma(\frac{n-1}{2} - \frac{j}{2})} \frac{\Gamma(\frac{n+1}{2} + \frac{j}{p-1})}{\Gamma(\frac{n+1}{2} + h + \frac{j}{p-1})}. \quad (ii)$$

Write down the conditions for the existence of the moment in (ii). [For the null and non-null distributions of Wilks' L_{mvc} criterion, see Mathai (1978).]

6.12. Let the $(p + q) \times 1$ vector X have a $(p + q)$ -variate nonsingular real Gaussian distribution, $X \sim N_{p+q}(\mu, \Sigma)$, $\Sigma > O$. Let

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2' & \Sigma_3 \end{bmatrix}$$

where Σ_1 is $p \times p$ with all its diagonal elements equal to σ_{aa} and all other elements equal to $\sigma_{aa'}$, Σ_2 has all elements equal to σ_{ab} , Σ_3 has all diagonal elements equal to σ_{bb} and all other elements equal to $\sigma_{bb'}$ where σ_{aa} , $\sigma_{aa'}$, σ_{bb} , $\sigma_{bb'}$ are unknown. Then, Σ is known as bipolar. Let λ be the likelihood ratio criterion for testing the hypothesis that Σ is bipolar. Then show that the h -th null moment is the following:

$$E[\lambda^h | H_0] = [(p-1)^{p-1} (q-1)^{q-1}] \frac{\Gamma[\frac{(q-1)(n-1)}{2}] \Gamma[\frac{(p-1)(n-1)}{2}]}{\Gamma[(p-1)(h + \frac{n-1}{2})] \Gamma[(q-1)(h + \frac{n-1}{2})]} \\ \times \prod_{j=0}^{p+q-3} \frac{\Gamma[h + \frac{n-3}{2} - \frac{j}{2}]}{\Gamma[\frac{n-3}{2} - \frac{j}{2}]}$$

where n is the sample size. Write down the conditions for the existence of this h -th null moment.

6.13. Let X be $m \times n$ real matrix having the matrix-variate Gaussian density

$$f(X) = \frac{1}{|2\pi \Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}[\Sigma^{-1}(X-M)(X-M)']}, \quad \Sigma > O.$$

Letting $S = XX'$, S is a non-central Wishart matrix. Derive the density of S and show that this density, denoted by $f_s(S)$, is the following:

$$f_s(S) = \frac{1}{\Gamma_m(\frac{n}{2}) |2\Sigma|^{\frac{n}{2}}} |S|^{\frac{n}{2} - \frac{m+1}{2}} e^{-\text{tr}(\Omega) - \frac{1}{2} \text{tr}(\Sigma^{-1}S)} \\ \times {}_0F_1\left(\ ; \frac{n}{2}; \frac{1}{2} \Sigma^{-1}S\right)$$

where $\Omega = \frac{1}{2} MM' \Sigma^{-1}$ is the non-centrality parameter and ${}_0F_1$ is a Bessel function of matrix argument.

6.14. Show that the h -th moment of the determinant of S , the non-central Wishart matrix specified in Exercise 6.13, is given by

$$E[|S|^h] = \frac{\Gamma_m(h + \frac{n}{2})}{\Gamma(\frac{n}{2}) |2\Sigma|^h} e^{-\text{tr}(\Omega)} {}_1F_1\left(h + \frac{n}{2}; \frac{n}{2}; \Omega\right)$$

where ${}_1F_1$ is a hypergeometric function of matrix argument and Ω is the non-centrality parameter defined in Exercise 6.13.

6.15. Letting $v = \lambda^{\frac{2}{n}}$ in Eq. (6.3.10), show that, under the null hypothesis H_0 , v is distributed as a real scalar type-1 beta with the parameters $(\frac{n-1}{2}, \frac{1}{2})$ and that $\frac{nA'\bar{X}\bar{X}'A}{A'SA}$ is real scalar type-2 beta distributed with the parameters $(\frac{1}{2}, \frac{n-1}{2})$.

6.16. Show that for an arbitrary h , the h -th null moment of the test statistic λ specified in Eq. (6.3.10) is

$$E[\lambda^h | H_0] = \frac{\Gamma(\frac{n-1}{2} + \frac{nh}{2})}{\Gamma(\frac{n-1}{2})} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2} + \frac{nh}{2})}, \Re(h) > -\frac{n-1}{n}.$$

References

T.W. Anderson (2003): *An Introduction to Multivariate Statistical Analysis*, Third Edition, Wiley, New York.

A.W. Davis (1971): Percentile approximations for a class of likelihood ratio criteria, *Biometrika*, **58**, 349–356.

A.W. Davis and J.B. Field (1971): Tables of some multivariate test criteria, *Technical Report No. 32*, Division of Mathematical Statistics, CSIRO, Canberra, Australia.

B.P. Korin (1968): On the distribution of a statistic used for testing a covariance matrix, *Biometrika*, **55**, 171–178.

A.M. Mathai (1970a): An expansion of Meijer's G-function in the logarithmic case with applications, *Mathematische Nachrichten*, **48**, 129–139.

A.M. Mathai (1970b): The exact distribution of Bartlett's criterion, *Publ. L'ISUP, Paris*, **19**, 1–15.

A.M. Mathai (1971): On the distribution of the likelihood ratio criterion for testing linear hypotheses on regression coefficients, *Annals of the Institute of Statistical Mathematics*, **23**, 181–197.

A.M. Mathai (1972a): The exact distribution of three tests associated with Wilks' concept of generalized variance, *Sankhya Series A*, **34**, 161–170.

A.M. Mathai (1972b): The exact non-central distribution of the generalized variance, *Annals of the Institute of Statistical Mathematics*, **24**, 53–65.

A.M. Mathai (1973): A review of the various methods of obtaining the exact distributions of multivariate test criteria, *Sankhya Series A*, **35**, 39–60.

A.M. Mathai (1977): A short note on the non-null distribution for the problem of sphericity, *Sankhya Series B*, **39**, 102.

A.M. Mathai (1978): On the non-null distribution of Wilks' L_{mvc} , *Sankhya Series B*, **40**, 43–48.

A.M. Mathai (1979a): The exact distribution and the exact percentage points for testing equality of variances in independent normal population, *Journal of Statistical Computations and Simulation*, **9**, 169–182.

A.M. Mathai (1979): On the non-null distributions of test statistics connected with exponential populations, *Communications in Statistics (Theory & Methods)*, **8(1)**, 47–55.

A.M. Mathai (1982): On a conjecture in geometric probability regarding asymptotic normality of a random simplex, *Annals of Probability*, **10**, 247–251.

A.M. Mathai (1984): On multi-sample sphericity (in Russian), *Investigations in Statistics (Russian)*, Tom **136**, 153–161.

A.M. Mathai (1985): Exact and asymptotic procedures in the distribution of multivariate test statistic, In *Statistical Theory and Data Analysis*, K. Matusita Editor, North-Holland, pp. 381–418.

A.M. Mathai (1986): Hypothesis of multisample sphericity, *Journal of Mathematical Sciences*, **33(1)**, 792–796.

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai and H.J. Haubold (2017): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai and R.S. Katiyar (1979a): Exact percentage points for testing independence, *Biometrika*, **66**, 353–356.

A.M. Mathai and R. S. Katiyar (1979b): The distribution and the exact percentage points for Wilks' L_{mvc} criterion, *Annals of the Institute of Statistical Mathematics*, **31**, 215–224.

A.M. Mathai and R.S. Katiyar (1980): Exact percentage points for three tests associated with exponential populations, *Sankhya Series B*, **42**, 333–341.

A.M. Mathai and P.N. Rathie (1970): The exact distribution of Votaw's criterion, *Annals of the Institute of Statistical Mathematics*, **22**, 89–116.

A.M. Mathai and P.N. Rathie (1971): Exact distribution of Wilks' criterion, *The Annals of Mathematical Statistics*, **42**, 1010–1019.

A.M. Mathai and R.K. Saxena (1973): *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Springer-Verlag, Lecture Notes No. 348, Heidelberg and New York.

A.M. Mathai and R.K. Saxena (1978): *The H-function with Applications in Statistics and Other Disciplines*, Wiley Halsted, New York and Wiley Eastern, New Delhi.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.

B.N. Nagarsenker and K.C.S. Pillai (1973): The distribution of the sphericity test criterion, *Journal of Multivariate Analysis*, **3**, 226–2235.

N. Sugiura and H. Nagao (1968): Unbiasedness of some test criteria for the equality of one or two covariance matrices, *Annals of Mathematical Statistics*, **39**, 1686–1692.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 7

Rectangular Matrix-Variate Distributions



7.1. Introduction

Thus far, we have primarily been dealing with distributions involving real positive definite or Hermitian positive definite matrices. We have already considered rectangular matrices in the matrix-variate Gaussian case. In this chapter, we will examine rectangular matrix-variate gamma and beta distributions and also consider to some extent other types of distributions. We will begin with the rectangular matrix-variate real gamma distribution, a version of which was discussed in connection with the pathway model introduced in Mathai (2005). The notations will remain as previously specified. Lower-case letters such as x, y, z will denote real scalar variables, whether mathematical or random. Capital letters such as X, Y will be used for matrix-variate variables, whether square or rectangular. In the complex domain, a tilde will be placed above the corresponding scalar and matrix-variables; for instance, we will write $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$. Constant matrices will be denoted by upper-case letter such as A, B, C . A tilde will not be utilized for constant matrices except for stressing the point that the constant matrix is in the complex domain. When X is a $p \times p$ real positive definite matrix, then $A < X < B$ will imply that the constant matrices A and B are positive definite, that is, $A > O, B > O$, and further that $X > O, X - A > O, B - X > O$. Real positive definite matrices will be assumed to be symmetric. The corresponding notation for a $p \times p$ Hermitian positive definite matrix is $A < \tilde{X} < B$. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ whereas, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as being $p \times p$ unless specified otherwise. Whenever A is a real $p \times q, q \geq p$, rectangular matrix of full rank p, AA' is positive definite, a prime denoting the transpose. When A is in the complex domain, then AA^* is Hermitian positive definite where an A^* indicates the complex conjugate transpose of A . Note that all positive definite complex matrices are necessarily Hermitian. As well, dX will denote the wedge product of all differentials in the matrix X . If $X = (x_{ij})$ is a

real $p \times q$ matrix, then $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. Whenever $X = (x_{ij})'$ is a $p \times p$ real symmetric matrix, $dX = \wedge_{i \geq j} dx_{ij} = \wedge_{i \leq j} dx_{ij}$, that is, the wedge product of the $\frac{p(p+1)}{2}$ distinct differentials. As for the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

7.2. Rectangular Matrix-Variate Gamma Density, Real Case

The most commonly utilized real gamma type distributions are the gamma, generalized gamma and Wishart in Statistics and the Maxwell-Boltzmann and Raleigh in Physics. The first author has previously introduced real and complex matrix-variate analogues of the gamma, Maxwell-Boltzmann, Raleigh and Wishart densities where the matrices are $p \times p$ real positive definite or Hermitian positive definite. For the generalized gamma density in the real scalar case, a matrix-variate analogue can be written down but the associated properties cannot be studied owing to the problem of making a transformation of the type $Y = X^\delta$ for $\delta \neq \pm 1$; additionally, when X is real positive definite or Hermitian positive definite, the Jacobians will produce awkward forms that cannot be easily handled, see Mathai (1997) for an illustration wherein $\delta = 2$ and the matrix X is real and symmetric. Thus, we will provide extensions of the gamma, Wishart, Maxwell-Boltzmann and Raleigh densities to the rectangular matrix-variate cases for $\delta = 1$, in both the real and complex domains.

The Maxwell-Boltzmann and Raleigh densities are associated with numerous problems occurring in Physics. A multivariate analogue as well as a rectangular matrix-variate analogue of these densities may become useful in extending the usual theories giving rise to these univariate densities, to multivariate and matrix-variate settings. It will be shown that, as was explained in Mathai (1999), this problem is also connected to the volumes of parallelotopes determined by p linearly independent random points in the Euclidean n -space, $n \geq p$. Structural decompositions of the resulting random determinants and pathway extensions to gamma, Wishart, Maxwell-Boltzmann and Raleigh densities will also be considered.

In the current nuclear reaction-rate theory, the basic distribution being assumed for the relative velocity of reacting particles is the Maxwell-Boltzmann. One of the forms of this density for the real scalar positive variable case is

$$f_1(x) = \frac{4}{\sqrt{\pi}} \beta^{\frac{3}{2}} x^2 e^{-\beta x^2}, \quad 0 \leq x < \infty, \quad \beta > 0, \quad (7.2.1)$$

and $f_1(x) = 0$ elsewhere. The Raleigh density is given by

$$f_2(x) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, \quad 0 \leq x < \infty, \quad \alpha > 0, \quad (7.2.2)$$

and $f_2 = 0$ elsewhere, and the three-parameter generalized gamma density has the form

$$f_3(x) = \frac{\delta b^{\frac{\alpha}{\delta}}}{\Gamma(\frac{\alpha}{\delta})} x^{\alpha-1} e^{-bx^\delta}, \quad x \geq 0, \quad b > 0, \quad \alpha > 0, \quad \delta > 0, \quad (7.2.3)$$

and $f_3 = 0$ elsewhere. Observe that (7.2.1) and (7.2.2) are special cases of (7.2.3). For derivations of a reaction-rate probability integral based on Maxwell-Boltzmann velocity density, the reader is referred to Mathai and Haubold (1988). Various basic results associated with the Maxwell-Boltzmann distribution are provided in Barnes et al. (1982), Critchfield (1972), Fowler (1984), and Pais (1986), among others. The Maxwell-Boltzmann and Raleigh densities have been extended to the real positive definite matrix-variate and the real rectangular matrix-variate cases in Mathai and Princy (2017). These results will be included in this section, along with extensions of the gamma and Wishart densities to the real and complex rectangular matrix-variate cases. Extensions of the gamma and Wishart densities to the real positive definite and complex Hermitian positive definite matrix-variate cases have already been discussed in Chap. 5. The Jacobians that are needed and will be frequently utilized in our discussion are already provided in Chaps. 1 and 4, further details being available from Mathai (1997). The previously defined real matrix-variate gamma $\Gamma_p(\alpha)$ and complex matrix-variate gamma $\tilde{\Gamma}_p(\alpha)$ functions will also be utilized in this chapter.

7.2.1. Extension of the gamma density to the real rectangular matrix-variate case

Consider a $p \times q$, $q \geq p$, real matrix X of full rank p , whose rows are thus linearly independent, and a real-valued scalar function $f(XX')$ whose integral over X is convergent, that is, $\int_X f(XX')dX < \infty$. Letting $S = XX'$, S will be symmetric as well as real positive definite meaning that for every $p \times 1$ non-null vector Y , $Y'SY > 0$ for all $Y \neq O$ (a non-null vector). Then, $S = (s_{ij})$ will involve only $\frac{p(p+1)}{2}$ differential elements, that is, $dS = \wedge_{i \geq j=1}^p ds_{ij}$, whereas dX will contain pq differential elements dx_{ij} 's. As has previously been explained in Chap. 4, the connection between dX and dS can be established via a sequence of two or three matrix transformations.

Let the $X = (x_{ij})$ be a $p \times q$, $q \geq p$, real matrix of rank p where the x_{ij} 's are distinct real scalar variables. Let A be a $p \times p$ real positive definite constant matrix and B be a $q \times q$ real positive definite constant matrix, $A^{\frac{1}{2}}$ and $B^{\frac{1}{2}}$ denoting the respective positive definite square roots of the positive definite matrices A and B . We will now determine the value of c that satisfies the following integral equation:

$$\frac{1}{c} = \int_X |AXBX'|^\gamma e^{-\text{tr}(AXBX')} dX. \quad (i)$$

Note that $\text{tr}(AXBX') = \text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}})$. Letting $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$, it follows from Theorem 1.7.4 that $dY = |A|^{\frac{q}{2}}|B|^{\frac{p}{2}}dX$. Thus,

$$\frac{1}{c} = |A|^{-\frac{q}{2}}|B|^{-\frac{p}{2}} \int_Y |YY'|^\gamma e^{-\text{tr}(YY')} dY. \quad (ii)$$

Letting $S = YY'$, we note that S is a $p \times p$ real positive definite matrix, and on applying Theorem 4.2.3, we have $dY = \frac{\pi^{\frac{qp}{2}}}{\Gamma_p(\frac{q}{2})} |S|^{\frac{q}{2} - \frac{p+1}{2}} dS$ where $\Gamma_p(\cdot)$ is the real matrix-variate gamma function. Thus,

$$\frac{1}{c} = |A|^{-\frac{q}{2}}|B|^{-\frac{p}{2}} \frac{\pi^{\frac{qp}{2}}}{\Gamma_p(\frac{q}{2})} \int_{S>O} |S|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} e^{-\text{tr}(S)} dS, \quad A > O, B > O, \quad (iii)$$

the integral being a real matrix-variate gamma integral given by $\Gamma_p(\gamma + \frac{q}{2})$ for $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, where $\Re(\cdot)$ is the real part of (\cdot) , so that

$$c = \frac{|A|^{\frac{q}{2}}|B|^{\frac{p}{2}}\Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}}\Gamma_p(\gamma + \frac{q}{2})} \text{ for } \Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}, \quad A > O, B > O. \quad (7.2.4)$$

Let

$$f_4(X) = c |AXBX'|^\gamma e^{-\text{tr}(AXBX')} \quad (7.2.5)$$

for $A > O$, $B > O$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, $X = (x_{ij})$, $-\infty < x_{ij} < \infty$, $i = 1, \dots, p$, $j = 1, \dots, q$, where c is as specified in (7.2.4). Then, $f_4(X)$ is a statistical density that will be referred to as the rectangular real matrix-variate gamma density with shape parameter γ and scale parameter matrices $A > O$ and $B > O$. Although the parameters are usually real in a statistical density, the above conditions apply to the general complex case.

For $p = 1$, $q = 1$, $\gamma = 1$, $A = 1$ and $B = \beta > 0$, we have $|AXBX'| = \beta x^2$ and

$$\frac{|A|^{\frac{q}{2}}|B|^{\frac{p}{2}}\Gamma(\frac{q}{2})}{\pi^{\frac{qp}{2}}\Gamma_p(\gamma + \frac{q}{2})} = \frac{(\beta)^{\frac{1}{2}}\Gamma(\frac{1}{2})}{\pi^{\frac{1}{2}}\Gamma(\frac{3}{2})} = \frac{2\sqrt{\beta}}{\sqrt{\pi}},$$

so that $c = \frac{2}{\sqrt{\pi}}\beta^{\frac{3}{2}}$ for $-\infty < x < \infty$. Note that when the support of $f(x)$ is restricted to the interval $0 \leq x < \infty$, the normalizing constant will be multiplied by 2, $f(x)$ being a symmetric function. Then, for this particular case, $f_4(X)$ in (7.2.5) agrees with the Maxwell-Boltzmann density for the real scalar positive variable x whose density is given in (7.2.1). Accordingly, when $\gamma = 1$, (7.2.5) with c as specified in (7.2.4) will be referred to as the real rectangular matrix-variate Maxwell-Boltzmann density. Observe that

for $\gamma = 0$, (7.2.5) is the real rectangular matrix-variate Gaussian density that was considered in Chap. 4. In the Raleigh case, letting $p = 1$, $q = 1$, $A = 1$, $B = \frac{1}{2\alpha^2}$ and $\gamma = \frac{1}{2}$,

$$|AXBX'|^\gamma = \left(\frac{x^2}{2\alpha^2}\right)^{\frac{1}{2}} = \frac{|x|}{\sqrt{2}|\alpha|} \text{ and } c = \frac{1}{\sqrt{2}\alpha}$$

which gives

$$f_5(x) = \frac{|x|}{2\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, \quad -\infty < x < \infty \text{ or } f_5(x) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, \quad 0 \leq x < \infty,$$

for $\alpha > 0$, and $f_5 = 0$ elsewhere where $|x|$ denotes the absolute value of x , which is the real positive scalar variable case of the Raleigh density given in (7.2.2). Accordingly, (7.2.5) with c as specified in (7.2.4) wherein $\gamma = \frac{1}{2}$ will be called the real rectangular matrix-variate Raleigh density.

From (7.2.5), which is the density for $X = (x_{ij})$, $p \times q$, $q \geq p$ of rank p , with $-\infty < x_{ij} < \infty$, $i = 1, \dots, p$, $j = 1, \dots, q$, we obtain the following density for $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$:

$$f_6(Y)dY = \frac{\Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}} \Gamma_p(\gamma + \frac{q}{2})} |YY'|^\gamma e^{-\text{tr}(YY')} dY \quad (7.2.6)$$

for $\gamma + \frac{q}{2} > \frac{p-1}{2}$, and $f_6 = 0$ elsewhere. We will refer to (7.2.6) as the standard form of the real rectangular matrix-variate gamma density. The density of $S = YY'$ is then

$$f_7(S) dS = \frac{1}{\Gamma_p(\gamma + \frac{q}{2})} |S|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} e^{-\text{tr}(S)} dS \quad (7.2.7)$$

for $S > O$, $\gamma + \frac{q}{2} > \frac{p-1}{2}$, and $f_7 = 0$ elsewhere.

Example 7.2.1. Specify the distribution of $u = \text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}})$, the exponent of the density given in (7.2.5).

Solution 7.2.1. Let us determine the moment generating function (mgf) of u with parameter t . That is,

$$\begin{aligned} M_u(t) &= E[e^{tu}] = E[e^{t \text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}})}] \\ &= c \int_X |A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^\gamma e^{-(1-t)\text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}})} dX \end{aligned}$$

where c is given in (7.2.4). Let us make the following transformations: $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$, $S = YY'$. Then, all factors, except $\Gamma_p(\gamma + \frac{q}{2})$, are canceled and the mgf becomes

$$M_u(t) = \frac{1}{\Gamma_p(\gamma + \frac{q}{2})} \int_{S>0} |S|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} e^{-(1-t)\text{tr}(S)} dS$$

for $1 - t > 0$. On making the transformation $(1 - t)S = S_1$ and then integrating out S_1 , we obtain the following representation of the moment generating function:

$$M_u(t) = (1 - t)^{-p(\gamma + \frac{q}{2})}, \quad 1 - t > 0,$$

which happens to be the mgf of a real scalar gamma random variables with the parameters $(\alpha = p(\gamma + \frac{q}{2}), \beta = 1)$, which owing to the uniqueness of the mgf, is the distribution of u .

Example 7.2.2. Let $U_1 = A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$, $U_2 = XBX'$, $U_3 = B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}$, $U_4 = X'AX$. Determine the corresponding densities when they exist.

Solution 7.2.2. Let us examine the exponent in the density (7.2.5). By making use of the commutative property of trace, one can write

$$\text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}) = \text{tr}[A(XBX')] = \text{tr}(B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}) = \text{tr}[B(X'AX)].$$

Observe that the exponent depends on the matrix $A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$, which is symmetric and positive definite, and that the functional part of the density also involves its determinant. Thus, the structure is that of real matrix-variate gamma density; however, (7.2.5) gives the density of X . Hence, one has to reach U_1 from X and derive the density of U_1 . Consider the transformation $Y = A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$. This will bring X to Y . Now, let $S = YY' = U_1$ so that the matrix U_1 has the real matrix-variate gamma distribution specified in (7.2.7), that is, U_1 is a real matrix-variate gamma variable with shape parameter $\gamma + \frac{q}{2}$ and scale parameter matrix I . Next, consider U_2 . Let us obtain the density of U_2 from the density (7.2.5) for X . Proceeding as above while ignoring A or taking $A = I$, (7.2.7) will become the following density, denoted by $f_{u_2}(U_2)$:

$$f_{u_2}(U_2)dU_2 = \frac{|A|^{\gamma + \frac{q}{2}}}{\Gamma_p(\gamma + \frac{q}{2})} |U_2|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} e^{-\text{tr}(AU_2)} dU_2,$$

which shows that U_2 is a real matrix-variate gamma variable with shape parameter $\gamma + \frac{q}{2}$ and scale parameter matrix A . With respect to U_3 and U_4 , when $q > p$, one has the positive semi-definite factor $X'BX$ whose determinant is zero; hence, in this singular case, the

densities do not exist for U_3 and U_4 . However, when $q = p$, U_3 has a real matrix-variate gamma distribution with shape parameter $\gamma + \frac{p}{2}$ and scale parameter matrix I and U_4 has a real matrix-variate gamma distribution with shape parameter $\gamma + \frac{p}{2}$ and scale parameter matrix B , observing that when $q = p$ both U_3 and U_4 are $q \times q$ and positive definite. This completes the solution.

The above findings are stated as a theorem:

Theorem 7.2.1. *Let $X = (x_{ij})$ be a real full rank $p \times q$ matrix, $q \geq p$, having the density specified in (7.2.5). Let U_1, U_2, U_3 and U_4 be as defined in Example 7.2.2. Then, U_1 is real matrix-variate gamma variable with scale parameter matrix I and shape parameter $\gamma + \frac{q}{2}$; U_2 is real matrix-variate gamma variable with shape parameter $\gamma + \frac{q}{2}$ and scale parameter matrix A ; U_3 and U_4 are singular and do not have densities when $q > p$; however, and when $q = p$, U_3 is real matrix-variate gamma distributed with shape parameter $\gamma + \frac{p}{2}$ and scale parameter matrix I , and U_4 is real matrix-variate gamma distributed with shape parameter $\gamma + \frac{p}{2}$ and scale parameter matrix B . Further $|I_p - A^{\frac{1}{2}}XBXA^{\frac{1}{2}}| = |I_q - B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}|$.*

Proof: All the results, except the last one, were obtained in Solution 7.2.2. Hence, we shall only consider the last part of the theorem. Observe that when $q > p$, $|A^{\frac{1}{2}}XBXA^{\frac{1}{2}}| > 0$, the matrix being positive definite, whereas $|B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}| = 0$, the matrix being positive semi-definite. The equality is established by noting that in accordance with results previously stated in Sect. 1.3, the determinant of the following partitioned matrix has two representations:

$$\begin{vmatrix} I_p & A^{\frac{1}{2}}XB^{\frac{1}{2}} \\ B^{\frac{1}{2}}X'A^{\frac{1}{2}} & I_q \end{vmatrix} = \begin{cases} |I_p| |I_q - (B^{\frac{1}{2}}X'A^{\frac{1}{2}})I_p^{-1}(A^{\frac{1}{2}}XB^{\frac{1}{2}})| = |I_q - B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}| \\ |I_q| |I_p - (A^{\frac{1}{2}}XB^{\frac{1}{2}})I_q^{-1}(B^{\frac{1}{2}}X'A^{\frac{1}{2}})| = |I_p - A^{\frac{1}{2}}XBXA^{\frac{1}{2}}| \end{cases} .$$

7.2.2. Multivariate gamma and Maxwell-Boltzmann densities, real case

Multivariate usually means a collection of scalar variables, real or complex. Many real scalar variable cases corresponding to (7.2.1) or a multivariate analogue of thereof can be obtained from (7.2.5) by taking $p = 1$ and $A = b > 0$. Note that in this case, X is $1 \times q$, that is, $X = (x_1, \dots, x_q)$, and XBX' is a positive definite quadratic form of the type

$$XBX' = (x_1, \dots, x_q)B \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} .$$

Thus, the density appearing in (7.2.5) becomes

$$f_8(X)dX = \frac{b^{\gamma+\frac{q}{2}}|B|^{\frac{1}{2}}\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}\Gamma(\gamma+\frac{q}{2})}[XBX']^\gamma e^{-b(XBX')}dX \quad (7.2.8)$$

for $X = (x_1, \dots, x_q)$, $-\infty < x_j < \infty$, $j = 1, \dots, q$, $B = B' > O$, $b > 0$, and $f_8 = 0$ elsewhere. Then, the density of $Y = B^{\frac{1}{2}}X'$ is given by

$$f_9(Y)dY = b^{\gamma+\frac{q}{2}} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}\Gamma(\gamma+\frac{q}{2})}(y_1^2 + \dots + y_q^2)^\gamma e^{-b(y_1^2 + \dots + y_q^2)}dY \quad (7.2.9)$$

where $Y' = (y_1, \dots, y_q)$, $-\infty < y_j < \infty$, $j = 1, \dots, q$, $b > 0$, $\gamma + \frac{q}{2} > 0$, and $f_9 = 0$ elsewhere. We will take (7.2.8) as the multivariate gamma as well as multivariate Maxwell-Boltzmann density, and (7.2.9) as the standard multivariate gamma as well as standard multivariate Maxwell-Boltzmann density.

How can we show that (7.2.9) is a statistical density? One way consists of writing $f_9(Y)dY$ as $f_9(S)dS$, applying Theorem 4.2.3 of Chap. 4 and writing dY in terms of dS for $p = 1$. This will yield the result. Another way is to integrate out variables y_1, \dots, y_q from $f_9(Y)dY$, which can be achieved via a general polar coordinate transformation such as the following: Consider the variables y_1, \dots, y_q , $-\infty < y_j < \infty$, $j = 1, \dots, q$, and the transformation,

$$\begin{aligned} y_1 &= r \sin \theta_1 \\ y_j &= r \cos \theta_1 \cos \theta_2 \cdots \cos \theta_{j-1} \sin \theta_j, \quad j = 2, 3, \dots, q-1, \\ y_q &= r \cos \theta_1 \cos \theta_2 \cdots \cos \theta_{q-1}, \end{aligned}$$

for $-\frac{\pi}{2} < \theta_j \leq \frac{\pi}{2}$, $j = 1, \dots, q-2$; $-\pi < \theta_{q-1} \leq \pi$, which was discussed in Mathai (1997). Its Jacobian is then given by

$$dy_1 \wedge \dots \wedge dy_q = r^{q-1} \left\{ \prod_{j=1}^{q-1} |\cos \theta_j|^{q-j-1} \right\} dr \wedge d\theta_1 \wedge \dots \wedge d\theta_{q-1}. \quad (7.2.10)$$

Under this transformation, $y_1^2 + \dots + y_q^2 = r^2$. Hence, integrating over r , we have

$$\int_{r=0}^{\infty} (r^2)^\gamma r^{q-1} e^{-br^2} dr = \frac{1}{2} b^{-(\gamma+\frac{q}{2})} \Gamma(\gamma+\frac{q}{2}), \quad \gamma+\frac{q}{2} > 0. \quad (7.2.11)$$

Note that the θ_j 's are present only in the Jacobian elements. There are formulae giving the integral over each differential element. We will integrate the θ_j 's one by one. Integrating over θ_1 gives

$$\begin{aligned} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_1)^{q-2} d\theta_1 &= 2 \int_0^{\frac{\pi}{2}} (\cos \theta_1)^{q-2} d\theta_1 = 2 \int_0^1 z^{q-2} (1-z^2)^{-\frac{1}{2}} dz \\ &= \frac{\Gamma(\frac{1}{2})\Gamma(\frac{q-1}{2})}{\Gamma(\frac{q}{2})}, \quad q > 1. \end{aligned}$$

The integrals over $\theta_2, \theta_3, \dots, \theta_{q-2}$ can be similarly evaluated as

$$\frac{\Gamma(\frac{1}{2})\Gamma(\frac{q-2}{2})}{\Gamma(\frac{q-1}{2})}, \frac{\Gamma(\frac{1}{2})\Gamma(\frac{q-3}{2})}{\Gamma(\frac{q-2}{2})}, \dots, \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{2})}$$

for $q > p-1$, the last integral $\int_{-\pi}^{\pi} d\theta_{q-1}$ giving 2π . On taking the product, several gamma functions cancel out, leaving

$$\frac{2\pi [\Gamma(\frac{1}{2})]^{q-2}}{\Gamma(\frac{q}{2})} = \frac{2\pi^{\frac{q}{2}}}{\Gamma(\frac{q}{2})}. \quad (7.2.12)$$

It follows from (7.2.11) and (7.2.12) that (7.2.9) is indeed a density which will be referred to as the standard real multivariate gamma or standard real Maxwell-Boltzmann density.

Example 7.2.3. Write down the densities specified in (7.2.8) and (7.2.9) explicitly if

$$B = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad b = 2 \text{ and } \gamma = 2.$$

Solution 7.2.3. Let us evaluate the normalizing constant in (7.2.8). Since in this case, $|B| = 2$,

$$c_8 = \frac{b^{\gamma+\frac{q}{2}} |B|^{\frac{1}{2}} \Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}} \Gamma(\gamma + \frac{q}{2})} = \frac{2^{2+\frac{3}{2}} 2^{\frac{1}{2}} \Gamma(\frac{3}{2})}{\pi^{\frac{3}{2}} \Gamma(2 + \frac{3}{2})} = \frac{2^6}{15\pi^{\frac{3}{2}}}. \quad (i)$$

The normalizing constant in (7.2.9) which will be denoted by c_9 , is the same as c_8 excluding $|B|^{\frac{1}{2}} = 2^{\frac{1}{2}}$. Thus,

$$c_9 = \frac{2^{\frac{11}{2}}}{15\pi^{\frac{3}{2}}}. \quad (ii)$$

Note that for $X = [x_1, x_2, x_3]$, $XBX' = 3x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 + 2x_2x_3$ and $YY' = y_1^2 + y_2^2 + y_3^2$. Hence the densities $f_8(X)$ and $f_9(Y)$ are the following, where c_8 and c_9 are given in (i) and (ii):

$$f_8(X) = c_8 \left[3x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 + 2x_2x_3 \right]^2 e^{-2[3x_1^2+2x_2^2+x_3^2-2x_1x_2+2x_2x_3]}$$

for $-\infty < x_j < \infty$, $j = 1, 2, 3$, and

$$f_9(Y) = c_9 \left[y_1^2 + y_2^2 + y_3^2 \right]^2 e^{-2[y_1^2 + y_2^2 + y_3^2]}, \quad \text{for } -\infty < y_j < \infty, \quad j = 1, 2, 3.$$

This completes the computations.

7.2.3. Some properties of the rectangular matrix-variate gamma density

For the real rectangular matrix-variate gamma and Maxwell-Boltzmann distribution whose density is specified in (7.2.5), what might be the h -th moment of the determinant $|AXBX'|$ for an arbitrary h ? This statistical quantity can be evaluated by looking at the normalizing constant c given in (7.2.4) since the integrand used to evaluate $E[|AXBX'|]^h$, where E denotes the expected value, is nothing but the density of X wherein γ is replaced by $\gamma + h$. Hence we have

$$E[|AXBX'|]^h = \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})}, \quad \Re(h) > -\gamma - \frac{q}{2} + \frac{p-1}{2}. \quad (7.2.13)$$

In many calculations involving the Maxwell-Boltzmann density for the real scalar variable case x , one has to integrate a function of x , say $v(x)$, over the Maxwell-Boltzmann density, as can be seen for example in equations (4.1) and (4.2) of Mathai and Haubold (1988) in connection with a certain reaction-rate probability integral. Thus, the expression appearing in (7.2.13) corresponds to the integral of a power function over the Maxwell-Boltzmann density.

This arbitrary h -th moment expression also reveals an interesting point. By expanding the matrix-variate gamma functions, we have the following:

$$\frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} = \prod_{j=1}^p \frac{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2})} = \prod_{j=1}^p E(t_j)^h$$

where t_j is a real scalar gamma random variable with parameter $(\gamma + \frac{q}{2} - \frac{j-1}{2}, 1)$, $j = 1, \dots, p$, whose density is

$$g_{(j)}(t_j) = \frac{1}{\Gamma(\gamma + \frac{q}{2} - \frac{(j-1)}{2})} t_j^{\gamma + \frac{q}{2} - \frac{(j-1)}{2} - 1} e^{-t_j}, \quad t_j \geq 0, \quad \gamma + \frac{q}{2} - \frac{(j-1)}{2} > 0, \quad (7.2.14)$$

and zero elsewhere. Thus structurally,

$$|AXBX'| = t_1 t_2 \cdots t_p \quad (7.2.15)$$

where t_1, \dots, t_p are independently distributed real scalar gamma random variables with t_j having the gamma density given in (7.2.14) for $j = 1, \dots, p$.

7.2.4. Connection to the volume of a random parallelotope

First, observe that $|AXBX'| = |(A^{\frac{1}{2}}XB^{\frac{1}{2}})(A^{\frac{1}{2}}XB^{\frac{1}{2}})'| \equiv |UU'|$ where $U = A^{\frac{1}{2}}XB^{\frac{1}{2}}$. Then, note that U is $p \times q$, $q \geq p$, and of full rank p , and that the p linearly independent rows of U , taken in the order, will then create a convex hull and a parallelotope in the q -dimensional Euclidean space. The p rows of U represent p linearly independent vectors in the Euclidean q -space as well as p points in the same space. In light of (7.2.14), these random points are gamma distributed, that is, the joint density of the p vectors or the p random points is the real rectangular matrix-variate density given in (7.2.5), and the volume content of the parallelotope created by these p random points is $|AXBX'|^{\frac{1}{2}}$. Accordingly, (7.2.13) represents the $(2h)$ -th moment of the random volume of the p -parallelotope generated by the p linearly independent rows of $A^{\frac{1}{2}}XB^{\frac{1}{2}}$. The geometrical probability problems considered in the literature usually pertain to random volumes generated by independently distributed isotropic random points, isotropic meaning that their associated density is invariant with respect to orthonormal transformations or rotations of the coordinate axes. For instance, the density given in (7.2.9) constitutes an example of isotropic form. The distributions of random geometrical configurations is further discussed in Chap. 4 of Mathai (1999).

7.2.5. Pathway to real matrix-variate gamma and Maxwell-Boltzmann densities

Consider a model of the following form for a $p \times q$, $q \geq p$, matrix X of full rank p :

$$f_{10}(X) = c_{10}|AXBX'|^\gamma |I - a(1 - \alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{-\frac{\eta}{1-\alpha}}, \alpha < 1, \tag{7.2.16}$$

for $A > O$, $B > O$, $a > 0$, $\eta > 0$, $I - a(1 - \alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}} > O$ (positive definite), and $f_{10}(X) = 0$ elsewhere. It will be determined later that the parameter γ is subject to the condition $\gamma + \frac{q}{2} > \frac{p-1}{2}$. When $\alpha > 1$, we let $1 - \alpha = -(\alpha - 1)$, $\alpha > 1$, so that the model specified in (7.2.16) shifts to the model

$$f_{11}(X) = c_{11}|AXBX'|^\gamma |I + a(\alpha - 1)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{-\frac{\eta}{\alpha-1}}, \alpha > 1 \tag{7.2.17}$$

for $\eta > 0$, $a > 0$, $A > O$, $B > O$, and $f_{11}(X) = 0$ elsewhere. Observe that $A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$ is symmetric as well as positive definite when X is of full rank p and $A > O$, $B > O$. For this model, the condition $\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} > \frac{p-1}{2}$ is required in addition to that applying to the parameter γ in (7.2.16). Note that when $f_{10}(X)$ and $f_{11}(X)$ are taken as statistical densities, c_{10} and c_{11} are the associated normalizing constants. Proceeding as in the evaluation of c in (7.2.4), we obtain the following representations for c_{10} and c_{11} :

$$c_{10} = |A|^{\frac{q}{2}} |B|^{\frac{p}{2}} [a(1 - \alpha)]^{p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}}} \frac{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2})}{\Gamma_p(\gamma + \frac{q}{2}) \Gamma_p(\frac{\eta}{1-\alpha} + \frac{p+1}{2})} \quad (7.2.18)$$

for $\eta > 0$, $\alpha < 1$, $a > 0$, $A > O$, $B > O$, $\gamma + \frac{q}{2} > \frac{p-1}{2}$,

$$c_{11} = \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}} [a(\alpha - 1)]^{p(\gamma + \frac{q}{2})} \Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}}} \frac{\Gamma_p(\frac{\eta}{\alpha-1})}{\Gamma_p(\gamma + \frac{q}{2}) \Gamma_p(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2})} \quad (7.2.19)$$

for $\alpha > 1$, $\eta > 0$, $a > 0$, $A > O$, $B > O$, $\gamma + \frac{q}{2} > \frac{p-1}{2}$, $\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} > \frac{p-1}{2}$. When $\alpha \rightarrow 1_-$ in (7.2.18) and $\alpha \rightarrow 1_+$ in (7.2.19), the models (7.2.16) and (7.2.17) converge to the real rectangular matrix-variate gamma or Maxwell-Boltzmann density specified in (7.2.5). This can be established by applying the following lemmas.

Lemma 7.2.1.

$$\lim_{\alpha \rightarrow 1_-} |I - a(1 - \alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{\frac{\eta}{1-\alpha}} = e^{-a\eta \text{tr}(AXBX')}$$

and

$$\lim_{\alpha \rightarrow 1_+} |I + a(\alpha - 1)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{-\frac{\eta}{\alpha-1}} = e^{-a\eta \text{tr}(AXBX')}. \quad (7.2.20)$$

Proof: Letting $\lambda_1, \dots, \lambda_p$ be the eigenvalues of the symmetric matrix $A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$, we have

$$|I - a(1 - \alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{\frac{\eta}{1-\alpha}} = \prod_{j=1}^p [1 - a(1 - \alpha)\lambda_j]^{\frac{\eta}{1-\alpha}}.$$

However, since

$$\lim_{\alpha \rightarrow 1_-} [1 - a(1 - \alpha)\lambda_j]^{\frac{\eta}{1-\alpha}} = e^{-a\eta\lambda_j},$$

the product gives the sum of the eigenvalues, that is, $\text{tr}(A^{\frac{1}{2}}XBX'A^{\frac{1}{2}})$ in the exponent, hence the result. The same result can be similarly obtained for the case $\alpha > 1$. We can also show that the normalizing constants c_{10} and c_{11} reduce to the normalizing constant in (7.2.4). This can be achieved by making use of an asymptotic expansion of gamma functions, namely,

$$\Gamma(z + \delta) \approx \sqrt{2\pi} z^{z+\delta-\frac{1}{2}} e^{-z} \text{ for } |z| \rightarrow \infty, \delta \text{ bounded.} \quad (7.2.21)$$

This first term approximation is also known as Stirling's formula.

Lemma 7.2.2.

$$\lim_{\alpha \rightarrow 1_-} [a(1 - \alpha)]^{p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2})}{\Gamma_p(\frac{\eta}{1-\alpha} + \frac{p+1}{2})} = (a\eta)^{p(\gamma + \frac{q}{2})}$$

and

$$\lim_{\alpha \rightarrow 1_+} [a(\alpha - 1)]^{p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\frac{\eta}{\alpha-1})}{\Gamma_p(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2})} = (a\eta)^{p(\gamma + \frac{q}{2})}. \tag{7.2.22}$$

Proof: On expanding $\Gamma_p(\cdot)$ using its definition, for $\alpha > 1$, we have

$$\frac{[a(\alpha - 1)]^{p(\gamma + \frac{q}{2})} \Gamma_p(\frac{\eta}{\alpha-1})}{\Gamma_p(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2})} = [a(\alpha - 1)]^{p(\gamma + \frac{q}{2})} \prod_{j=1}^p \frac{\Gamma(\frac{\eta}{\alpha-1} - \frac{j-1}{2})}{\Gamma(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - \frac{j-1}{2})}.$$

Now, on applying the Stirling’s formula as given in (7.2.21) to each of the gamma functions by taking $z = \frac{\eta}{\alpha-1} \rightarrow \infty$ when $\alpha \rightarrow 1_+$, it is seen that the right-hand side of the above equality reduces to $(a\eta)^{p(\gamma + \frac{q}{2})}$. The result can be similarly established for the case $\alpha < 1$.

This shows that c_{10} and c_{11} of (7.2.18) and (7.2.19) converge to the normalizing constant in (7.2.4). This means that the models specified in (7.2.16), (7.2.17), and (7.2.5) are all available from either (7.2.16) or (7.2.17) via the pathway parameter α . Accordingly, the combined model, either (7.2.16) or (7.2.17), is referred to as the pathway generalized real rectangular matrix-variate gamma density. The Maxwell-Boltzmann case corresponds to $\gamma = 1$ and the Raleigh case, to $\gamma = \frac{1}{2}$. If either of the Maxwell-Boltzmann or Raleigh densities is the ideal or stable density in a physical system, then these stable densities as well as the unstable neighborhoods, described through the pathway parameter $\alpha < 1$ and $\alpha > 1$, and the transitional stages, are given by (7.2.16) or (7.2.17). The original pathway model was introduced in Mathai (2005).

For addressing other problems occurring in physical situations, one may have to integrate functions of X over the densities (7.2.16), (7.2.17) or (7.2.5). Consequently, we will evaluate an arbitrary h -th moment of $|AXBX'|$ in the models (7.2.16) and (7.2.17). For example, let us determine the h -th moment of $|AXBX'|$ with respect to the model specified in (7.2.16):

$$E[|AXBX'|^h] = c_{10} \int_X |AXBX'|^{\gamma+h} |I - a(1 - \alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{\frac{\eta}{1-\alpha}} dX.$$

Note that the only change in the integrand, as compared to (7.2.16), is that γ is replaced by $\gamma + h$. Hence the result is available from the normalizing constant c_{10} , and the answer is the following:

$$E[|AXBX'|^h] = [a(1 - \alpha)]^{-ph} \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2})}{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2} + h)} \quad (7.2.23)$$

for $\Re(\gamma + \frac{q}{2} + h) > \frac{p-1}{2}$, $a > 0$, $\alpha < 1$. Therefore

$$\begin{aligned} E[|a(1 - \alpha)AXBX'|^h] &= \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2})}{\Gamma_p(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2} + h)} \\ &= \prod_{j=1}^p \left\{ \frac{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2})} \frac{\Gamma(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2} - \frac{j-1}{2})}{\Gamma(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + \frac{p+1}{2} - \frac{j-1}{2} + h)} \right\} \\ &= \prod_{j=1}^p E(y_j^h) \end{aligned} \quad (7.2.24)$$

where y_j is a real scalar type-1 beta random variable with the parameters $(\gamma + \frac{q}{2} - \frac{j-1}{2}, \frac{\eta}{1-\alpha} + \frac{p+1}{2})$, $j = 1, \dots, p$, the y_j 's being mutually independently distributed. Hence, we have the structural relationship

$$|a(1 - \alpha)AXBX'| = y_1 \cdots y_p. \quad (7.2.25)$$

Proceeding the same way for the model (7.2.17), we have

$$E[|AXBX'|^h] = [a(\alpha - 1)]^{-ph} \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - h)}{\Gamma_p(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2})} \quad (7.2.26)$$

for $\Re(\gamma + \frac{q}{2} + h) > \frac{p-1}{2}$, $\Re(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - h) > \frac{p-1}{2}$ or $-(\gamma + \frac{q}{2}) + \frac{p-1}{2} < \Re(h) < \frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - \frac{p-1}{2}$. Thus,

$$\begin{aligned} E[|a(\alpha - 1)AXBX'|^h] &= \prod_{j=1}^p \frac{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2})} \frac{\Gamma(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - \frac{j-1}{2} - h)}{\Gamma(\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - \frac{j-1}{2})} \\ &= \prod_{j=1}^p E(z_j^h) \end{aligned} \quad (7.2.27)$$

where z_j is a real scalar type-2 beta random variable with the parameters $(\gamma + \frac{q}{2} - \frac{j-1}{2}, \frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} - \frac{j-1}{2})$ for $j = 1, \dots, p$, the z_j 's being mutually independently distributed. Thus, for $\alpha > 1$, we have the structural representation

$$|a(\alpha - 1)AXBX'| = z_1 \cdots z_p. \tag{7.2.28}$$

As previously explained, one can consider the p linearly independent rows of $A^{\frac{1}{2}}XB^{\frac{1}{2}}$ as p vectors in the Euclidean q -space. Then, these p vectors are jointly distributed as rectangular matrix-variate type-2 beta, and $E[|AXBX'|^h] = E[|AXBX'|^{\frac{1}{2}}]^{2h}$ is the $(2h)$ -th moment of the volume of the random parallelotope generated by these p q -vectors for $q > p$. In this case, the random points will be called type-2 beta distributed random points.

The real Maxwell-Boltzmann case will correspond $\gamma = 1$ and the Raleigh case, to $\gamma = \frac{1}{2}$, and all the above extensions and properties will apply to both of these distributions.

7.2.6. Multivariate gamma and Maxwell-Boltzmann densities, pathway model

Consider the density given in (7.2.16) for the case $p = 1$. In this instance, the $p \times p$ constant matrix A is 1×1 and we shall let $A = b > 0$, a positive real scalar quantity. Then for $\alpha < 1$, (7.2.16) reduces to the following where X is $1 \times q$ of the form $X = (x_1, \dots, x_q)$, $-\infty < x_j < \infty$, $j = 1, \dots, q$:

$$f_{12}(X) = b^{\frac{q}{2}}|B|^{\frac{1}{2}}[a(1 - \alpha)]^{(\gamma+\frac{q}{2})} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + 1)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\frac{\eta}{1-\alpha} + 1)} \times [bXBX']^\gamma [1 - a(1 - \alpha)bXBX']^{\frac{\eta}{1-\alpha}} \tag{7.2.29}$$

for $b > 0$, $B = B' > O$, $a > 0$, $\eta > 0$, $\gamma + \frac{q}{2} > 0$, $-\infty < x_j < \infty$, $j = 1, \dots, q$, $1 - a(1 - \alpha)bXBX' > 0$, $\alpha < 1$, and $f_{12} = 0$ elsewhere. Note that

$$XBX' = (x_1, \dots, x_q)B \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}$$

is a real quadratic form whose associated matrix B is positive definite. Letting the $1 \times q$ vector $Y = XB^{\frac{1}{2}}$, the density of Y when $\alpha < 1$ is given by

$$f_{13}(Y) dY = b^{\gamma+\frac{q}{2}}[a(1 - \alpha)]^{(\gamma+\frac{q}{2})} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\gamma + \frac{q}{2} + \frac{\eta}{1-\alpha} + 1)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\frac{\eta}{1-\alpha} + 1)} \times [(y_1^2 + \cdots + y_q^2)]^\gamma [1 - a(1 - \alpha)b(y_1^2 + \cdots + y_q^2)]^{\frac{\eta}{1-\alpha}} dY, \tag{7.2.30}$$

for $b > 0$, $\gamma + \frac{q}{2} > 0$, $\eta > 0$, $-\infty < y_j < \infty$, $j = 1, \dots, q$, $1 - a(1 - \alpha)b(y_1^2 + \dots + y_q^2) > 0$, and $f_{13} = 0$ elsewhere, which will be taken as the standard form of the real multivariate gamma density in its pathway generalized form, and for $\gamma = 1$, it will be the real pathway generalized form of the Maxwell-Boltzmann density in the standard multivariate case. For $\alpha > 1$, the corresponding standard form of the real multivariate gamma and Maxwell-Boltzmann densities is given by

$$f_{14}(Y)dY = b^{\gamma + \frac{q}{2}} [a(\alpha - 1)]^{(\gamma + \frac{q}{2})} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\frac{\eta}{\alpha - 1})}{\Gamma(\gamma + \frac{q}{2})\Gamma(\frac{\eta}{\alpha - 1} - \gamma - \frac{q}{2})} \\ \times [(y_1^2 + \dots + y_q^2)]^\gamma [1 + a(\alpha - 1)b(y_1^2 + \dots + y_q^2)]^{-\frac{\eta}{\alpha - 1}} dY. \quad (7.2.31)$$

for $b > 0$, $\gamma + \frac{q}{2} > 0$, $a > 0$, $\eta > 0$, $\frac{\eta}{\alpha - 1} - \gamma - \frac{q}{2} > 0$, $-\infty < y_j < \infty$, $j = 1, \dots, q$, and $f_{14} = 0$ elsewhere. This will be taken as the pathway generalized real multivariate gamma density for $\alpha > 1$, and for $\gamma = 1$, it will be the standard form of the real pathway extended Maxwell-Boltzmann density for $\alpha > 1$. Note that when $\alpha \rightarrow 1_-$ in (7.2.30) and $\alpha \rightarrow 1_+$ in (7.2.31), we have

$$f_{15}(Y)dY = b^{\gamma + \frac{q}{2}} (a\eta)^{(\gamma + \frac{q}{2})} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}} \Gamma(\gamma + \frac{q}{2})} \\ \times [(y_1^2 + \dots + y_q^2)]^\gamma e^{-a\eta b(y_1^2 + \dots + y_q^2)} dY, \quad (7.2.32)$$

for $b > 0$, $a > 0$, $\eta > 0$, $\gamma + \frac{q}{2} > 0$, and $f_{15} = 0$ elsewhere, which for $\gamma = 1$, is the real multivariate Maxwell-Boltzmann density in the standard form. From (7.2.30), (7.2.31), and thereby from (7.2.32), one can obtain the density of $u = y_1^2 + \dots + y_q^2$, either by using the general polar coordinate transformation or the transformation of variables technique, that is, going from dY to dS with $S = YY'$, Y being $1 \times p$. Then, the density of u for the case $\alpha < 1$ is

$$f_{16}(u) = b^{\gamma + \frac{q}{2}} [a(1 - \alpha)]^{\gamma + \frac{q}{2}} \frac{\Gamma(\gamma + \frac{q}{2} + \frac{\eta}{1 - \alpha} + 1)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\frac{\eta}{1 - \alpha} + 1)} u^{\gamma + \frac{q}{2} - 1} [1 - a(1 - \alpha)bu]^{-\frac{\eta}{1 - \alpha}}, \quad \alpha < 1, \quad (7.2.33)$$

for $b > 0$, $a > 0$, $\eta > 0$, $\alpha < 1$, $\gamma + \frac{q}{2} > 0$, $1 - a(1 - \alpha)bu > 0$, and $f_{16} = 0$ elsewhere, the density of u for $\alpha > 1$ being

$$f_{17}(u) = b^{\gamma + \frac{q}{2}} [a(\alpha - 1)]^{\gamma + \frac{q}{2}} \frac{\Gamma(\frac{\eta}{\alpha - 1})}{\Gamma(\gamma + \frac{q}{2})\Gamma(\frac{\eta}{\alpha - 1} - \gamma - \frac{q}{2})} u^{\gamma + \frac{q}{2} - 1} [1 + a(\alpha - 1)bu]^{-\frac{\eta}{\alpha - 1}}, \quad \alpha > 1, \quad (7.2.34)$$

for $b > 0$, $a > 0$, $\eta > 0$, $\gamma + \frac{q}{2} > 0$, $\frac{\eta}{\alpha-1} - \gamma - \frac{q}{2} > 0$, $u \geq 0$, and $f_{17} = 0$ elsewhere. Observe that as $\alpha \rightarrow 1$, both (7.2.33) and (7.2.34) converge to the form

$$f_{18}(u) = \frac{(a\eta b)^{\gamma + \frac{q}{2}}}{\Gamma(\gamma + \frac{q}{2})} u^{\gamma + \frac{q}{2} - 1} e^{-ab\eta u} \tag{7.2.35}$$

for $a > 0$, $b > 0$, $\eta > 0$, $u \geq 0$, and $f_{18} = 0$ elsewhere. For $\gamma = \frac{1}{2}$, we have the corresponding Raleigh cases.

Letting $\gamma = 1$ and $q = 1$ in (7.2.32), we have

$$\begin{aligned} f_{19}(y_1) &= \frac{b^{\frac{3}{2}}}{\Gamma(\gamma + \frac{1}{2})} (y_1^2)^\gamma e^{-by_1^2} = \frac{2b^{\frac{3}{2}}}{\sqrt{\pi}} y_1^2 e^{-by_1^2}, \quad -\infty < y_1 < \infty, \quad b > 0 \\ &= \frac{4b^{\frac{3}{2}}}{\sqrt{\pi}} y_1^2 e^{-by_1^2}, \quad 0 \leq y_1 < \infty, \quad b > 0, \end{aligned} \tag{7.2.36}$$

and $f_{19} = 0$ elsewhere. This is the real Maxwell-Boltzmann case. For the Raleigh case, we let $\gamma = \frac{1}{2}$ and $p = 1$, $q = 1$ in (7.2.32), which results in the following density:

$$\begin{aligned} f_{20}(y_1) &= b(y_1^2)^{\frac{1}{2}} e^{-by_1^2}, \quad -\infty < y_1 < \infty, \quad b > 0 \\ &= 2b|y_1| e^{-by_1^2}, \quad 0 \leq y_1 < \infty, \quad b > 0, \end{aligned} \tag{7.2.37}$$

and $f_{20} = 0$ elsewhere.

7.2.7. Concluding remarks

There exist natural phenomena that are suspected to involve an underlying distribution which is not Maxwell-Boltzmann but may be some deviation therefrom. In such instances, it is preferable to model the collected data by means of the pathway extended model previously specified for $p = 1$, $q = 1$ (real scalar case), $p = 1$ (real multivariate case) and the general matrix-variate case. The pathway parameter α will capture the Maxwell-Boltzmann case, the neighboring models described by the pathway model for $\alpha < 1$ and for $\alpha > 1$ and the transitional stages when moving from one family of functions to another, and thus, to all three different families of functions. Incidentally, for $\gamma = 0$, one has the rectangular matrix-variate Gaussian density given in (7.2.5) and its pathway extension in (7.2.16) and (7.2.17) or the general extensions in the standard forms in (7.2.30), (7.2.31), and (7.2.32) wherein $\gamma = 0$. The structures in (7.2.24), (7.2.27), and (7.2.28) suggest that the corresponding densities can also be written in terms of G- and H-functions. For the theory and applications of the G- and H-functions, the reader is referred to Mathai (1993)

and Mathai et al. (2010), respectively. The complex analogues of some matrix-variate distributions, including the matrix-variate Gaussian, were introduced in Mathai and Provost (2006). Certain bivariate distributions are discussed in Balakrishnan and Lai (2009) and some general method of generating real multivariate distributions are presented in Marshall and Olkin (1967).

Example 7.2.4. Let $X = (x_{ij})$ be a real $p \times q$, $q \geq p$, matrix of rank p , where the x_{ij} 's are distinct real scalar variables. Let the constant matrices $A = b > 0$ be 1×1 and $B > O$ be $q \times q$. Consider the following generalized multivariate Maxwell-Boltzmann density

$$f(X) = c |AXBX'|^\gamma e^{-[\text{tr}(AXBX')]^\delta}$$

for $\delta > 0$, $A = b > 0$, $X = [x_1, \dots, x_q]$. Evaluate c if $f(X)$ is a density.

Solution 7.2.4. Since X is $1 \times q$, $|AXBX'| = b[XX']$ where XX' is a real quadratic form. For $f(X)$ to be a density, we must have

$$1 = \int_X f(X) dX = c b^\gamma \int_X [XX']^\gamma e^{-[bXX']^\delta} dX. \quad (i)$$

Let us make the transformations $Y = XB^{\frac{1}{2}}$ and $s = YY'$. Then (i) reduces to the following:

$$1 = c b^\gamma \frac{\pi^{\frac{q}{2}}}{\Gamma(\frac{q}{2})} |B|^{-\frac{1}{2}} \int_0^\infty s^{\gamma + \frac{q}{2} - 1} e^{-(bs)^\delta} ds. \quad (ii)$$

Letting $t = b^\delta s^\delta$, $b > 0$, $s > 0 \Rightarrow ds = \frac{1}{\delta} \frac{t^{\frac{1}{\delta} - 1}}{b} dt$, (ii) becomes

$$\begin{aligned} 1 &= c \frac{\pi^{\frac{q}{2}}}{\delta b^{\frac{q}{2}} \Gamma(\frac{q}{2}) |B|^{\frac{1}{2}}} \int_0^\infty t^{\frac{\gamma}{\delta} + \frac{q}{2\delta} - 1} e^{-t} dt \\ &= c \frac{\pi^{\frac{q}{2}} \Gamma(\frac{\gamma}{\delta} + \frac{q}{2\delta})}{\delta b^{\frac{q}{2}} \Gamma(\frac{q}{2}) |B|^{\frac{1}{2}}}. \end{aligned}$$

Hence,

$$c = \frac{\delta b^{\frac{q}{2}} \Gamma(\frac{q}{2}) |B|^{\frac{1}{2}}}{\pi^{\frac{q}{2}} \Gamma(\frac{\gamma}{\delta} + \frac{q}{2\delta})}.$$

No additional conditions are required other than $\gamma > 0$, $\delta > 0$, $q > 0$, $B > O$. This completes the solution.

7.2a. Complex Matrix-Variate Gamma and Maxwell-Boltzmann Densities

The matrix-variate gamma density in the real positive definite matrix case was defined in equation (5.2.4) of Sect. 5.2. The corresponding matrix-variate gamma density in the complex domain was given in Eq. (5.2a.4). Those distributions will be extended to the rectangular matrix-variate cases in this section. A particular case of the rectangular matrix-variate gamma in the complex domain will be called the Maxwell-Boltzmann density in the complex matrix-variate case. Let $\tilde{X} = (\tilde{x}_{ij})$ be a $p \times q$, $q \geq p$, rectangular matrix of rank p in the complex domain whose elements \tilde{x}_{ij} are distinct scalar complex variables. Let $|\det(\cdot)|$ denote the absolute value of the determinant of (\cdot) . Let A of order $p \times p$ and B of order $q \times q$ be real positive definite or Hermitian positive definite constant matrices. The conjugate transpose of \tilde{X} will be denoted by \tilde{X}^* . Consider the function:

$$\tilde{f}(\tilde{X})d\tilde{X} = \tilde{c} |\det(A\tilde{X}B\tilde{X}^*)|^\gamma e^{-\text{tr}(A\tilde{X}B\tilde{X}^*)} d\tilde{X} \quad (7.2a.1)$$

for $A > O$, $B > O$, $\Re(\gamma + q) > p - 1$ where \tilde{c} is the normalizing constant so that $\tilde{f}(\tilde{X})$ is a statistical density. One can evaluate \tilde{c} by proceeding as was done in the real case. Let

$$\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}} \Rightarrow d\tilde{Y} = |\det(A)|^q |\det(B)|^p d\tilde{X},$$

the Jacobian of this matrix transformation being provided in Chap. 1 or Mathai (1997). Then, $\tilde{f}(\tilde{X})$ becomes

$$\tilde{f}_1(\tilde{Y}) d\tilde{Y} = \tilde{c} |\det(A)|^{-q} |\det(B)|^{-p} |\det(\tilde{Y}\tilde{Y}^*)|^\gamma e^{-\text{tr}(\tilde{Y}\tilde{Y}^*)} d\tilde{Y}. \quad (7.2a.2)$$

Now, letting

$$\tilde{S} = \tilde{Y}\tilde{Y}^* \Rightarrow d\tilde{Y} = \frac{\pi^{qp}}{\tilde{\Gamma}_p(q)} |\det(\tilde{S})|^{q-p} d\tilde{S}$$

by applying Result 4.2a.3 where $\tilde{\Gamma}_p(q)$ is the complex matrix-variate gamma function, \tilde{f}_1 changes to

$$\begin{aligned} \tilde{f}_2(\tilde{S}) d\tilde{S} &= \tilde{c} |\det(A)|^{-q} |\det(B)|^{-p} \frac{\pi^{qp}}{\tilde{\Gamma}_p(q)} \\ &\times |\det(\tilde{S})|^{\gamma+q-p} e^{-\text{tr}(\tilde{S})} d\tilde{S}. \end{aligned} \quad (7.2a.3)$$

Finally, integrating out \tilde{S} by making use of a complex matrix-variate gamma integral, we have $\tilde{\Gamma}_p(\gamma + q)$ for $\Re(\gamma + q) > p - 1$. Hence, the normalizing constant \tilde{c} is the following:

$$\tilde{c} = \frac{|\det(A)|^q |\det(B)|^p}{\pi^{qp}} \frac{\tilde{\Gamma}_p(q)}{\tilde{\Gamma}_p(\gamma + q)}, \quad \Re(\gamma + q) > p - 1, \quad A > O, \quad B > O. \quad (7.2a.4)$$

Example 7.2a.1. Evaluate the normalizing constant in the density in (7.2a.1) if $\gamma = 2$, $q = 3$, $p = 2$,

$$A = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 0 & i \\ 0 & 2 & 1+i \\ -i & 1-i & 2 \end{bmatrix}.$$

Solution 7.2a.1. Note that A and B are both Hermitian matrices since $A = A^*$ and $B = B^*$. The leading minors of A are $|(3)| = 3 > 0$, $\begin{vmatrix} 3 & 1+i \\ 1-i & 2 \end{vmatrix} = (3)(2) - (1+i)(1-i) = 4 > 0$ and hence, $A > O$ (positive definite). The leading minors of B are $|(3)| = 3 > 0$, $\begin{vmatrix} 3 & 0 \\ 0 & 2 \end{vmatrix} = 6 > 0$, $|B| = 3 \begin{vmatrix} 2 & 1+i \\ 1-i & 2 \end{vmatrix} + 0 + i \begin{vmatrix} 0 & 2 \\ -i & 1-i \end{vmatrix} = 3(4-2) + i(2i) = 4 > 0$. Hence $B > O$ and $|B| = 4$. The normalizing constant

$$\begin{aligned} \tilde{c} &= \frac{|\det(A)|^q |\det(B)|^p}{\pi^{pq}} \frac{\tilde{\Gamma}_p(q)}{\tilde{\Gamma}_p(\gamma + q)} \\ &= \frac{(4)^3 (4)^2 \tilde{\Gamma}_2(3)}{\pi^6 \tilde{\Gamma}_2(5)} = \frac{4^5 \pi \Gamma(3) \Gamma(2)}{\pi^6 \pi \Gamma(5) \Gamma(4)} \\ &= \frac{2^7}{3^2 \pi^6}. \end{aligned}$$

This completes the computations.

7.2a.1. Extension of the Matrix-Variate Gamma Density in the Complex Domain

Consider the density of \tilde{X} is given in (7.2a.1) with \tilde{c} given in (7.2a.4). The density of $\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}}$ is given by

$$f_1(\tilde{Y}) = \frac{\tilde{\Gamma}_p(q)}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q)} |\det(\tilde{Y} \tilde{Y}^*)|^\gamma e^{-\text{tr}(\tilde{Y} \tilde{Y}^*)} \quad (7.2a.5)$$

for $\Re(\gamma + q) > p - 1$, and $f_1 = 0$ elsewhere, and the density of $\tilde{S} = \tilde{Y} \tilde{Y}^*$ is

$$\tilde{f}_2(\tilde{S}) = \frac{1}{\tilde{\Gamma}_p(\gamma + q)} |\det(\tilde{S})|^{\gamma+q-p} e^{-\text{tr}(\tilde{S})}, \Re(\gamma + q) > p - 1, \quad (7.2a.6)$$

and $\tilde{f}_2 = 0$ elsewhere. Then, the density given in (7.2a.1), namely, $\tilde{f}(\tilde{X})$ for $\gamma = 1$ will be called the Maxwell-Boltzmann density for the complex rectangular matrix-variate case since for $p = 1$ and $q = 1$ in the real scalar case, the density corresponds to the case $\gamma = 1$, and (7.2a.1) for $\gamma = \frac{1}{2}$ will be called complex rectangular matrix-variate Raleigh density.

7.2a.2. The multivariate gamma density in the complex matrix-variate case

Consider the case $p = 1$ and $A = b > 0$ where b is a real positive scalar as the $p \times p$ matrix A is assumed to be Hermitian positive definite. Then, \tilde{X} is $1 \times q$ and

$$A\tilde{X}B\tilde{X}^* = b\tilde{X}B\tilde{X}^* = b(\tilde{x}_1, \dots, \tilde{x}_q)B \begin{pmatrix} \tilde{x}_1^* \\ \vdots \\ \tilde{x}_q^* \end{pmatrix}$$

is a positive definite Hermitian form, an asterisk denoting only the conjugate when the elements are scalar quantities. Thus, when $p = 1$ and $A = b > 0$, the density $\tilde{f}(\tilde{X})$ reduces to

$$\tilde{f}_3(\tilde{X}) = b^{\gamma+q} |\det(B)| \frac{\tilde{\Gamma}(q)}{\pi^q \tilde{\Gamma}(\gamma + q)} |\det(\tilde{X}B\tilde{X}^*)|^\gamma e^{-b(\tilde{X}B\tilde{X}^*)} \quad (7.2a.7)$$

for $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_q)$, $B = B^* > O$, $b > 0$, $\Re(\gamma + q) > 0$, and $\tilde{f}_3 = 0$ elsewhere. Letting $\tilde{Y}^* = B^{\frac{1}{2}} \tilde{X}^*$, the density of \tilde{Y} is the following:

$$\tilde{f}_4(\tilde{Y}) = \frac{b^{\gamma+q} \tilde{\Gamma}(q)}{\pi^q \tilde{\Gamma}(\gamma + q)} (|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2)^\gamma e^{-b(|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2)} \quad (7.2a.8)$$

for $b > 0$, $\Re(\gamma + q) > 0$, and $\tilde{f}_4 = 0$ elsewhere, where $|\tilde{y}_j|$ is the absolute value or modulus of the complex quantity \tilde{y}_j . We will take (7.2a.8) as the complex multivariate gamma density; when $\gamma = 1$, it will be referred to as the complex multivariate Maxwell-Boltzmann density, and when $\gamma = \frac{1}{2}$, it will be called complex multivariate Raleigh density. These densities are believed to be new.

Let us verify by integration that (7.2a.8) is indeed a density. First, consider the transformation $s = \tilde{Y}\tilde{Y}^*$. In view of Theorem 4.2a.3, the integral over the Stiefel manifold gives $d\tilde{Y} = \frac{\pi^q}{\tilde{\Gamma}(q)} \tilde{s}^{q-1} d\tilde{s}$, so that $\frac{\pi^q}{\tilde{\Gamma}(q)}$ is canceled. Then, the integral over \tilde{s} yields

$b^{-(\gamma+q)} \tilde{\Gamma}(\gamma + q)$, $\Re(\gamma + q) > 0$, and hence it is verified that (7.2a.8) is a statistical density.

7.2a.3. Arbitrary moments, complex case

Let us determine the h -th moment of $u = |\det(A\tilde{X}B\tilde{X}^*)|$ for an arbitrary h , that is, the h -th moment of the absolute value of the determinant of the matrix $A\tilde{X}B\tilde{X}^*$ or its symmetric form $A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}}$, which is

$$E[|\det(A\tilde{X}B\tilde{X}^*)|^h] = \tilde{c} \int_{\tilde{X}} |\det(A\tilde{X}B\tilde{X}^*)|^{h+\gamma} e^{-\text{tr}(A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})} d\tilde{X}. \quad (7.2a.9)$$

Observe that the only change, as compared to the total integral, is that γ is replaced by $\gamma + h$, so that the h -th moment is available from the normalizing constant \tilde{c} . Accordingly,

$$E[u^h] = \frac{\tilde{\Gamma}_p(\gamma + q + h)}{\tilde{\Gamma}_p(\gamma + q)}, \quad \Re(\gamma + q + h) > p - 1, \quad (7.2a.10)$$

$$\begin{aligned} &= \prod_{j=1}^p \frac{\Gamma(\gamma + q + h - (j - 1))}{\Gamma(\gamma + q - (j - 1))} \\ &= E(u_1^h) E(u_2^h) \cdots E(u_p^h) \end{aligned} \quad (7.2a.11)$$

where the u_j 's are independently distributed real scalar gamma random variables with parameters $(\gamma + q - (j - 1), 1)$, $j = 1, \dots, p$. Thus, structurally $u = |\det(A\tilde{X}B\tilde{X}^*)|$ is a product of independently distributed real scalar gamma random variables with parameters $(\gamma + q - (j - 1), 1)$, $j = 1, \dots, p$. The corresponding result in the real case is that $|AXBX'|$ is structurally a product of independently distributed real gamma random variables with parameters $(\gamma + \frac{q}{2} - \frac{j-1}{2}, 1)$, $j = 1, \dots, p$, which can be seen from (7.2.15).

7.2a.4. A pathway extension in the complex case

A pathway extension is also possible in the complex case. The results and properties are parallel to those obtained in the real case. Hence, we will only mention the pathway extended density. Consider the following density:

$$\tilde{f}_5(\tilde{X}) = \tilde{c}_1 |\det(A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})|^\gamma |\det(I - a(1 - \alpha)A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})|^{\frac{\eta}{1-\alpha}} \quad (7.2a.12)$$

for $a > 0$, $\alpha < 1$, $I - a(1 - \alpha)A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}} > O$ (positive definite), $A > O$, $B > O$, $\eta > 0$, $\Re(\gamma + q) > p - 1$, and $\tilde{f}_5 = 0$ elsewhere. Observe that (7.2a.12) remains in the generalized type-1 beta family of functions for $\alpha < 1$ (type-1 and type-2 beta densities

in the complex rectangular matrix-variate cases will be considered in the next sections). If $\alpha > 1$, then on writing $1 - \alpha = -(\alpha - 1)$, $\alpha > 1$, the model in (7.2a.12) shifts to the model

$$\tilde{f}_6(\tilde{X}) = \tilde{c}_2 |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^\gamma |\det(I + a(\alpha - 1)A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^{-\frac{\eta}{\alpha-1}} \quad (7.2a.13)$$

for $a > 0$, $\alpha > 1$, $A > O$, $B > O$, $A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}} > O$, $\eta > 0$, $\Re(\frac{\eta}{\alpha-1} - \gamma - q) > p - 1$, $\Re(\gamma + q) > p - 1$, and $\tilde{f}_6 = 0$ elsewhere, where \tilde{c}_2 is the normalizing constant, different from \tilde{c}_1 . When $\alpha \rightarrow 1$, both models (7.2a.12) and (7.2a.13) converge to the model \tilde{f}_7 where

$$\tilde{f}_7(\tilde{X}) = \tilde{c}_3 |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^\gamma e^{-a \eta \operatorname{tr}(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})} \quad (7.2a.14)$$

for $a > 0$, $\eta > 0$, $A > O$, $B > O$, $\Re(\gamma + q) > p - 1$, $A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}} > O$, and $\tilde{f}_7 = 0$ elsewhere. The normalizing constants can be evaluated by following steps parallel to those used in the real case. They respectively are:

$$\tilde{c}_1 = |\det(A)|^q |\det(B)|^p [a(1 - \alpha)]^{p(\gamma+q)} \frac{\tilde{\Gamma}_p(q) \tilde{\Gamma}_p(\gamma + q + \frac{\eta}{1-\alpha} + p)}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\frac{\eta}{1-\alpha} + p)} \quad (7.2a.15)$$

for $\eta > 0$, $a > 0$, $\alpha < 1$, $A > O$, $B > O$, $\Re(\gamma + q) > p - 1$;

$$\tilde{c}_2 = |\det(A)|^q |\det(B)|^p [a(\alpha - 1)]^{p(\gamma+q)} \frac{\tilde{\Gamma}_p(q) \tilde{\Gamma}_p(\frac{\eta}{\alpha-1})}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\frac{\eta}{\alpha-1} - \gamma - q)} \quad (7.2a.16)$$

for $a > 0$, $\alpha > 1$, $\eta > 0$, $A > O$, $B > O$, $\Re(\gamma + q) > p - 1$, $\Re(\frac{\eta}{\alpha-1} - \gamma - q) > p - 1$;

$$\tilde{c}_3 = (a\eta)^{p(\gamma+q)} |\det(A)|^q |\det(B)|^p \frac{\tilde{\Gamma}_p(q)}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q)} \quad (7.2a.17)$$

for $a > 0$, $\eta > 0$, $A > O$, $B > O$, $\Re(\gamma + q) > p - 1$.

7.2a.5. The Maxwell-Boltzmann and Raleigh cases in the complex domain

The complex counterparts of the Maxwell-Boltzmann and Raleigh cases may not be available in the literature. Their densities can be derived from (7.2a.8). Letting $p = 1$ and $q = 1$ in (7.2a.8), we have

$$\begin{aligned} \tilde{f}_8(\tilde{y}_1) &= \frac{b^{\gamma+1}}{\pi \tilde{\Gamma}(\gamma + 1)} [|\tilde{y}_1|^2]^\gamma e^{-b|\tilde{y}_1|^2}, \quad \tilde{y}_1 = y_{11} + iy_{12}, \quad i = \sqrt{-1}, \\ &= \frac{b^{\gamma+1}}{\pi \tilde{\Gamma}(\gamma + 1)} [y_{11}^2 + y_{12}^2]^\gamma e^{-b(y_{11}^2 + y_{12}^2)} \end{aligned} \quad (7.2a.18)$$

for $b > 0$, $\Re(\gamma + 1) > 0$, $-\infty < y_{1j} < \infty$, $j = 1, 2$, and $\tilde{f}_8 = 0$ elsewhere. We may take $\gamma = 1$ as the Maxwell-Boltzmann case and $\gamma = \frac{1}{2}$ as the Raleigh case. Then, for $\gamma = 1$, we have

$$\tilde{f}_9(\tilde{y}_1) = \frac{b^2}{\pi} |\tilde{y}_1|^2 e^{-b|\tilde{y}_1|^2}, \quad \tilde{y}_1 = y_{11} + iy_{12}$$

for $b > 0$, $-\infty < y_{1j} < \infty$, $j = 1, 2$, and $\tilde{f}_9 = 0$ elsewhere. Note that in the real case $y_{12} = 0$ so that the functional part of \tilde{f}_6 becomes $y_{11}^2 e^{-by_{11}^2}$, $-\infty < y_{11} < \infty$. However, the normalizing constants in the real and complex cases are evaluated in different domains. Observe that, corresponding to (7.2a.18), the normalizing constant in the real case is $b^{\gamma+\frac{1}{2}}/[\pi^{\frac{1}{2}}\Gamma(\gamma+\frac{1}{2})]$. Thus, the normalizing constant has to be evaluated separately. Consider the integral

$$\int_{-\infty}^{\infty} y_{11}^2 e^{-by_{11}^2} dy_{11} = 2 \int_0^{\infty} y_{11}^2 e^{-by_{11}^2} dy_{11} = \int_0^{\infty} u^{\frac{3}{2}-1} e^{-bu} du = \frac{\sqrt{\pi}}{2b^{\frac{3}{2}}}.$$

Hence,

$$\begin{aligned} f_{10}(y_{11}) &= \frac{2b^{\frac{3}{2}}}{\sqrt{\pi}} y_{11}^2 e^{-by_{11}^2}, \quad -\infty < y_{11} < \infty, \quad b > 0, \\ &= \frac{4b^{\frac{3}{2}}}{\sqrt{\pi}} y_{11}^2 e^{-by_{11}^2}, \quad 0 \leq y_{11} < \infty, \quad b > 0, \end{aligned} \quad (7.2a.19)$$

and $f_{10} = 0$ elsewhere. This is the real Maxwell-Boltzmann case. For the Raleigh case, letting $\gamma = \frac{1}{2}$ in (7.2a.18) yields

$$\begin{aligned} \tilde{f}_{11}(\tilde{y}_1) &= \frac{b^{\frac{3}{2}}}{\pi \Gamma(\frac{3}{2})} [|\tilde{y}_1|^2]^{\frac{1}{2}} e^{-b(|\tilde{y}_1|^2)}, \quad b > 0, \\ &= \frac{2b^{\frac{3}{2}}}{\pi^{\frac{3}{2}}} [y_{11}^2 + y_{12}^2]^{\frac{1}{2}} e^{-b(y_{11}^2 + y_{12}^2)}, \quad -\infty < y_{1j} < \infty, \quad j = 1, 2, \quad b > 0, \end{aligned} \quad (7.2a.20)$$

and $\tilde{f}_{11} = 0$ elsewhere. Then, for $y_{12} = 0$, the functional part of \tilde{f}_{11} is $|y_{11}| e^{-by_{11}^2}$ with $-\infty < y_{11} < \infty$. The integral over y_{11} gives

$$\int_{-\infty}^{\infty} |y_{11}| e^{-by_{11}^2} dy_{11} = 2 \int_0^{\infty} y_{11} e^{-by_{11}^2} dy_{11} = b^{-1}.$$

Thus, in the Raleigh case,

$$\begin{aligned} f_{12}(y_{11}) &= b |y_{11}| e^{-by_{11}^2}, \quad -\infty < y_{11} < \infty, \quad b > 0, \\ &= 2b y_{11} e^{-by_{11}^2}, \quad 0 \leq y_{11} < \infty, \quad b > 0, \end{aligned} \quad (7.2a.21)$$

and $f_{12} = 0$ elsewhere. The normalizing constant in (7.2a.18) can be verified by making use of the polar coordinate transformation: $y_{11} = r \cos \theta$, $y_{12} = r \sin \theta$, so that $dy_{11} \wedge dy_{12} = r dr \wedge d\theta$, $0 < \theta \leq 2\pi$, $0 \leq r < \infty$. Then,

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y_{11}^2 + y_{12}^2]^{\gamma} e^{-(y_{11}^2 + y_{12}^2)} dy_{11} \wedge dy_{12} &= 2\pi \int_0^{\infty} (r^2)^{\gamma} e^{-br^2} r dr \\ &= \pi b^{-(\gamma+1)} \Gamma(\gamma + 1) \end{aligned}$$

for $b > 0$, $\Re(\gamma + 1) > 0$.

Exercises 7.2

- 7.2.1. Supply a proof to (7.2.9) by using Theorem 4.2.3.
- 7.2.2. Derive the exact density of the determinant in (7.2.15) for $p = 2$.
- 7.2.3. Verify the results in (7.2.18) and (7.2.19).
- 7.2.4. Derive the normalizing constants \tilde{c}_1 in (7.2a.12) and \tilde{c}_2 in (7.2a.13).
- 7.2.5. Derive \tilde{c}_3 in (7.2a.14) by integrating out over \tilde{X} .
- 7.2.6. Approximate \tilde{c}_1 and \tilde{c}_2 of Exercise 7.2.4 by making use of Stirling's approximation, and then show that the result agrees with that in Exercise 7.2.5.
- 7.2.7. Derive (state and prove) for the complex case the lemmas corresponding to Lemmas 7.2.1 and 7.2.2.

7.3. Real Rectangular Matrix-Variate Type-1 and Type-2 Beta Densities

Let us begin with the real case. Let $A > O$ be $p \times p$ and $B > O$ be $q \times q$ where A and B are real constant matrices. Let $X = (x_{ij})$ be a $p \times q$, $q \geq p$, matrix of distinct real scalar variables x_{ij} 's as its elements, X being of full rank p . Then, $A^{\frac{1}{2}} X B X' A^{\frac{1}{2}} > O$ is real positive definite where $A^{\frac{1}{2}}$ is the positive definite square root of the positive definite matrix A . Let $|\cdot|$ represent the determinant of (\cdot) when (\cdot) is real or complex, and $|\det(\cdot)|$

be the absolute value of the determinant of (\cdot) when (\cdot) is in the complex domain. Consider the following density:

$$g_1(X) = C_1 |A^{\frac{1}{2}} X B X' A^{\frac{1}{2}}|^{\gamma} |I - A^{\frac{1}{2}} X B X' A^{\frac{1}{2}}|^{\beta - \frac{p+1}{2}} \quad (7.3.1)$$

for $A > O$, $B > O$, $I - A^{\frac{1}{2}} X B X' A^{\frac{1}{2}} > O$, $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, and $g_1 = 0$ elsewhere, where C_1 is the normalizing constant. Accordingly, $U = A^{\frac{1}{2}} X B X' A^{\frac{1}{2}} > O$ and $I - U > O$ or U and $I - U$ are both positive definite. We now make the transformations $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}}$ and $S = Y Y'$. Then, proceeding as in the case of the rectangular matrix-variate gamma density discussed in Sect. 7.2, and evaluating the final part involving S with the help of a real positive definite matrix-variate type-1 beta integral, we obtain the following normalizing constant:

$$C_1 = |A|^{\frac{q}{2}} |B|^{\frac{p}{2}} \frac{\Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}}} \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\beta) \Gamma_p(\gamma + \frac{q}{2})} \quad (7.3.2)$$

for $A > O$, $B > O$, $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$. Usually the parameters associated with a statistical density are real, which is the case for γ and β . Nonetheless, the conditions will be stated for general complex parameters. When the density of X is as given in g_1 , the density of $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}}$ is given by

$$g_2(Y) = \frac{\Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}}} \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\beta) \Gamma_p(\gamma + \frac{q}{2})} |Y Y'|^{\gamma} |I - Y Y'|^{\beta - \frac{p+1}{2}} \quad (7.3.3)$$

for $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, $Y Y' > O$, $I - Y Y' > O$, and $g_2 = 0$ elsewhere. When X has the density specified in (7.3.1), the density of $S = Y Y'$ is given by

$$g_3(S) = \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\beta) \Gamma_p(\gamma + \frac{q}{2})} |S|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} |I - S|^{\beta - \frac{p+1}{2}} \quad (7.3.4)$$

for $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, $S > O$, $I - S > O$, and $g_3 = 0$ elsewhere, which is the usual real matrix-variate type-1 beta density. Observe that the density $g_1(X)$ is also available from the pathway form of the real matrix-variate gamma case introduced in Sect. 7.2.

Example 7.3.1. Let $U_1 = A^{\frac{1}{2}} X B X' A^{\frac{1}{2}}$, $U_2 = X B X'$, $U_3 = B^{\frac{1}{2}} X' A X B^{\frac{1}{2}}$ and $U_4 = X' A X$. If X has the rectangular matrix-variate type-1 beta density given in (7.3.1), evaluate the densities of U_1 , U_2 , U_3 and U_4 whenever possible.

Solution 7.3.1. The matrix U_1 is already present in the density of X , namely (7.3.1). Now, we have to convert the density of X into the density of U_1 . Consider the transformations $Y = A^{\frac{1}{2}}XB^{\frac{1}{2}}$, $S = YY' = U_1$ and the density of S is given in (7.3.4). Thus, U_1 has a real matrix-variate type-1 beta density with the parameters $\gamma + \frac{q}{2}$ and β . Now, on applying the same transformations as above with $A = I$, the density appearing in (7.3.4), which is the density of U_2 , becomes

$$g_3(U_2)dU_2 = \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\gamma + \frac{q}{2})\Gamma(\beta)} |A|^{\gamma + \frac{q}{2}} |U_2|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} |I - AU_2|^{\beta - \frac{p+1}{2}} dU_2 \quad (i)$$

for $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, $A > O$, $U_2 > O$, $I - A^{\frac{1}{2}}U_2A^{\frac{1}{2}} > O$, and zero elsewhere, so that U_2 has a scaled real matrix-variate type-1 beta distribution with parameters $(\gamma + \frac{q}{2}, \beta)$ and scaling matrix $A > O$. For $q > p$, both $X'AX$ and $B^{\frac{1}{2}}X'AXB^{\frac{1}{2}}$ are positive semi-definite matrices whose determinants are thus equal to zero. Accordingly, the densities do not exist whenever $q > p$. When $q = p$, U_3 has a $q \times q$ real matrix-variate type-1 beta distribution with parameters $(\gamma + \frac{p}{2}, \beta)$ and U_4 is a scaled version of a type-1 beta matrix variable whose density is of the form given in (i) wherein B is the scaling matrix and p and q are interchanged. This completes the solution.

7.3.1. Arbitrary moments

The h -th moment of the determinant of $U = A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$ with h being arbitrary, is available from the normalizing constant given in (7.3.2) on observing that when the h -th moment is taken, the only change is that γ turns into $\gamma + h$. Thus,

$$E[|U|^h] = E[|YY'|^h] = E[|S|^h] = \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\gamma + \frac{q}{2} + \beta + h)} \quad (7.3.5)$$

$$= \prod_{j=1}^p \frac{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2})} \frac{\Gamma(\gamma + \frac{q}{2} + \beta - \frac{j-1}{2})}{\Gamma(\gamma + \frac{q}{2} + \beta - \frac{j-1}{2} + h)} \quad (7.3.6)$$

$$= E[u_1^h]E[u_2^h] \cdots E[u_p^h] \quad (7.3.7)$$

where u_1, \dots, u_p are mutually independently distributed real scalar type-1 beta random variables with the parameters $(\gamma + \frac{q}{2} - \frac{j-1}{2}, \beta)$, $j = 1, \dots, p$, provided $\Re(\beta) > \frac{p-1}{2}$ and $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$.

7.3.2. Special case: $p = 1$

For the case $p = 1$, let the positive definite 1×1 matrix A be the scalar $b > 0$ and X which is $1 \times q$, be equal to (x_1, \dots, x_q) . Then,

$$AXBX' = b(x_1, \dots, x_q)B \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix}$$

is a real quadratic form, the matrix of the quadratic form being $B > O$. Letting $Y = XB^{\frac{1}{2}}$, $dY = |B|^{\frac{1}{2}}dX$, and the density of Y , denoted by $g_4(Y)$, is then given by

$$\begin{aligned} g_4(Y) &= b^{\gamma + \frac{q}{2}} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\gamma + \frac{q}{2} + \beta)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\beta)} |YY'|^\gamma \\ &\quad \times |I - bYY'|^{\beta-1}, \quad YY' > O, \quad I - bYY' > O, \quad \Re(\gamma + \frac{q}{2}) > 0, \quad \Re(\beta) > 0 \\ &= b^{\gamma + \frac{q}{2}} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\gamma + \frac{q}{2} + \beta)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\beta)} [y_1^2 + \dots + y_q^2]^\gamma \\ &\quad \times [1 - b(y_1^2 + \dots + y_q^2)]^{\beta-1}, \quad Y = (y_1, \dots, y_q), \end{aligned} \quad (7.3.8)$$

for $b > 0$, $\Re(\gamma + 1) > 0$, $\Re(\beta) > 0$, $1 - b(y_1^2 + \dots + y_q^2) > 0$, and $g_4 = 0$ elsewhere. The form of the density in (7.3.8) presents some interest as it appears in various areas of research. In reliability studies, a popular model for the lifetime of components corresponds to (7.3.8) wherein $\gamma = 0$ in both the scalar and multivariate cases. When independently distributed isotropic random points are considered in connection with certain geometrical probability problems, a popular model for the distribution of the random points is the type-1 beta form or (7.3.8) for $\gamma = 0$. Earlier results obtained assuming that $\gamma = 0$ and the new case where $\gamma \neq 0$ in geometrical probability problems are discussed in Chapter 4 of Mathai (1999). We will take (7.3.8) as the standard form of the real rectangular matrix-variate type-1 beta density for the case $p = 1$ in a $p \times q$, $q \geq p$, real matrix X of rank p . For verifying the normalizing constant in (7.3.8), one can apply Theorem 4.2.3. Letting $S = YY'$, $dY = \frac{\pi^{\frac{q}{2}}}{\Gamma(\frac{q}{2})} |S|^{\frac{q}{2}-1} dS$, which once substituted to dY in (7.3.8) yields a total integral equal to one upon integrating out S with the help of a real matrix-variate type-1 beta integral (in this case a real scalar type-1 beta integral); accordingly, the constant part in (7.3.8) is indeed the normalizing constant. In this case, the density of $S = YY'$ is given by

$$g_5(S) = b^{\gamma + \frac{q}{2}} \frac{\Gamma(\gamma + \frac{q}{2} + \beta)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\beta)} |S|^{\gamma + \frac{q}{2} - 1} |I - bS|^{\beta-1} \quad (7.3.9)$$

for $S > O$, $b > 0$, $\Re(\gamma + \frac{q}{2}) > 0$, $\Re(\beta) > 0$, and $g_5 = 0$ elsewhere. Observe that this S is actually a real scalar variable.

As obtained from (7.3.8), the type-1 beta form of the density in the real scalar case, that is, for $p = 1$ and $q = 1$, is

$$g_6(y_1) = b^{\gamma + \frac{1}{2}} \frac{\Gamma(\gamma + \frac{1}{2} + \beta)}{\Gamma(\gamma + \frac{1}{2})\Gamma(\beta)} [y_1^2]^\gamma [1 - by_1^2]^{\beta-1}, \quad (7.3.10)$$

for $b > 0$, $\beta > 0$, $\gamma + \frac{1}{2} > 0$, $-\frac{1}{\sqrt{b}} < y_1 < \frac{1}{\sqrt{b}}$, and $g_6 = 0$ elsewhere. When the support is $0 < y_1 < \frac{1}{\sqrt{b}}$, the above density which is symmetric, is multiplied by two.

7.3a. Rectangular Matrix-Variate Type-1 Beta Density, Complex Case

Consider the following function:

$$\tilde{g}_1(\tilde{X}) = \tilde{C}_1 |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^\gamma |\det(I - A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^{\beta-p} \quad (7.3a.1)$$

for $A > O$, $B > O$, $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, $I - A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}} > O$, and $\tilde{g}_1 = 0$ elsewhere. The normalizing constant can be evaluated by proceeding as in the real rectangular matrix-variate case. Let $\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}}$ so that $\tilde{S} = \tilde{Y} \tilde{Y}^*$, and then integrate out \tilde{S} by using a complex matrix-variate type-1 beta integral, which yields the following normalizing constant:

$$\tilde{C}_1 = |\det(A)|^q |\det(B)|^p \frac{\tilde{\Gamma}_p(q) \tilde{\Gamma}_p(\gamma + q + \beta)}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)} \quad (7.3a.2)$$

for $\Re(\gamma + q) > p - 1$, $\Re(\beta) > p - 1$, $A > O$, $B > O$.

7.3a.1. Different versions of the type-1 beta density, the complex case

The densities that follow can be obtained from that specified in (7.3a.1) and certain related transformations. The density of $\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}}$ is given by

$$\tilde{g}_2(\tilde{Y}) = \frac{\tilde{\Gamma}_p(q) \tilde{\Gamma}_p(\gamma + q + \beta)}{\pi^{qp} \tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)} |\det(\tilde{Y} \tilde{Y}^*)|^\gamma |\det(I - \tilde{Y} \tilde{Y}^*)|^{\beta-p} \quad (7.3a.3)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, $I - \tilde{Y} \tilde{Y}^* > O$, and $\tilde{g}_2 = 0$ elsewhere. The density of $\tilde{S} = \tilde{Y} \tilde{Y}^*$ is the following:

$$\tilde{g}_3(\tilde{S}) = \frac{\tilde{\Gamma}_p(\gamma + q + \beta)}{\tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)} |\det(\tilde{S})|^{\gamma+q-p} |\det(I - \tilde{S})|^{\beta-p} \quad (7.3a.4)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, $\tilde{S} > O$, $I - \tilde{S} > O$, and $\tilde{g}_3 = 0$ elsewhere.

7.3a.2. Multivariate type-1 beta density, the complex case

When $p = 1$, \tilde{X} is the $1 \times q$ vector $(\tilde{x}_1, \dots, \tilde{x}_q)$ and the 1×1 matrix A will be denoted by $b > 0$. The resulting density will then have the same structure as that given in (7.3a.1) with p replaced by 1 and A replaced by $b > 0$:

$$\tilde{g}_4(\tilde{X}) = \tilde{C}_2 |\det(\tilde{X} B \tilde{X}^*)|^\gamma |\det(I - b \tilde{X} B \tilde{X}^*)|^{\beta-1} \quad (7.3a.5)$$

for $B > O$, $b > 0$, $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, $I - b \tilde{X} B \tilde{X}^* > O$, and $\tilde{g}_4 = 0$ elsewhere, where the normalizing constant \tilde{C}_2 is

$$\tilde{C}_2 = b^{\gamma+q} |\det(B)| \frac{\tilde{\Gamma}(q) \tilde{\Gamma}(\gamma + q + \beta)}{\pi^q \tilde{\Gamma}(\gamma + q) \tilde{\Gamma}(\beta)} \quad (7.3a.6)$$

for $b > 0$, $B > O$, $\Re(\beta) > 0$, $\Re(\gamma + q) > 0$. Letting $\tilde{Y} = \tilde{X} B^{\frac{1}{2}}$ so that $d\tilde{Y} = |\det(B)| d\tilde{X}$, the density of \tilde{Y} reduces to

$$\tilde{g}_5(\tilde{Y}) = b^{\gamma+q} \frac{\tilde{\Gamma}(q) \tilde{\Gamma}(\gamma + q + \beta)}{\pi^q \tilde{\Gamma}(\gamma + q) \tilde{\Gamma}(\beta)} |\det(\tilde{Y} \tilde{Y}^*)|^\gamma |\det(I - b \tilde{Y} \tilde{Y}^*)|^{\beta-1} \quad (7.3a.7)$$

$$= b^{\gamma+q} \frac{\tilde{\Gamma}(q) \tilde{\Gamma}(\gamma + q + \beta)}{\pi^q \tilde{\Gamma}(\gamma + q) \tilde{\Gamma}(\beta)} [|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2]^\gamma [1 - b(|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2)]^{\beta-1} \quad (7.3a.8)$$

for $\Re(\beta) > 0$, $\Re(\gamma + q) > 0$, $1 - b(|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2) > 0$, and $\tilde{g}_5 = 0$ elsewhere. The form appearing in (7.3a.8) is applicable to several problems occurring in various areas, as was the case for (7.3.8) in the real domain. However, geometrical probability problems do not appear to have yet been formulated in the complex domain. Let $\tilde{S} = \tilde{Y} \tilde{Y}^*$, \tilde{S} being in this case a real scalar denoted by s whose density is

$$\tilde{g}_6(s) = b^{\gamma+q} \frac{\tilde{\Gamma}(\gamma + q + \beta)}{\tilde{\Gamma}(\gamma + q) \tilde{\Gamma}(\beta)} s^{\gamma+q-1} (1 - bs)^{\beta-1}, \quad b > 0 \quad (7.3a.9)$$

for $\Re(\beta) > 0$, $\Re(\gamma + q) > 0$, $s > 0$, $1 - bs > 0$, and $\tilde{g}_6 = 0$ elsewhere. Thus, s is real scalar type-1 beta random variable with parameters $(\gamma + q, \beta)$ and scaling factor $b > 0$. Note that in the real case, the distribution was also a scalar type-1 beta random variable, but having a different first parameter, namely, $\gamma + \frac{q}{2}$, its second parameter and scaling factor remaining β and b .

Example 7.3a.1. Express the density (7.3a.5) explicitly for $b = 5$, $p = 1$, $q = 2$, $\beta = 3$, $\gamma = 4$, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2] = [x_1 + iy_1, x_2 + iy_2]$, where $x_j, y_j, j = 1, 2$, are real variables, $i = \sqrt{-1}$, and

$$B = \begin{bmatrix} 3 & 1+i \\ 1-i & 1 \end{bmatrix}.$$

Solution 7.3a.1. Observe that since $B = B^*$, B is Hermitian. Its leading minors being $|(3)| = 3 > 0$ and $\begin{vmatrix} 3 & 1+i \\ 1-i & 1 \end{vmatrix} = (3)(1) - (1+i)(1-i) = 3 - 2 = 1 > 0$, B is also positive definite. Letting $Q = \tilde{X}B\tilde{X}^*$,

$$\begin{aligned} Q &= 3(x_1 + iy_1)(x_1 - iy_1) + (1)(x_2 + iy_2)(x_2 - iy_2) + (1+i)(x_1 + iy_1)(x_2 - iy_2) \\ &\quad + (1-i)(x_2 + iy_2)(x_1 - iy_1) \\ &= 3(x_1^2 + y_1^2) + (x_2^2 + y_2^2) + (1+i)[x_1x_2 + y_1y_2 - i(x_1y_2 - x_2y_1)] \\ &\quad + (1-i)[x_1x_2 + y_1y_2 - i(x_2y_1 - x_1y_2)] \\ &= 3(x_1^2 + y_1^2) + (x_2^2 + y_2^2) + 2(x_1x_2 + y_1y_2) + 2(x_1y_2 - x_2y_1). \end{aligned} \quad (i)$$

The normalizing constant being

$$\begin{aligned} b^{\gamma+q} |\det(B)| \frac{\tilde{\Gamma}(q)}{\pi^q} \frac{\tilde{\Gamma}(\gamma+q+\beta)}{\tilde{\Gamma}(\gamma+q)\tilde{\Gamma}(\beta)} &= 5^6(1) \frac{\Gamma(2)}{\pi^2} \frac{\Gamma(9)}{\Gamma(6)\Gamma(3)} \\ &= \frac{5^6(1!)(8!)}{\pi^2(5!)(2!)} = \frac{5^6(168)}{\pi^2}. \end{aligned} \quad (ii)$$

The explicit form of the density (7.3a.5) is thus the following:

$$\tilde{g}_4(\tilde{X}) = \frac{5^6(168)}{\pi^2} Q^3 [1 - 5Q]^2, \quad 1 - 3Q > 0, \quad Q > 0,$$

and zero elsewhere, where Q is given in (i). It is a multivariate generalized type-1 complex-variate beta density whose scaling factor is 5. Observe that even though \tilde{X} is complex, $\tilde{g}_4(\tilde{X})$ is real-valued.

7.3a.3. Arbitrary moments in the complex case

Consider again the density of the complex $p \times q$, $q \geq p$, matrix-variate random variable \tilde{X} of full rank p having the density specified in (7.3a.1). Let $\tilde{U} = A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}}$. The h -th moment of the absolute value of the determinant of \tilde{U} , that is, $E[|\det(\tilde{U})|^h]$, will now be determined for arbitrary h . As before, note that when the expected value is taken,

the only change is that the parameter γ is replaced by $\gamma + h$, so that the moment is available from the normalizing constant present in (7.3a.2). Thus,

$$E[|\det(\tilde{U})|^h] = \frac{\tilde{\Gamma}_p(\gamma + q + h)}{\tilde{\Gamma}_p(\gamma + q)} \frac{\tilde{\Gamma}_p(\gamma + q + \beta)}{\tilde{\Gamma}_p(\gamma + q + \beta + h)} \quad (7.3a.10)$$

$$= \prod_{j=1}^p \frac{\Gamma(\gamma + q + h - (j - 1))}{\Gamma(\gamma + q - (j - 1))} \frac{\Gamma(\gamma + q + \beta - (j - 1))}{\Gamma(\gamma + q + \beta - (j - 1) + h)} \quad (7.3a.11)$$

$$= E[u_1]^h E[u_2]^h \cdots E[u_p]^h \quad (7.3a.12)$$

where u_1, \dots, u_p are independently distributed real scalar type-1 beta random variables with the parameters $(\gamma + q - (j - 1), \beta)$ for $j = 1, \dots, p$. The results are the same as those obtained in the real case except that the parameters are slightly different, the parameters being $(\gamma + \frac{q}{2} - \frac{j-1}{2}, \beta)$, $j = 1, \dots, p$, in the real domain. Accordingly, the absolute value of the determinant of \tilde{U} in the complex case has the following structural representation:

$$|\det(\tilde{U})| = |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})| = |\det(\tilde{Y} \tilde{Y}^*)| = |\det(\tilde{S})| = u_1 \cdots u_p \quad (7.3a.13)$$

where u_1, \dots, u_p are mutually independently distributed real scalar type-1 beta random variables with the parameters $(\gamma + q - (j - 1), \beta)$, $j = 1, \dots, p$.

We now consider the scalar type-1 beta density in the complex case. Thus, letting $p = 1$ and $q = 1$ in (7.3a.8), we have

$$\begin{aligned} \tilde{g}_7(\tilde{y}_1) &= b^{\gamma+1} \frac{1}{\pi} \frac{\tilde{\Gamma}(\gamma + 1 + \beta)}{\tilde{\Gamma}(\gamma + 1)\tilde{\Gamma}(\beta)} [|\tilde{y}_1|^2]^\gamma [1 - b|\tilde{y}_1|^2]^{\beta-1}, \quad \tilde{y}_1 = y_{11} + iy_{12} \\ &= b^{\gamma+1} \frac{1}{\pi} \frac{\Gamma(\gamma + 1 + \beta)}{\Gamma(\gamma + 1)\Gamma(\beta)} [y_{11}^2 + y_{12}^2]^\gamma [1 - b(y_{11}^2 + y_{12}^2)]^{\beta-1} \end{aligned} \quad (7.3a.14)$$

for $b > 0$, $\Re(\beta) > 0$, $\Re(\gamma) > -1$, $-\infty < y_{1j} < \infty$, $j = 1, 2$, $1 - b(y_{11}^2 + y_{12}^2) > 0$, and $\tilde{g}_7 = 0$ elsewhere. The normalizing constant in (7.3a.14) can be verified by making the polar coordinate transformation $y_{11} = r \cos \theta$, $y_{12} = r \sin \theta$, as was done earlier.

Exercises 7.3

7.3.1. Derive the normalizing constant C_1 in (7.3.2) and verify the normalizing constants in (7.3.3) and (7.3.4).

7.3.2. From $E[|A^{\frac{1}{2}} X B X' A^{\frac{1}{2}}|^h]$ or otherwise, derive the h -th moment of $|X B X'|$. What is then the structural representation corresponding to (7.3.7)?

7.3.3. From (7.3.7) or otherwise, derive the exact density of $|U|$ for the cases (1): $p = 2$; (2): $p = 3$.

7.3.4. Write down the conditions on the parameters γ and β in (7.3.6) so that the exact density of $|U|$ can easily be evaluated for some $p \geq 4$.

7.3.5. Evaluate the normalizing constant in (7.3.8) by making use of the general polar coordinate transformation.

7.3.6. Evaluate the normalizing constant in (7.3a.2).

7.3.7. Derive the exact density of $|\det(\tilde{U})|$ in (7.3a.13) for (1): $p = 2$; (2): $p = 3$.

7.4. The Real Rectangular Matrix-Variate Type-2 Beta Density

Let us consider a $p \times q$, $q \geq p$, matrix X of full rank p and the following associated density:

$$g_8(X) = C_3 |AXBX'|^\gamma |I + AXBX'|^{-(\gamma + \frac{q}{2} + \beta)} \quad (7.4.1)$$

for $A > O$, $B > O$, $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, and $g_8 = 0$ elsewhere. The normalizing constant can be seen to be the following:

$$C_3 = |A|^{\frac{q}{2}} |B|^{\frac{p}{2}} \frac{\Gamma_p(\frac{q}{2}) \Gamma_p(\gamma + \frac{q}{2} + \beta)}{\pi^{\frac{qp}{2}} \Gamma_p(\gamma + \frac{q}{2}) \Gamma_p(\beta)} \quad (7.4.2)$$

for $A > O$, $B > O$, $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$. Letting $Y = A^{\frac{1}{2}} X B^{\frac{1}{2}}$, its density denoted by $g_9(Y)$, is

$$g_9(Y) = \frac{\Gamma_p(\frac{q}{2}) \Gamma_p(\gamma + \frac{q}{2} + \beta)}{\pi^{\frac{qp}{2}} \Gamma_p(\gamma + \frac{q}{2}) \Gamma_p(\beta)} |YY'|^\gamma |I + YY'|^{-(\gamma + \frac{q}{2} + \beta)} \quad (7.4.3)$$

for $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, and $g_9 = 0$ elsewhere. The density of $S = YY'$ then reduces to

$$g_{10}(S) = \frac{\Gamma_p(\gamma + \frac{q}{2} + \beta)}{\Gamma_p(\gamma + \frac{q}{2}) \Gamma_p(\beta)} |S|^{\gamma + \frac{q}{2} - \frac{p+1}{2}} |I + S|^{-(\gamma + \frac{q}{2} + \beta)}, \quad (7.4.4)$$

for $\Re(\beta) > \frac{p-1}{2}$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, and $g_{10} = 0$ elsewhere.

7.4.1. The real type-2 beta density in the multivariate case

Consider the case $p = 1$ and $A = b > 0$ in (7.4.1). The resulting density has the same structure, with A replaced by b and X being the $1 \times q$ vector (x_1, \dots, x_q) . Letting $Y = XB^{\frac{1}{2}}$, the following density of $Y = (y_1, \dots, y_q)$, denoted by $g_{11}(Y)$, is obtained:

$$g_{11} = b^{\gamma + \frac{q}{2}} \frac{\Gamma(\frac{q}{2})}{\pi^{\frac{q}{2}}} \frac{\Gamma(\gamma + \frac{q}{2} + \beta)}{\Gamma(\gamma + \frac{q}{2})\Gamma(\beta)} [y_1^2 + \dots + y_q^2]^\gamma \times [1 + b(y_1^2 + \dots + y_q^2)]^{-(\gamma + \frac{q}{2} + \beta)} \quad (7.4.5)$$

for $b > 0$, $\Re(\gamma + \frac{q}{2}) > 0$, $\Re(\beta) > 0$, and $g_{11} = 0$ elsewhere. The density appearing in (7.4.5) will be referred to as the standard form of the real rectangular matrix-variate type-2 beta density. In this case, A is taken as $A = b > 0$ and $Y = XB^{\frac{1}{2}}$. What might be the standard form of the real type-2 beta density in the real scalar case, that is, when it is assumed that $p = 1$, $q = 1$, $A = b > 0$ and $B = 1$ in (7.4.1)? In this case, it is seen from (7.4.5) that

$$g_{12}(y_1) = b^{\gamma + \frac{1}{2}} \frac{\Gamma(\gamma + \frac{1}{2} + \beta)}{\Gamma(\gamma + \frac{1}{2})\Gamma(\beta)} [y_1^2]^\gamma [1 + by_1^2]^{-(\gamma + \frac{1}{2} + \beta)}, \quad -\infty < y_1 < \infty, \quad (7.4.6)$$

for $\Re(\beta) > 0$, $\Re(\gamma + \frac{1}{2}) > 0$, $b > 0$, and $g_{12} = 0$ elsewhere.

7.4.2. Moments in the real rectangular matrix-variate type-2 beta density

Letting $U = A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$, what would be the h -th moment of the determinant of U , that is, $E[|U|^h]$ for arbitrary h ? Upon determining $E[|U|^h]$, the parameter γ is replaced by $\gamma + h$ while the other parameters remain unchanged. The h -th moment which is thus available from the normalizing constant, is given by

$$E[|U|^h] = \frac{\Gamma_p(\gamma + \frac{q}{2} + h)}{\Gamma_p(\gamma + \frac{q}{2})} \frac{\Gamma_p(\beta - h)}{\Gamma_p(\beta)} \quad (7.4.7)$$

$$= \prod_{j=1}^p \frac{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{j-1}{2})} \frac{\Gamma(\beta - \frac{j-1}{2} - h)}{\Gamma(\beta - \frac{j-1}{2})} \quad (7.4.8)$$

$$= E[u_1^h]E[u_2^h] \cdots E[u_p^h] \quad (7.4.9)$$

where u_1, \dots, u_p are mutually independently distributed real scalar type-2 beta random variables with the parameters $(\gamma + \frac{q}{2} - \frac{j-1}{2}, \beta - \frac{j-1}{2})$, $j = 1, \dots, p$. That is, (7.4.9) or (7.4.10) gives a structural representation to the determinant of U as

$$|U| = |A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}| = |YY'| = |S| = u_1 \cdots u_p \quad (7.4.10)$$

where the u_j 's are mutually independently distributed real scalar type-2 beta random variables as specified above.

Example 7.4.1. Evaluate the density of $u = |A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|$ for $p = 2$ and the general parameters γ, q, β where X has a real rectangular matrix-variate type-2 beta density with the parameter matrices $A > O$ and $B > O$ where A is $p \times p$, B is $q \times q$ and X is a $p \times q$, $q \geq p$, rank p matrix.

Solution 7.4.1. The general h -th moment of u can be determined from (7.4.8). Letting $p = 2$, we have

$$E[u^h] = \frac{\Gamma(\gamma + \frac{q}{2} - \frac{1}{2} + h)\Gamma(\gamma + \frac{q}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{1}{2})\Gamma(\gamma + \frac{q}{2})} \frac{\Gamma(\beta - \frac{1}{2} - h)\Gamma(\beta - h)}{\Gamma(\beta - \frac{1}{2})\Gamma(\beta)} \quad (i)$$

for $-\gamma - \frac{q}{2} + \frac{1}{2} < \Re(h) < \beta - \frac{1}{2}$, $\beta > \frac{1}{2}$, $\gamma + \frac{q}{2} > \frac{1}{2}$. Since four pairs of gamma functions differ by $\frac{1}{2}$, we can combine them by applying the duplication formula for gamma functions, namely,

$$\Gamma(z)\Gamma(z + 1/2) = \pi^{\frac{1}{2}} 2^{1-2z} \Gamma(2z). \quad (ii)$$

Take $z = \gamma + \frac{q}{2} - \frac{1}{2} + h$ and $z = \gamma + \frac{q}{2} - \frac{1}{2}$ in the first set of gamma ratios in (i) and $z = \beta - \frac{1}{2} - h$ and $z = \beta - \frac{1}{2}$ in the second set of gamma ratios in (i). Then, we have the following:

$$\begin{aligned} \frac{\Gamma(\gamma + \frac{q}{2} - \frac{1}{2} + h)\Gamma(\gamma + \frac{q}{2} + h)}{\Gamma(\gamma + \frac{q}{2} - \frac{1}{2})\Gamma(\gamma + \frac{q}{2})} &= \frac{\pi^{\frac{1}{2}} 2^{1-2\gamma-q+1-2h} \Gamma(2\gamma + q - 1 + 2h)}{\pi^{\frac{1}{2}} 2^{1-2\gamma-q+1} \Gamma(2\gamma + q - 1)} \\ &= 2^{-2h} \frac{\Gamma(2\gamma + q - 1 + 2h)}{\Gamma(2\gamma + q - 1)} \end{aligned} \quad (iii)$$

$$\begin{aligned} \frac{\Gamma(\beta - \frac{1}{2} - h)\Gamma(\beta - h)}{\Gamma(\beta - \frac{1}{2})\Gamma(\beta)} &= \frac{\pi^{\frac{1}{2}} 2^{1-2\beta+1+2h} \Gamma(2\beta - 1 - 2h)}{\pi^{\frac{1}{2}} 2^{1-2\beta+1} \Gamma(2\beta - 1)} \\ &= 2^{2h} \frac{\Gamma(2\beta - 1 - 2h)}{\Gamma(2\beta - 1)}, \end{aligned} \quad (iv)$$

the product of (iii) and (iv) yielding the simplified representation of the h -th moment of u that follows:

$$E[u^h] = \frac{\Gamma(2\gamma + q - 1 + 2h)}{\Gamma(2\gamma + q - 1)} \frac{\Gamma(2\beta - 1 - 2h)}{\Gamma(2\beta - 1)}.$$

Now, since $E[u^h] = E[u^{\frac{1}{2}}]^{2h} \equiv E[y^t]$ with $y = u^{\frac{1}{2}}$ and $t = 2h$, we have

$$E[y^t] = \frac{\Gamma(2\gamma + q - 1 + t)}{\Gamma(2\gamma + q - 1)} \frac{\Gamma(2\beta - 1 - t)}{\Gamma(2\beta - 1)}. \quad (v)$$

As t is arbitrary in (ν) , the moment expression will uniquely determine the density of y . Accordingly, y has a real scalar type-2 beta distribution with the parameters $(2\gamma + q - 1, 2\beta - 1)$, and so, its density denoted by $f(y)$, is

$$\begin{aligned} f(y)dy &= \frac{\Gamma(2\gamma + q + 2\beta - 2)}{\Gamma(2\gamma + q - 1)\Gamma(2\beta - 1)} y^{2\gamma+q-2} (1+y)^{-(2\gamma+q+2\beta-2)} dy \\ &= \frac{\Gamma(2\gamma + q + 2\beta - 2)}{\Gamma(2\gamma + q - 1)\Gamma(2\beta - 1)} \frac{1}{2} u^{-\frac{1}{2}} u^{\gamma+\frac{q}{2}-1} (1+u^{\frac{1}{2}})^{-(2\gamma+q+2\beta-2)} du. \end{aligned}$$

Thus, the density of u , denoted by $g(u)$, is the following:

$$g(u) = \frac{1}{2} \frac{\Gamma(2\gamma + q + 2\beta - 2)}{\Gamma(2\gamma + q - 1)\Gamma(2\beta - 1)} u^{\gamma+\frac{q-1}{2}-1} (1+u^{\frac{1}{2}})^{-(2\gamma+q+2\beta-2)}$$

for $0 \leq u < \infty$, and zero elsewhere, where the original conditions on the parameters remain the same. It can be readily verified that $g(u)$ is a density.

7.4.3. A pathway extension in the real case

Let us relabel $f_{10}(X)$ as specified in (7.2.16), as $g_{13}(X)$ in this section:

$$g_{13}(X) = C_4 |AXBX'|^\gamma |I - a(1-\alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{\frac{\eta}{1-\alpha}} \quad (7.4.11)$$

for $A > O$, $B > O$, $\eta > 0$, $a > 0$, $\alpha < 1$, $1 - a(1-\alpha)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}} > O$, and $g_{13} = 0$ elsewhere, where C_4 is the normalizing constant. Observe that for $\alpha < 1$, $a(1-\alpha) > 0$ and hence the model in (7.4.11) is a generalization of the real rectangular matrix-variate type-1 beta density considered in (7.3.1). When $\alpha < 1$, the normalizing constant C_4 is of the form given in (7.2.18). For $\alpha > 1$, we may write $1 - \alpha = -(\alpha - 1)$, so that $-a(1-\alpha) = a(\alpha-1) > 0$, $\alpha > 1$ in (7.4.11) and the exponent $\frac{\eta}{1-\alpha}$ changes to $-\frac{\eta}{\alpha-1}$; thus, the model appearing in (7.4.11) becomes the following generalization of the rectangular matrix-variate type-2 beta density given in (7.4.1):

$$g_{14}(X) = C_5 |AXBX'|^\gamma |I + a(\alpha-1)A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}|^{-\frac{\eta}{\alpha-1}} \quad (7.4.12)$$

for $A > O$, $B > O$, $\eta > 0$, $a > 0$, $\alpha > 1$ and $g_{14} = 0$ elsewhere. The normalizing constant C_5 will then be different from that associated with the type-1 case. Actually, in the type-2 case, the normalizing constant is available from (7.2.19). The model appearing in (7.4.12) is a generalization of the real rectangular matrix-variate type-2 beta model considered in (7.4.1). When $\alpha \rightarrow 1$, the model in (7.4.11) converges to a generalized form of the real rectangular matrix-variate gamma model in (7.2.5), namely,

$$g_{15}(X) = C_6 |AXBX'|^\gamma e^{-a\eta \text{tr}(AXBX')} \quad (7.4.13)$$

where

$$C_6 = (a\eta)^{p(\gamma + \frac{q}{2})} \frac{|A|^{\frac{q}{2}} |B|^{\frac{p}{2}} \Gamma_p(\frac{q}{2})}{\pi^{\frac{qp}{2}} \Gamma_p(\gamma + \frac{q}{2})} \quad (7.4.14)$$

for $a > 0$, $\eta > 0$, $A > O$, $B > O$, $\Re(\gamma + \frac{q}{2}) > \frac{p-1}{2}$, and $g_{15} = 0$ elsewhere. More properties of the model given in (7.4.11) have already been provided in Sect. 4.2. The real rectangular matrix-variate pathway model was introduced in Mathai (2005).

7.4a. Complex Rectangular Matrix-Variate Type-2 Beta Density

Let us consider a full rank $p \times q$, $q \geq p$, matrix \tilde{X} in the complex domain and the following associated density:

$$\tilde{g}_8(\tilde{X}) = \tilde{C}_3 |\det(A\tilde{X}B\tilde{X}^*)|^\gamma |\det(I + A\tilde{X}B\tilde{X}^*)|^{-(\beta + \gamma + q)} \quad (7.4a.1)$$

for $A > O$, $B > O$, $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, and $\tilde{g}_8 = 0$ elsewhere, where \tilde{C}_3 is the normalizing constant. Let

$$\tilde{Y} = A^{\frac{1}{2}} \tilde{X} B^{\frac{1}{2}} \Rightarrow d\tilde{Y} = |\det(A)|^q |\det(B)|^p d\tilde{X},$$

and make the transformation

$$\tilde{S} = \tilde{Y}\tilde{Y}^* \Rightarrow d\tilde{Y} = \frac{\pi^{qp}}{\tilde{\Gamma}_p(q)} |\det(\tilde{S})|^{q-p} d\tilde{S}.$$

Then, the integral over \tilde{S} can be evaluated by means of a complex matrix-variate type-2 beta integral. That is,

$$\int_{\tilde{S}} |\det(\tilde{S})|^{\gamma + q - p} |\det(I + \tilde{S})|^{-(\beta + \gamma + q)} d\tilde{S} = \frac{\tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)}{\tilde{\Gamma}_p(\gamma + q + \beta)} \quad (7.4a.2)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$. The normalizing constant \tilde{C}_3 as well as the densities of \tilde{Y} and \tilde{S} can be determined from the previous steps. The normalizing constant is

$$\tilde{C}_3 = |\det(A)|^q |\det(B)|^p \frac{\tilde{\Gamma}_p(q)}{\pi^{qp}} \frac{\tilde{\Gamma}_p(\gamma + q + \beta)}{\tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)} \quad (7.4a.3)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$. The density of \tilde{Y} , denoted by $\tilde{g}_9(\tilde{Y})$, is given by

$$\tilde{g}_9(\tilde{Y}) = \frac{\tilde{\Gamma}_p(q)}{\pi^{qp}} \frac{\tilde{\Gamma}_p(\gamma + q + \beta)}{\tilde{\Gamma}_p(\gamma + q) \tilde{\Gamma}_p(\beta)} |\det(\tilde{Y}\tilde{Y}^*)|^\gamma |\det(I + \tilde{Y}\tilde{Y}^*)|^{-(\gamma + q + \beta)} \quad (7.4a.4)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, and $\tilde{g}_9 = 0$ elsewhere. The density of \tilde{S} , denoted by $\tilde{g}_{10}(\tilde{S})$, is the following:

$$\tilde{g}_{10}(\tilde{S}) = \frac{\tilde{\Gamma}_p(\gamma + q + \beta)}{\tilde{\Gamma}_p(\gamma + q)\tilde{\Gamma}_p(\beta)} |\det(\tilde{S})|^{\gamma+q-p} |\det(I + \tilde{S})|^{-(\beta+\gamma+q)} \quad (7.4a.5)$$

for $\Re(\beta) > p - 1$, $\Re(\gamma + q) > p - 1$, and $\tilde{g}_{10} = 0$ elsewhere.

7.4a.1. Multivariate complex type-2 beta density

As in Sect. 7.3a, let us consider the special case $p = 1$ in (7.4a.1). So, let the 1×1 matrix A be denoted by $b > 0$ and the $1 \times q$ vector $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_q)$. Then,

$$A\tilde{X}B\tilde{X}^* = b\tilde{X}B\tilde{X}^* = b(\tilde{x}_1, \dots, \tilde{x}_q)B \begin{pmatrix} \tilde{x}_1^* \\ \vdots \\ \tilde{x}_q^* \end{pmatrix} \equiv b\tilde{U} \quad (a)$$

where, in the case of a scalar, an asterisk only designates a complex conjugate. Note that when $p = 1$, $\tilde{U} = \tilde{X}B\tilde{X}^*$ is a positive definite Hermitian form whose density, denoted by $\tilde{g}_{11}(\tilde{U})$, is obtained as:

$$\tilde{g}_{11}(\tilde{U}) = b^{\gamma+q} |\det(B)| \frac{\tilde{\Gamma}(q)}{\pi^q} \frac{\tilde{\Gamma}(\gamma + q + \beta)}{\tilde{\Gamma}(\gamma + q)\tilde{\Gamma}(\beta)} |\det(\tilde{U})|^\gamma |\det(I + b\tilde{U})|^{-(\gamma+q+\beta)} \quad (7.4a.6)$$

for $\Re(\beta) > 0$, $\Re(\gamma + q) > 0$ and $b > 0$, and $\tilde{g}_{11} = 0$ elsewhere. Now, letting $\tilde{X}B^{\frac{1}{2}} = \tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_q)$, the density of \tilde{Y} , denoted by $\tilde{g}_{12}(\tilde{Y})$, is obtained as

$$\tilde{g}_{12}(\tilde{Y}) = b^{\gamma+q} \frac{\tilde{\Gamma}(q)}{\pi^q} \frac{\tilde{\Gamma}(\gamma + q + \beta)}{\tilde{\Gamma}(\gamma + q)\tilde{\Gamma}(\beta)} [|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2]^\gamma [1 + b(|\tilde{y}_1|^2 + \dots + |\tilde{y}_q|^2)]^{-(\gamma+q+\beta)} \quad (7.4a.7)$$

for $\Re(\beta) > 0$, $\Re(\gamma + q) > 0$, $b > 0$, and $\tilde{g}_{12} = 0$ elsewhere. The constant in (7.4a.7) can be verified to be a normalizing constant, either by making use of Theorem 4.2a.3 or a $(2n)$ -variate real polar coordinate transformation, which is left as an exercise to the reader.

Example 7.4a.1. Provide an explicit representation of the complex multivariate density in (7.4a.7) for $p = 2$, $\gamma = 2$, $q = 3$, $b = 3$ and $\beta = 2$.

Solution 7.4a.1. The normalizing constant, denoted by \tilde{c} , is the following:

$$\begin{aligned} \tilde{c} &= b^{\gamma+q} \frac{\tilde{\Gamma}(q)}{\pi^q} \frac{\tilde{\Gamma}(\gamma + q + \beta)}{\tilde{\Gamma}(\gamma + q)\tilde{\Gamma}(\beta)} \\ &= 3^5 \frac{\Gamma(3)}{\pi^3} \frac{\Gamma(7)}{\Gamma(5)\Gamma(2)} = 3^5 \frac{(2!)}{\pi^3} \frac{(6!)}{(4!)(1!)} = \frac{(60)3^5}{\pi^3}. \end{aligned} \quad (i)$$

Letting $\tilde{y}_1 = y_{11} + iy_{12}$, $\tilde{y}_2 = y_{21} + iy_{22}$, $\tilde{y}_3 = y_{31} + iy_{32}$, y_{1j} , y_{2j} , y_{3j} , $j = 1, 2$, being real and $i = \sqrt{-1}$,

$$Q = |\tilde{y}_1|^2 + |\tilde{y}_2|^2 + |\tilde{y}_3|^2 = (y_{11}^2 + y_{12}^2) + (y_{21}^2 + y_{22}^2) + (y_{31}^2 + y_{32}^2). \quad (ii)$$

Thus, the required density, denoted by $\tilde{g}_{12}(\tilde{Y})$ as in (7.4a.7), is given by

$$\tilde{g}_{12}(\tilde{Y}) = \tilde{c}Q^2(1 + 3Q)^{-7}, \quad -\infty < y_{ij} < \infty, \quad i = 1, 2, 3, \quad j = 1, 2.$$

where \tilde{c} is specified in (i) and Q , in (ii). This completes the solution.

The density in (7.4a.6) is called a complex multivariate type-2 beta density in the general form and (7.4a.7) is referred to as a complex multivariate type-2 beta density in its standard form. Observe that these constitute only one form of the multivariate case of a type-2 beta density. When extending a univariate function to a multivariate one, there is no such thing as a unique multivariate analogue. There exist a multitude of multivariate functions corresponding to specified marginal functions, or marginal densities in statistical problems. In the latter case for instance, there are countless possible copulas associated with some specified marginal distributions. Copulas actually encapsulate the various dependence relationships existing between random variables. We have already seen that one set of generalizations to the multivariate case for univariate type-1 and type-2 beta densities are the type-1 and type-2 Dirichlet densities and their extensions. The densities appearing in (7.4a.6) and (7.4a.7) are yet another version of a multivariate type-2 beta density in the complex case.

What will be the resulting distribution when $q = 1$ in (7.4a.7)? The standard form of this density then becomes the following, denoted by $\tilde{g}_{13}(\tilde{y}_1)$:

$$\tilde{g}_{13}(\tilde{y}_1) = b^{\gamma+1} \frac{1}{\pi} \frac{\Gamma(\gamma + \beta + 1)}{\Gamma(\gamma + 1)\Gamma(\beta)} [|\tilde{y}_1|^2]^\gamma [1 + b|\tilde{y}_1|^2]^{-(\gamma+1+\beta)} \quad (7.4a.8)$$

for $\Re(\beta) > 0$, $\Re(\gamma + 1) > 0$, $b > 0$, and $\tilde{g}_{13} = 0$ elsewhere. We now verify that this is indeed a density function. Let $\tilde{y}_1 = y_{11} + iy_{12}$, y_{11} and y_{12} being real scalar quantities and $i = \sqrt{-1}$. When \tilde{y}_1 is in the complex plane, $-\infty < y_{1j} < \infty$, $j = 1, 2$. Let us make a polar coordinate transformation. Letting $y_{11} = r \cos \theta$ and $y_{12} = r \sin \theta$, $dy_{11} \wedge dy_{12} = r dr \wedge d\theta$, $0 \leq r < \infty$, $0 < \theta \leq 2\pi$. The integral over the functional part of (7.4a.8) yields

$$\begin{aligned}
\int_{\tilde{y}_1} |\tilde{y}_1|^{2\gamma} [1 + |\tilde{y}_1|^2]^{-(\gamma+1+\beta)} d\tilde{y}_1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y_{11}^2 + y_{12}^2]^\gamma \\
&\quad \times [1 + b(y_{11}^2 + y_{12}^2)]^{-(\gamma+1+\beta)} dy_{11} \wedge dy_{12} \\
&= \int_{\theta=0}^{2\pi} \int_{r=0}^{\infty} [r^2]^\gamma [1 + br^2]^{-(\gamma+1+\beta)} r d\theta \wedge dr \\
&= (2\pi) \left(\frac{1}{2}\right) \int_{t=0}^{\infty} t^\gamma (1 + bt)^{-(\gamma+1+\beta)} dt
\end{aligned}$$

which is equal to

$$\pi b^{-(\gamma+1)} \frac{\Gamma(\gamma+1)\Gamma(\beta)}{\Gamma(\gamma+1+\beta)}.$$

This establishes that the function specified by (7.4a.8) is a density.

7.4a.2. Arbitrary moments in the complex type-2 beta density

Let us consider the h -th moment of $|\det(\tilde{U})| = |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|$ in (7.4a.1). Since the only change upon integration is that γ is replaced by $\gamma+h$, the h -th moment is available from the normalizing constant in (7.4a.2):

$$E[|\det(\tilde{U})|^h] = \frac{\Gamma_p(\gamma+q+h)}{\Gamma_p(\gamma+q)} \frac{\Gamma_p(\beta-h)}{\Gamma_p(\beta)} \quad (7.4a.9)$$

$$= \prod_{j=1}^p \frac{\Gamma(\gamma+q-(j-1)+h)}{\Gamma(\gamma+q-(j-1))} \frac{\Gamma(\beta-(j-1)-h)}{\Gamma(\beta-(j-1))} \quad (7.4a.10)$$

$$= E[u_1^h] \cdots E[u_p^h] \quad (7.4a.11)$$

where u_1, \dots, u_p mutually independently distributed real scalar type-2 beta random variables with the parameters $(\gamma+q-(j-1), \beta-(j-1))$, $j = 1, \dots, p$. Thus, $|\det(\tilde{U})|$ has the structural representation

$$|\det(\tilde{U})| = |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})| = |\det(\tilde{Y} \tilde{Y}^*)| = |\det(\tilde{S})| = u_1 \cdots u_p \quad (7.4a.12)$$

where the u_1, \dots, u_p are as previously defined. The density for a complex scalar type-2 beta random variable is provided in (7.4a.8).

7.4a.3. A pathway version of the complex rectangular matrix-variate type-1 beta density

Consider the model specified in (7.2a.12), that is,

$$\tilde{g}_{14}(\tilde{X}) = \tilde{C}_4 |\det(A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^\gamma |\det(I - a(1-\alpha)A^{\frac{1}{2}} \tilde{X} B \tilde{X}^* A^{\frac{1}{2}})|^{\frac{\eta}{1-\alpha}} \quad (7.4a.13)$$

for $a > 0$, $\alpha < 1$, $A > O$, $B > O$, $\eta > 0$, $I - a(1 - \alpha)A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}} > O$, and $\tilde{g}_{14} = 0$ elsewhere, where \tilde{C}_4 is the normalizing constant given in (7.2a.15). When $\alpha < 1$, the model appearing in (7.4a.13) is a generalization of the complex rectangular matrix-variate type-1 beta model considered in (7.3a.1). When $\alpha > 1$, we write $1 - \alpha = -(\alpha - 1)$, $\alpha > 1$ and re-express \tilde{g}_{14} as

$$\tilde{g}_{15}(\tilde{X}) = \tilde{C}_5 |\det(A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})|^\gamma |\det(I + a(\alpha - 1)A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})|^{-\frac{\eta}{\alpha-1}} \quad (7.4a.14)$$

for $a > 0$, $\alpha > 1$, $\eta > 0$, $A > O$, $B > O$, and $\tilde{g}_{15} = 0$ elsewhere, where the normalizing constant \tilde{C}_5 is the same as the one in (7.2a.16). Observe that the model in (7.4a.14) is a generalization of the complex rectangular matrix-variate type-2 beta model in (7.4.1). When $q \rightarrow 1$, the models in (7.4a.13) and (7.4a.14) both converge to the following model:

$$\tilde{g}_{16}(\tilde{X}) = \tilde{C}_6 |\det(A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})|^\gamma e^{-a\eta \operatorname{tr}(A^{\frac{1}{2}}\tilde{X}B\tilde{X}^*A^{\frac{1}{2}})} \quad (7.4a.15)$$

for $a > 0$, $\eta > 0$, $A > O$, $B > O$, and $\tilde{g}_{16} = 0$ elsewhere, where the normalizing constant \tilde{C}_6 is the same as that in (7.2a.17). The model specified in (7.4a.15) is a generalization of the complex rectangular matrix-variate gamma model considered in (7.2a.1). Thus, model in (7.4a.13) contains all the three models (7.4a.13), (7.4a.14), and (7.4a.15), which are generalizations of the models given in (7.3a.1), (7.4a.1), and (7.2a.1), respectively. The pathway model in the complex domain, namely (7.4a.13), was introduced in Mathai and Provost (2006). Additional properties of the pathway model have already been discussed in Sect. 7.2a.

Exercises 7.4

- 7.4.1. Following the instructions or otherwise, derive the normalizing constant \tilde{C}_3 in (7.4a.3).
- 7.4.2. By integrating over \tilde{Y} , show that (7.4a.4) is a density.
- 7.4.3. Evaluate the normalizing constant in (7.4a.7) by using (1): Theorem 4.2a.3; (2): a $(2n)$ -variate real polar coordinate transformation.
- 7.4.4. Given the standard real matrix-variate type-2 beta model in (7.4.5), evaluate the marginal joint density of y_1, \dots, y_r , $r < p$.
- 7.4.5. Evaluate the density in (7.4a.4) explicitly for $p = 1$ and $q = 2$.

7.4.6. Given the standard complex matrix-variate type-2 beta model in (7.4a.7), evaluate the joint marginal density of $\tilde{y}_1, \dots, \tilde{y}_r$, $r < p$.

7.4.7. Derive the density of $|\det(\tilde{U})|$ in (7.4a.12) for the cases (1): $p = 1$; (2): $p = 2$.

7.4.8. Derive the density of $|U|$ in (7.4.10) for (1): $p = 2$; (2): $p = 3$.

7.5,7.5a. Ratios Involving Rectangular Matrix-Variate Random Variables

Since scalar variables such as type-1 beta, type-2 beta, F, Student- t and Cauchy variables are all associated with ratios of independently distributed random variables, we will explore ratios involving rectangular matrix-variate random variables. Such ratios will yield the rectangular matrix-variate versions of the aforementioned ratios of scalar variables. Let the $p \times n_1$, $p \leq n_1$, full rank matrix X_1 and the $p \times n_2$, $p \leq n_2$, full rank matrix X_2 be independently distributed real matrix-variate random variables having the rectangular matrix-variate gamma densities specified in (7.2.5), that is,

$$f_j(X_j) = \frac{|A_j|^{\frac{n_j}{2}} |B_j|^{\frac{p}{2}} \Gamma_p(\frac{n_j}{2})}{\pi^{\frac{n_j p}{2}} \Gamma_p(\gamma_j + \frac{n_j}{2})} |A_j X_j B_j X_j'|^{\gamma_j} e^{-\text{tr}(A_j X_j B_j X_j')}, \quad j = 1, 2, \quad (7.5.1)$$

for $A_j > O$, $B_j > O$, $\Re(\gamma_j + \frac{n_j}{2}) > \frac{p-1}{2}$, and $n_j \geq p$ where A_j is $p \times p$ and B_j is $n_j \times n_j$. Then, owing to the statistical independence of the variables, the joint density of X_1 and X_2 is $f(X_1, X_2) = f_1(X_1) f_2(X_2)$. Consider the ratios

$$U_1 = \left(\sum_{j=1}^2 \left(A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} \right) \right)^{-\frac{1}{2}} \left(A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}} \right) \left(\sum_{j=1}^2 \left(A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} \right) \right)^{-\frac{1}{2}} \quad (i)$$

and

$$U_2 = \left(A_2^{\frac{1}{2}} X_2 B_2 X_2' A_2^{\frac{1}{2}} \right)^{-\frac{1}{2}} \left(A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}} \right) \left(A_2^{\frac{1}{2}} X_2 B_2 X_2' A_2^{\frac{1}{2}} \right)^{-\frac{1}{2}}. \quad (ii)$$

Let us derive the densities of U_1 and U_2 . Letting $V_j = A_j^{\frac{1}{2}} X_j B_j^{\frac{1}{2}}$, we have $dX_j = |A_j|^{-\frac{n_j}{2}} |B_j|^{-\frac{p}{2}} dV_j$. Denoting the joint density of V_1 and V_2 by $g(V_1, V_2)$, it follows that $f(X_1, X_2) dX_1 \wedge dX_2 = g(V_1, V_2) dV_1 \wedge dV_2$ and so,

$$g(V_1, V_2) dV_1 \wedge dV_2 = \left\{ \prod_{j=1}^2 \frac{\Gamma_p(\frac{n_j}{2})}{\pi^{\frac{n_j p}{2}} \Gamma_p(\gamma_j + \frac{n_j}{2})} \right\} |V_1 V_1'|^{\gamma_1} |V_2 V_2'|^{\gamma_2} e^{-\text{tr}(V_1 V_1' + V_2 V_2')} dV_1 \wedge dV_2.$$

Letting $W_j = V_j V_j'$, $dV_j = \frac{\pi^{\frac{n_j p}{2}}}{\Gamma_p(\frac{n_j}{2})} |W_j|^{\frac{n_j}{2} - \frac{p+1}{2}} dW_j$, and the joint density of W_1 and W_2 , denoted by $h(W_1, W_2)$, is the following:

$$h(W_1, W_2) = \left\{ \prod_{j=1}^2 \frac{1}{\Gamma_p(\gamma_j + \frac{n_j}{2})} \right\} |W_1|^{\gamma_1 + \frac{n_1}{2} - \frac{p+1}{2}} |W_2|^{\gamma_2 + \frac{n_2}{2} - \frac{p+1}{2}} e^{-\text{tr}(W_1 + W_2)}. \quad (7.5.2)$$

Note that $U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}}$ and $U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$. Then, given the relationship between independently distributed matrix-variate gamma variables and a type-1 matrix-variate beta variable and a type-2 matrix-variate beta variable, U_1 and U_2 are distributed as real matrix-variate type-1 beta and type-2 beta random variables, respectively, both with the parameters $(\gamma_1 + \frac{n_1}{2}, \gamma_2 + \frac{n_2}{2})$, that is,

$$U_1 \sim \text{type-1 beta}\left(\gamma_1 + \frac{n_1}{2}, \gamma_2 + \frac{n_2}{2}\right) \text{ and } U_2 \sim \text{type-2 beta}\left(\gamma_1 + \frac{n_1}{2}, \gamma_2 + \frac{n_2}{2}\right).$$

Thus, we have the following result:

Theorem 7.5.1. *Let X_1 of dimension $p \times n_1$, $p \leq n_1$, and X_2 of dimension $p \times n_2$, $p \leq n_2$, be rank p matrices that are independently distributed rectangular real matrix-variate gamma random variables whose densities are specified in (7.5.1). Then, as defined in (i) and (ii), U_1 and U_2 are respectively real matrix-variate type-1 beta and type-2 beta distributed with the same parameters $(\gamma_1 + \frac{n_1}{2}, \gamma_2 + \frac{n_2}{2})$. Thus they have the following densities, denoted by $g_j(U_j)$, $j = 1, 2$:*

$$g_1(U_1) dU_1 = c |U_1|^{\gamma_1 + \frac{n_1}{2} - \frac{p+1}{2}} |I - U_1|^{\gamma_2 + \frac{n_2}{2} - \frac{p+1}{2}} dU_1, \quad O < U_1 < I, \quad (7.5.3)$$

and zero elsewhere, and

$$g_2(U_2) dU_2 = c |U_2|^{\gamma_1 + \frac{n_1}{2} - \frac{p+1}{2}} |I + U_2|^{-(\gamma_1 + \gamma_2 + \frac{n_1}{2} + \frac{n_2}{2})} dU_2, \quad U_2 > O, \quad (7.5.4)$$

where

$$c = \frac{\Gamma_p(\gamma_1 + \gamma_2 + \frac{n_1}{2} + \frac{n_2}{2})}{\Gamma_p(\gamma_1 + \frac{n_1}{2}) \Gamma_p(\gamma_2 + \frac{n_2}{2})}, \quad \Re\left(\gamma_j + \frac{n_j}{2}\right) > \frac{p-1}{2}, \quad j = 1, 2.$$

Analogous derivations will yield the densities of \tilde{U}_1 and \tilde{U}_2 , the corresponding matrix-variate random variables in the complex domain:

Theorem 7.5a.1. Let \tilde{X}_1 of dimension $p \times n_1$, $p \leq n_1$, and \tilde{X}_2 of dimension $p \times n_2$, $p \leq n_2$, be full rank rectangular matrix-variate complex gamma random variables that are independently distributed whose densities are

$$\tilde{f}_j(\tilde{X}_j)d\tilde{X}_j = \frac{|A_j|^{n_j}|B_j|^p \tilde{\Gamma}_p(n_j)}{\pi^{n_j p} \tilde{\Gamma}_p(\gamma_j + n_j)} |\det(A_j \tilde{X}_j B_j \tilde{X}_j^*)|^{\gamma_j} e^{-\text{tr}(A_j \tilde{X}_j B_j \tilde{X}_j^*)} d\tilde{X}_j \quad (7.5a.1)$$

where $A_j = A_j^* > O$ and $B_j = B_j^* > O$ with A_j being $p \times p$ and B_j , $n_j \times n_j$, $p \leq n_j$, $j = 1, 2$. Letting

$$\tilde{U}_1 = \left(\sum_{j=1}^2 A_j^{\frac{1}{2}} \tilde{X}_j B_j \tilde{X}_j^* A_j^{\frac{1}{2}} \right)^{-\frac{1}{2}} (A_1^{\frac{1}{2}} \tilde{X}_1 B_1 \tilde{X}_1^* A_1^{\frac{1}{2}}) \left(\sum_{j=1}^2 A_j^{\frac{1}{2}} \tilde{X}_j B_j \tilde{X}_j^* A_j^{\frac{1}{2}} \right)^{-\frac{1}{2}}$$

and

$$\tilde{U}_2 = (A_2^{\frac{1}{2}} \tilde{X}_2 B_2 \tilde{X}_2^* A_2^{\frac{1}{2}})^{-\frac{1}{2}} (A_1^{\frac{1}{2}} \tilde{X}_1 B_1 \tilde{X}_1^* A_1^{\frac{1}{2}}) (A_2^{\frac{1}{2}} \tilde{X}_2 B_2 \tilde{X}_2^* A_2^{\frac{1}{2}})^{-\frac{1}{2}}, \quad (7.5a.2)$$

the densities of \tilde{U}_1 and \tilde{U}_2 , denoted by $\tilde{g}_j(\tilde{U}_j)$, $j = 1, 2$, are respectively given by

$$\tilde{g}_1(\tilde{U}_1)d\tilde{U}_1 = \tilde{c} |\det(\tilde{U}_1)|^{\gamma_1 + n_1 - p} |\det(I - \tilde{U}_1)|^{\gamma_2 + n_2 - p} d\tilde{U}_1, \quad O < \tilde{U}_1 < I, \quad (7.5a.3)$$

and

$$\tilde{g}_2(\tilde{U}_2)d\tilde{U}_2 = \tilde{c} |\det(\tilde{U}_2)|^{\gamma_1 + n_1 - p} |\det(I + \tilde{U}_2)|^{-(\gamma_1 + \gamma_2 + n_1 + n_2)} d\tilde{U}_2, \quad \tilde{U}_2 > O, \quad (7.5a.4)$$

where

$$\tilde{c} = \frac{\tilde{\Gamma}_p(\gamma_1 + \gamma_2 + n_1 + n_2)}{\tilde{\Gamma}_p(\gamma_1 + n_1) \tilde{\Gamma}_p(\gamma_2 + n_2)}, \quad \Re(\gamma_j + n_j) > p - 1, \quad j = 1, 2.$$

The densities specified in (7.5.4) and (7.5a.4) happen to be quite useful in real-life applications. Connections of the type-2 beta distribution to the F-distribution, the Student- t^2 distribution and the distribution of the sample correlation coefficient when the population is Gaussian, have already been pointed out in the course of our previous discussions with respect to the scalar, vector variable and matrix-variate cases. Some further relationships are next pointed out. Let $\{Y_1, \dots, Y_n\}$ constitutes a simple random sample where $Y_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$, and the sample matrix be denoted by $\mathbf{Y} = [Y_1, Y_2, \dots, Y_n]$; letting $\bar{Y} = \frac{1}{n}[Y_1 + \dots + Y_n]$ and the matrix of sample means be $\bar{\mathbf{Y}} = [\bar{Y}, \dots, \bar{Y}]$, the sample sum of products (corrected) matrix is $S = (\mathbf{Y} - \bar{\mathbf{Y}})(\mathbf{Y} - \bar{\mathbf{Y}})'$, which is unaffected by μ . We have determined that S follows a real Wishart distribution having $m = n - 1$ degrees of freedom, and that when μ is

known to be a null vector, $\mathbf{Y}\mathbf{Y}'$ is real Wishart matrix with n degrees of freedom. Now, consider $A^{\frac{1}{2}}Y_j \stackrel{iid}{\sim} N_p(A^{\frac{1}{2}}\mu, A^{\frac{1}{2}}\Sigma A^{\frac{1}{2}})$; when $\mu = O$, the sample sum of products matrix is $A^{\frac{1}{2}}YY'A^{\frac{1}{2}}$, which can be expressed in the form of the product of matrices appearing in (7.5.1) with $B = I$. Hence, we can regard $A^{\frac{1}{2}}XBX'A^{\frac{1}{2}}$ (or equivalently $AXBX'$ in the determinant in (7.5.1)) as a weighted sample sum of products matrix with sample sizes n_1 and n_2 . Then, the type-2 beta density in (7.5.4) with U_2 replaced by $\frac{n_1}{n_2}U_2$ corresponds to a generalized real rectangular matrix-variate F -density having n_1 and n_2 degrees of freedom where U_2 is as defined in (ii). Moreover, for $\gamma_1 = 0 = \gamma_2$, this density will correspond to a rectangular matrix-variate Student- t density. The material included in Sect. 7.5, 7.5a may not be available in the literature.

7.5.1. Multivariate F , Student- t and Cauchy densities

The densities appearing in (7.5.4) and (7.5a.4) for the $p \times p$ positive definite matrices U_2 and \tilde{U}_2 have dU_2 and $d\tilde{U}_2$ as differential elements. A positive definite matrix such as U_2 , can be expressed as $U_2 = TT'$ where T of dimension $p \times n_1$, $p \leq n_1$, has rank p , and we can write dU_2 in terms of dT . We can also consider the format $U_2 = TCT'$ where $C > O$ is an $n_1 \times n_1$ positive definite constant matrix. In other words, we can arrive at the format in (7.4.1) from (7.5.4), and correspondingly obtain (7.4a.1) from (7.5a.4). Let us re-examine the expressions given in (7.4.1) and (7.4a.1), which could be referred to as rectangular matrix-variate F and Student- t densities in the real and complex cases for specific values of the parameters β and γ . Now, let $p = 1$ and $A = a > 0$ in (7.4.1) wherein a location parameter vector μ is inserted. The resulting density is

$$h(X)dX = \frac{a^{\gamma+\frac{q}{2}}|B|^{\frac{1}{2}}\Gamma(\frac{q}{2})\Gamma(\gamma+\frac{q}{2}+\beta)}{\pi^{\frac{q}{2}}\Gamma(\gamma+\frac{q}{2})\Gamma(\beta)}[(X-\mu)B(X-\mu)']^\gamma \times [1+a(X-\mu)B(X-\mu)']^{-(\gamma+\frac{q}{2}+\beta)}dX \tag{7.5.5}$$

where X and μ are $1 \times q$ row vectors, the corresponding density in the complex domain, denoted by $\tilde{h}(\tilde{X})$, being the following:

$$\tilde{h}(\tilde{X})d\tilde{X} = \frac{|a|^{\gamma+q}|\det(B)|\Gamma(q)\Gamma(\gamma+q+\beta)}{\pi^q\Gamma(\gamma+q)\Gamma(\beta)}[(\tilde{X}-\tilde{\mu})B(\tilde{X}-\tilde{\mu})^*]^\gamma \times [1+a(\tilde{X}-\tilde{\mu})B(\tilde{X}-\tilde{\mu})^*]^{-(\gamma+q+\beta)}d\tilde{X}. \tag{7.5a.5}$$

For specific values of the parameters, the densities appearing in (7.5.5) and (7.5a.5) can be respectively called the multivariate F and Student- t densities in the real and complex domains. With a view to model certain types of signal processes, (Kondo et al., 2020) made use of a special form of the complex multivariate Student- t wherein $\gamma = 0$, $a = \frac{2}{\nu}$ and $\beta = \frac{\nu}{2}$, which is given next.

7.5a.1. A complex multivariate Student- t having ν degrees of freedom

$$\tilde{h}_1(\tilde{X})d\tilde{X} = \frac{2^q \Gamma(\frac{\nu}{2} + q)}{(\nu\pi)^q \Gamma(\frac{\nu}{2}) |\det(\Sigma)|} \left[1 + \frac{2}{\nu} (\tilde{X} - \tilde{\mu}) \Sigma^{-1} (\tilde{X} - \tilde{\mu})^* \right]^{-(\frac{\nu}{2} + q)} d\tilde{X}. \quad (7.5a.6)$$

A complex multivariate Cauchy density that, as well, is mentioned in Kondo et al. (2020), can be obtained by letting $\nu = 1$ in (7.5a.6). We conclude this section with its representation, denoted by $\tilde{h}_2(\tilde{X})$:

7.5a.2. A complex multivariate Cauchy density

$$\tilde{h}_2(\tilde{X})d\tilde{X} = \frac{2^q \Gamma(\frac{1}{2} + q)}{\pi^q \Gamma(\frac{1}{2}) |\det(\Sigma)|} [1 + 2(\tilde{X} - \tilde{\mu}) \Sigma^{-1} (\tilde{X} - \tilde{\mu})^*]^{-(q + \frac{1}{2})} d\tilde{X}. \quad (7.5a.7)$$

Exercises 7.5

7.5.1. Derive the complex densities in (7.5a.3) and (7.5a.4).

7.5.2. Derive the normalizing constant in (7.5a.6) by integrating out the functional portion of this density.

7.5.3. Derive the normalizing constant in (7.5a.7) by integrating out the functional portion of this density.

7.5.4. Derive the density in (7.5a.6) from complex q -variate Gaussian densities.

7.5.5. Derive the density in (7.5a.7) from complex q -variate Gaussian densities.

7.6. Rectangular Matrix-Variate Dirichlet Density, Real Case

For the real matrix-variate type-1 and type-2 Dirichlet models involving sets of real positive definite matrices, the reader is referred to Sects. 5.8.6 and 5.8.7. The corresponding rectangular matrix-variate cases will be considered in this section. Let $A_j > O$, $j = 1, \dots, k$, be $p \times p$ real positive definite constant matrices, and B_j , $j = 1, \dots, k$, be $q_j \times q_j$ real positive definite constant matrices. Let X_j , $j = 1, \dots, k$, be $p \times q_j$, $q_j \geq p$,

rank p real matrices whose elements are distinct real scalar variables. Then, consider the real-valued scalar function of X_1, \dots, X_k ,

$$f_1(X_1, \dots, X_k) = C_k |A_1 X_1 B_1 X_1'|^{\gamma_1} \cdots |A_k X_k B_k X_k'|^{\gamma_k} \times |I - A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}} - \cdots - A_k^{\frac{1}{2}} X_k B_k X_k' A_k^{\frac{1}{2}}|^{\gamma_{k+1} - \frac{p+1}{2}} \quad (7.6.1)$$

for $A_j > O$, $B_j > O$, $A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} > O$, $j = 1, \dots, k$, $I - \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} > O$, $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, and $f_1 = 0$ elsewhere, where C_k is the normalizing constant. This normalizing constant can be evaluated as follows: Letting

$$Y_j = A_j^{\frac{1}{2}} X_j B_j^{\frac{1}{2}} \Rightarrow dY_j = |A_j|^{\frac{q_j}{2}} |B_j|^{\frac{p}{2}} dX_j, \quad j = 1, \dots, k, \quad (i)$$

the joint density of Y_1, \dots, Y_k , denoted by $f_2(Y_1, \dots, Y_k)$, is given by

$$f_2(Y_1, \dots, Y_k) = \left\{ \prod_{j=1}^k |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} \right\} C_k |Y_1 Y_1'|^{\gamma_1} \cdots |Y_k Y_k'|^{\gamma_k} \times |I - \sum_{j=1}^k Y_j Y_j'|^{\gamma_{k+1} - \frac{p+1}{2}}. \quad (7.6.2)$$

Now, let

$$S_j = Y_j Y_j' \Rightarrow dY_j = \frac{\pi^{\frac{q_j p}{2}}}{\Gamma_p(\frac{q_j}{2})} |S_j|^{\frac{q_j}{2} - \frac{p+1}{2}} dS_j, \quad j = 1, \dots, k. \quad (ii)$$

Then, the joint density of S_1, \dots, S_k , which follows, is a real matrix-variate type-1 Dirichlet density:

$$f_3(S_1, \dots, S_k) = C_k \left\{ \prod_{j=1}^k |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} \frac{\pi^{\frac{q_j p}{2}}}{\Gamma_p(\frac{q_j}{2})} \right\} \times \left\{ \prod_{j=1}^k |S_j|^{\gamma_j + \frac{q_j}{2} - \frac{p+1}{2}} \right\} |I - S_1 - \cdots - S_k|^{\gamma_{k+1} - \frac{p+1}{2}} \quad (7.6.3)$$

for $S_j > O$, $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$. Next, on integrating out S_1, \dots, S_k , by making use of a type-1 real matrix-variate Dirichlet integral that was defined in Sect. 5.8.6, we have

$$\frac{\{\prod_{j=1}^k \Gamma_p(\gamma_j + \frac{q_j}{2})\} \Gamma_p(\gamma_{k+1})}{\Gamma_p(\sum_{j=1}^{k+1} \gamma_j + \sum_{j=1}^k \frac{q_j}{2})}, \quad \Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}, \quad j = 1, \dots, k, \quad (iii)$$

and $\Re(\gamma_{k+1}) > \frac{p-1}{2}$. Then, as obtained from (i), (ii) and (iii), the normalizing constant is

$$C_k = \left\{ \prod_{j=1}^k |A_j|^{\frac{q_j}{2}} |B_j|^{\frac{p}{2}} \frac{\Gamma_p(\frac{q_j}{2})}{\pi^{\frac{q_j p}{2}}} \frac{1}{\Gamma_p(\gamma_j + \frac{q_j}{2})} \right\} \\ \times \frac{\Gamma_p(\gamma_1 + \cdots + \gamma_{k+1} + \frac{q_1}{2} + \cdots + \frac{q_k}{2})}{\{\prod_{j=1}^k \Gamma_p(\gamma_j + \frac{q_j}{2})\} \Gamma_p(\gamma_{k+1})} \quad (7.6.4)$$

for $A_j > O$, $B_j > O$, $q_j \geq p$, $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, and $\Re(\gamma_{k+1}) > \frac{p-1}{2}$.

7.6.1. Certain properties, real rectangular matrix-variate type-1 Dirichlet density

Letting $U = \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}$, what might be the distributions of U and $I - U$? In the real scalar case, one could have easily evaluated the moments $E[(1-u)^h]$ for arbitrary h , which would have automatically determined the distribution of $1-u$, and therefrom that of u . In the matrix-variate case as well, one can readily determine the h -th moment of the determinant of $I - U$, $E[|I - U|^h]$, and the unique resulting distribution. However, the distribution of a determinant being unique does not imply that the distribution of the corresponding matrix is unique. Thus, we have to resort to other approaches for obtaining the distributions of U and $I - U$. Consider the following transformation:

$$V_j = A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}, \quad j = 1, \dots, k-1, \quad V_k = \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} = U.$$

Then $A_k^{\frac{1}{2}} X_k B_k X_k' A_k^{\frac{1}{2}} = U - V_1 - \cdots - V_{k-1}$ and $I - \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}} = I - V_k = I - U$. Noting that

$$dX_1 \wedge \cdots \wedge dX_k = \left\{ \prod_{j=1}^k |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} \frac{\pi^{\frac{q_j p}{2}}}{\Gamma_p(\frac{q_j}{2})} \right\} dV_1 \wedge \cdots \wedge dV_{k-1} \wedge dU, \quad (7.6.5)$$

the joint density of V_1, \dots, V_{k-1}, U , denoted by $f_3(V_1, \dots, V_{k-1}, U)$, is seen to be

$$f_3(V_1, \dots, V_{k-1}, U) = \frac{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1})}{\{\prod_{j=1}^k \Gamma_p(\gamma_j + \frac{q_j}{2})\} \Gamma_p(\gamma_{k+1})} \left\{ \prod_{j=1}^{k-1} |V_j|^{\gamma_j + \frac{q_j}{2} - \frac{p+1}{2}} \right\} \\ \times |U - V_1 - \cdots - V_{k-1}|^{\gamma_k + \frac{q_k}{2} - \frac{p+1}{2}} |I - U|^{\gamma_{k+1} - \frac{p+1}{2}}, \quad (7.6.6)$$

where

$$|U - V_1 - \cdots - V_{k-1}| = |U| |I - U^{-\frac{1}{2}} V_1 U^{-\frac{1}{2}} - \cdots - U^{-\frac{1}{2}} V_{k-1} U^{-\frac{1}{2}}|.$$

Letting $W_j = U^{-\frac{1}{2}} V_j U^{-\frac{1}{2}}$, $j = 1, \dots, k - 1$, for fixed U we have

$$dV_1 \wedge \dots \wedge dV_{k-1} = |U|^{(k-1)(\frac{p+1}{2})} dW_1 \wedge \dots \wedge dW_{k-1}.$$

Now, the joint density of W_1, \dots, W_{k-1} and U , denoted by $f_4(W_1, \dots, W_{k-1}, U)$, is the following:

$$f_4(W_1, \dots, W_{k-1}, U) = C'_k |U|^{\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) - \frac{p+1}{2}} |I - U|^{\gamma_{k+1} - \frac{p+1}{2}} \left\{ \prod_{j=1}^{k-1} |W_j|^{\gamma_j + \frac{q_j}{2} - \frac{p+1}{2}} \right\} \\ \times |I - W_1 - \dots - W_{k-1}|^{\gamma_k + \frac{q_k}{2} - \frac{p+1}{2}} \tag{7.6.7}$$

where C'_k is the normalizing constant. We then integrate out W_1, \dots, W_{k-1} by using a $(k - 1)$ -variate type-1 Dirichlet integral, this yielding the result:

$$\left\{ \prod_{j=1}^k \Gamma_p(\gamma_j + \frac{q_j}{2}) \right\} / \Gamma_p\left(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2})\right) \text{ for } \Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}, j = 1, \dots, k.$$

Accordingly, the marginal density of U is the following:

$$f_5(U) = \frac{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1})}{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2})) \Gamma_p(\gamma_{k+1})} |U|^{\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) - \frac{p+1}{2}} |I - U|^{\gamma_{k+1} - \frac{p+1}{2}} \tag{7.6.8}$$

for $O < U < I$, $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, $\Re(\gamma_{k+1}) > \frac{p-1}{2}$, and $f_5 = 0$ elsewhere. Thus, U is a real matrix-variate type-1 beta with the parameters $(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}), \gamma_{k+1})$ and therefore that $I - U$ is a real matrix-variate type-1 beta with the parameters $(\gamma_{k+1}, \sum_{j=1}^k (\gamma_j + \frac{q_j}{2}))$. These results are now stated as a theorem.

Theorem 7.6.1. Consider the density given in (7.6.1). Let $U = \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}$. Then, U has a real matrix-variate type-1 beta distribution whose parameters are $(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}), \gamma_{k+1})$ and $I - U$ is distributed as a real matrix-variate type-1 beta with the parameters $(\gamma_{k+1}, \sum_{j=1}^k (\gamma_j + \frac{q_j}{2}))$.

The h -th moment of the determinant of the matrix $I - U$ can be evaluated either from Theorem 7.6.1 or from Eq. (7.6.1). This h -th moment of the determinant, which can be worked out from the normalizing constant appearing in (7.6.8), is

$$E[|I - U|^h] = \frac{\Gamma_p(\gamma_{k+1} + h)}{\Gamma_p(\gamma_{k+1})} \frac{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1})}{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1} + h)} \tag{7.6.9}$$

for $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $\Re(\gamma_{k+1}) > \frac{p-1}{2}$. Observe that a representation of the h -th moment of the determinant of U cannot be derived from (7.6.9). However, $E[|U|^h]$ can be readily evaluated from Theorem 7.6.1:

$$E[|U|^h] = \frac{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + h)}{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}))} \frac{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1})}{\Gamma_p(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1} + h)} \quad (7.6.10)$$

for $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, $\Re(\gamma_{k+1}) > \frac{p-1}{2}$. Upon expanding the $\Gamma_p(\cdot)$'s in terms of $\Gamma(\cdot)$'s, the following structural representations are obtained:

$$|I - U| = u_1 \cdots u_p, \quad (7.6.11)$$

$$|U| = v_1 \cdots v_p, \quad (7.6.12)$$

where u_1, \dots, u_p are independently distributed real scalar type-1 beta random variables with the parameters $(\gamma_{k+1} - \frac{j-1}{2}, \sum_{j=1}^k (\gamma_j + \frac{q_j}{2}))$, $j = 1, \dots, k$, and v_1, \dots, v_k are independently distributed real scalar type-1 beta random variables with the parameters $(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) - \frac{j-1}{2}, \gamma_{k+1})$, $j = 1, \dots, k$.

7.6.2. A multivariate version of the real matrix-variate type-1 Dirichlet density

For $p = 1$, consider the joint density of Y_1, \dots, Y_k in $f_2(Y_1, \dots, Y_k)$, which shall be denoted by $f_6(Y_1, \dots, Y_k)$. Then,

$$f_6(Y_1, \dots, Y_k) = \left\{ \prod_{j=1}^k \frac{\Gamma(\frac{q_j}{2})}{\pi^{\frac{q_j}{2}}} \right\} \frac{\Gamma(\sum_{j=1}^k (\gamma_j + \frac{q_j}{2}) + \gamma_{k+1})}{\left\{ \prod_{j=1}^k \Gamma(\gamma_j + \frac{q_j}{2}) \right\} \Gamma(\gamma_{k+1})} \left\{ \prod_{j=1}^k |Y_j Y_j'|^{\gamma_j} \right\} \\ \times |I - Y_1 Y_1' - \cdots - Y_k Y_k'|^{\gamma_{k+1} - \frac{p+1}{2}}, \quad (7.6.13)$$

the conditions on the parameters remaining as previously stated. Note that Y_j is of the form $Y_j = (y_{j1}, \dots, y_{jq_j})$, so that $Y_j Y_j' = y_{j1}^2 + \cdots + y_{jq_j}^2$. Thus, in light of its structure, the density appearing in (7.6.13) has interesting properties. For instance, it can be observed that all the subsets of Y_1, \dots, Y_k also have densities belonging to the same family. Accordingly, the marginal density of Y_1 is the following:

$$\begin{aligned}
 f_7(Y_1) &= \frac{\Gamma(\frac{q_1}{2})}{\pi^{\frac{q_1}{2}}} \frac{\Gamma(\gamma_1 + \frac{q_1}{2} + \gamma_2)}{\Gamma(\gamma_1 + \frac{q_1}{2})\Gamma(\gamma_2)} [y_{11}^2 + \dots + y_{1q_1}^2]^{\gamma_1} \\
 &\times [1 - y_{11}^2 - \dots - y_{1q_1}^2]^{\gamma_2 + \gamma_1 + \frac{q_1}{2} - \frac{p+1}{2}}
 \end{aligned} \tag{7.6.14}$$

for $-\infty < y_{1r} < \infty$, $r = 1, \dots, q_1$, $0 < y_{11}^2 + \dots + y_{1q_1}^2 < 1$, $\Re(\gamma_1 + \frac{q_1}{2}) > 0$, $\Re(\gamma_2) > 0$, and $f_7 = 0$ elsewhere. As has already been mentioned, the structure in (7.6.14) is related to geometrical probability problems involving type-1 beta distributed isotropic random points. Thus, (7.6.13) suggests the possibility of generalizing such geometrical probability problems in connection with a type-1 Dirichlet density as the underlying density for the random points. This does not appear to have yet been discussed in the literature on geometrical probability.

The complex case of the type-1 rectangular matrix-variate Dirichlet density, the real and complex cases of the rectangular matrix-variate type-2 Dirichlet density and their generalized forms can be similarly handled; hence, they will not be further discussed. Certain of these cases are brought up in this section’s exercises.

Note 7.6.1. One could also consider a pathway version of the model appearing in Eq. (7.6.1). Let us replace the second line in (7.6.1) by

$$|I - a(1 - \alpha)(A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}} + \dots + A_k^{\frac{1}{2}} X_k B_k X_k' A_k^{\frac{1}{2}})|^{\frac{\eta}{1-\alpha} - \frac{p+1}{2}}$$

where $a > 0$, $\alpha < 1$, $\eta > 0$ are real scalar and γ_{k+1} by $\frac{\eta}{1-\alpha}$, and denote the resulting model by f_8 whose corresponding equation number will be referred to as (7.6.15). Observe that (7.6.15) belongs to a generalized type-1 Dirichlet family of models and that the new normalizing constant will be denoted C_{k1} . For $\alpha > 1$, write $-a(1-\alpha) = a(\alpha-1) > 0$, and then $\frac{\eta}{1-\alpha} = -\frac{\eta}{\alpha-1}$. Number the resulting model of (7.6.1) as f_9 , with (7.6.16) as the associated equation number. Note that (7.6.16) is actually a generalized type-2 Dirichlet model whose normalizing constant, denoted C_{k2} , will be different. Taking the limits as $\alpha \rightarrow 1_-$ in (7.6.15) and $\alpha \rightarrow 1_+$ in (7.6.16), both the models f_8 in (7.6.15) and f_9 in (7.6.16) will converge to a model f_{10} whose associated equation number will be (7.6.17), wherein the second line corresponding to the second line in (7.6.1) will be

$$e^{-a \eta \operatorname{tr}(A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}} + \dots + A_k^{\frac{1}{2}} X_k B_k X_k' A_k^{\frac{1}{2}})},$$

this limiting model having its own normalizing constant denoted by C_{k3} . As well, it can be established that, under the above limiting process, both C_{k1} and C_{k2} will converge to C_{k3} . Now, observe that the matrices X_1, \dots, X_k in model f_{10} are mutually independently distributed real rectangular matrix-variate gamma random variables. This turns out to be

an unforeseen result as, in this case, the pathway parameter α is also seen to control the dependence to independence transitional stages. Results analogous to those obtained in Sects. 7.6, 7.6.1, and 7.6.2 could similarly be derived within the complex domain.

Exercises 7.6

7.6.1. Construct, in the complex domain, the rectangular matrix-variate type-1 Dirichlet density corresponding to the density specified in (7.6.1) and determine the associated normalizing constant.

7.6.2. Establish, in the complex domain, a theorem corresponding to Theorem 7.6.1.

7.6.3. Establish, for the complex case, the structural representations corresponding to (7.6.11) and (7.6.12).

7.6.4. Construct a real rectangular matrix-variate type-2 Dirichlet density corresponding to the density in (7.6.1).

7.6.5. Construct a complex rectangular matrix-variate type-2 Dirichlet density corresponding to the density in (7.6.1).

7.6.6. When the $p \times q_j$ matrices X_j 's jointly have a real type-2 Dirichlet density with the parameter matrices $A_j > O$, $B_j > O$ as in (7.6.1) where A_j is $p \times p$, B_j is $q_j \times q_j$ and X_j is $p \times q_j$, $q_j \geq p$, $j = 1, \dots, k$, of full rank p , establish that $U = [I + \sum_{j=1}^k (A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}})]^{-1}$ has a real matrix-variate type-1 beta distribution and specify its parameters. What about the density of $\sum_{j=1}^k (A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}})$ in this case?

7.6.7. Answer the questions in Exercise 7.6.6 for the corresponding type-2 Dirichlet density in the complex domain, replacing X_j by \tilde{X}_j and X_j' by \tilde{X}_j^* .

7.6.8. For the real type-2 Dirichlet density in Exercise 7.6.4, determine $E[|U|^h]$ for U as specified in Exercise 7.6.6.

7.6.9. Extend all the results obtained in Sect. 7.6 to the complex domain.

7.6.10. Derive, in the complex domain, results that are analogous to those obtained for the real case in Note 7.6.1, while keeping a , η and α real.

7.7. Generalizations of the Real Rectangular Dirichlet Models

The first author and his collaborators have considered several types of generalizations to the type-1 and type-2 Dirichlet models for real positive definite matrices and Hermitian positive definite matrices. We will propose certain extensions of those results to rectangular matrix-variate cases, both in the real and complex domains. Again, let X_j be a

$p \times q_j$, $q_j \geq p$, matrix of full rank p having distinct real scalar variables as its elements, for $j = 1, \dots, k$. Let the constant real positive definite matrices $A_j > O$ and $B_j > O$, where A_j is $p \times p$ and B_j is $q_j \times q_j$, $j = 1, \dots, k$, be as defined in Sect. 7.6. Consider the real model

$$\begin{aligned}
 f_{11}(X_1, \dots, X_k) &= D_k |A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}}|^{\gamma_1} |I - A_1^{\frac{1}{2}} X_1 B_1 X_1' A_1^{\frac{1}{2}}|^{\beta_1} \\
 &\times |A_2^{\frac{1}{2}} X_2 B_2 X_2' A_2^{\frac{1}{2}}|^{\gamma_2} |I - \sum_{j=1}^2 A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}|^{\beta_2} \dots \\
 &\times |A_k^{\frac{1}{2}} X_k B_k X_k' A_k^{\frac{1}{2}}|^{\gamma_k} |I - \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}|^{\gamma_{k+1} + \beta_k - \frac{p+1}{2}} \quad (7.7.1)
 \end{aligned}$$

for $\Re(\gamma_j + \frac{q_j}{2}) > \frac{p-1}{2}$, $j = 1, \dots, k$, $\Re(\alpha_{k+1}) > \frac{p-1}{2}$, and other conditions to be specified later, where D_k is the normalizing constant. For evaluating the normalizing constant, consider the following transformations:

$$Z_j = A_j^{\frac{1}{2}} X_j B_j^{\frac{1}{2}} \Rightarrow dX_j = |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} dZ_j, \quad j = 1, \dots, k, \quad (i)$$

so that the model f_{11} changes to f_{12} where

$$\begin{aligned}
 f_{12}(Z_1, \dots, Z_k) &= D_k \left\{ \prod_{j=1}^k |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} \right\} |Z_1 Z_1'|^{\gamma_1} \\
 &\times |I - Z_1 Z_1'|^{\beta_1} |Z_2 Z_2'|^{\gamma_2} \\
 &\times |I - Z_1 Z_1' - Z_2 Z_2'|^{\beta_2} \dots |Z_k Z_k'|^{\gamma_k} |I - \sum_{j=1}^k Z_j Z_j'|^{\gamma_{k+1} + \beta_k - \frac{p+1}{2}}. \quad (7.7.2)
 \end{aligned}$$

Now, letting

$$Z_j Z_j' = S_j \Rightarrow dZ_j = \frac{\pi^{\frac{q_j p}{2}}}{\Gamma_p(\frac{q_j}{2})} |S_j|^{\frac{q_j}{2} - \frac{p+1}{2}} dS_j, \quad j = 1, \dots, k, \quad (ii)$$

the model becomes

$$\begin{aligned}
 f_{13}(S_1, \dots, S_k) &= \left\{ \prod_{j=1}^k |A_j|^{-\frac{q_j}{2}} |B_j|^{-\frac{p}{2}} \frac{\pi^{\frac{q_j p}{2}}}{\Gamma_p(\frac{q_j}{2})} \right\} |S_1|^{\gamma_1 + \frac{q_1}{2} - \frac{p+1}{2}} \\
 &\quad \times |I - S_1|^{\beta_1} |S_2|^{\gamma_2 + \frac{q_2}{2} - \frac{p+1}{2}} |I - S_1 - S_2|^{\beta_2} \dots \\
 &\quad \times |S_k|^{\gamma_k + \frac{q_k}{2} - \frac{p+1}{2}} |I - \sum_{j=1}^k S_j|^{\gamma_k + \beta_k - \frac{p+1}{2}}. \tag{7.7.3}
 \end{aligned}$$

Now, consider the transformation (5.8.20), namely,

$$\begin{aligned}
 S_1 &= Y_1 \\
 S_2 &= (I - Y_1)^{\frac{1}{2}} Y_2 (I - Y_1)^{\frac{1}{2}} \\
 S_j &= (I - Y_1)^{\frac{1}{2}} \dots (I - Y_{j-1})^{\frac{1}{2}} Y_j (I - Y_{j-1})^{\frac{1}{2}} \dots (I - Y_1)^{\frac{1}{2}} \tag{7.7.4}
 \end{aligned}$$

for $j = 2, \dots, k$. Then Y_1, \dots, Y_k will be independently distributed real matrix-variate type-1 beta random variables with the parameters $(\alpha_j = \gamma_j + \frac{q_j}{2}, \delta_j)$, $j = 1, \dots, k$, where

$$\delta_j = \gamma_{j+1} + \frac{q_{j+1}}{2} + \dots + \gamma_{k+1} + \frac{q_{k+1}}{2} + \beta_j + \dots + \beta_k, \quad j = 1, \dots, k, \quad \text{and } q_{k+1} = 0. \quad (iii)$$

The normalizing constant D_k is thus the following:

$$D_k = \left\{ \prod_{j=1}^k |A_j|^{\frac{q_j}{2}} |B_j|^{\frac{p}{2}} \frac{\Gamma_p(\frac{q_j}{2})}{\pi^{\frac{q_j p}{2}}} \right\} \frac{\Gamma_p(\sum_{j=1}^k (\delta_j + \alpha_j))}{\prod_{j=1}^k [\Gamma_p(\alpha_j) \Gamma_p(\delta_j)]} \tag{7.7.5}$$

for $\alpha_j > \frac{p-1}{2}$, $\delta_j > \frac{p-1}{2}$, $j = 1, \dots, k$, where the α_j 's and δ_j 's are as previously given. Properties parallel to those pointed out in Sects. 7.1–7.6 can also be studied for the model specified in (7.7.1). The marginal distributions of subsets of the matrices X_1, X_2, \dots, X_k , taken in the order, will belong to the same family of densities. There exist other generalizations of the type-1 and type-2 Dirichlet models. For all such generalizations, one can extend the results to the rectangular matrix-variate cases in both the real and complex domains.

Exercises 7.7

7.7.1. Develop the transformation corresponding to (7.7.4) for the real type-2 Dirichlet case. Specify the Jacobians of the transformation (7.7.4) and the corresponding transformation for the type-2 case.

7.7.2. Verify the result for δ_j in (iii) following (7.7.4) and develop the expression corresponding to δ_j for the type-2 Dirichlet case.

7.7.3. Derive the joint marginal density of X_1, \dots, X_r , $r < k$, by integrating out the matrices starting with X_k in (7.7.1).

7.7.4. Develop, in the complex domain, the model corresponding to (7.7.1) and derive its associated normalizing constant.

7.7.5. If possible, derive the density of $U = \sum_{j=1}^k A_j^{\frac{1}{2}} X_j B_j X_j' A_j^{\frac{1}{2}}$ where the X_j 's, $j = 1, \dots, k$, jointly have the density given in (7.7.1).

References

- N. Balakrishnan and C.D. Lai (2009): *Continuous Bivariate Distributions*, Springer, New York.
- C.A. Barnes, D.D. Clayton and D.N. Schramm (editors) (1982): *Essays in Nuclear Astrophysics*, Cambridge University Press, Cambridge.
- C.L. Critchfield (1972): In *Cosmology, Fusion and Other Matters*, George Gamow Memorial Volume (ed: F. Reines), University of Colorado Press, Colorado, pp.186–191.
- W.A. Fowler (1984): Experimental and Theoretical Nuclear Astrophysics: The Quest for the Origin of the Elements, *Review of Modern Physics*, **56**, 149–179.
- T. Kondo, K. Fukushige, N. Takamune, D. Kitamura, H. Saruwatari, R. Ikeshita and T. Nakatani (2020): Convergence-guaranteed independent positive semidefinite tensor analysis based on Student's-t distribution, *arXiv:2002.08582v1 [cs.LG]* 20 Feb 2020.
- A.W. Marshall and I. Olkin (1967): A multivariate exponential distribution, *Journal of the American Statistical Association*, **62**, 30–44.
- A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A.M. Mathai (1999): *An Introduction to Geometrical Probability, Distributional Aspects with Applications*, Gordon and Breach, Amsterdam.

A.M. Mathai (2005): A pathway to matrix-variate gamma and normal densities, *Linear Algebra and its Applications*, **396**, 317–328.

A.M. Mathai and H.J. Haubold (1988): *Modern Problems in Nuclear and Neutrino Astrophysics*, Akademie-Verlag, Berlin.

A.M. Mathai and T. Princy (2017): Analogues of reliability analysis for matrix-variate cases, *Linear Algebra and its Applications*, **532** (2017), 287–311.

A.M. Mathai and S.B. Provost (2006): Some complex matrix-variate statistical distributions on rectangular matrices, *Linear Algebra and its Applications*, **410**, 198–216.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.

A. Pais (1986): *Inward Bound of Matter and Forces in the Physical World*, Oxford University Press, Oxford.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 8

The Distributions of Eigenvalues and Eigenvectors



8.1. Introduction

We will utilize the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. Other notations will remain unchanged.

Our objective in this chapter is to examine the distributions of the eigenvalues and eigenvectors associated with a matrix-variate random variable. Letting W be such a $p \times p$ matrix-variate random variable, its determinant is the product of its eigenvalues and its trace, the sum thereof. Accordingly, the distributions of the determinant and the trace of W are available from the distributions of simple functions of its eigenvalues. Actually, several statistical quantities are associated with eigenvalues or eigenvectors. In order to delve into such problems, we will require certain additional properties of the matrix-variate gamma and beta distributions previously introduced in Chap. 5. As a preamble to the study of the distributions of eigenvalues and eigenvectors, these will be looked into in the next subsections for both the real and complex cases.

8.1.1. Matrix-variate gamma and beta densities, real case

Let W_1 and W_2 be statistically independently distributed $p \times p$ real matrix-variate gamma random variables whose respective parameters are (α_1, B) and (α_2, B) with $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, their common scale parameter matrix B being a real positive definite constant matrix. Then, the joint density of W_1 and W_2 , denoted by $f(W_1, W_2)$, is the following:

$$f(W_1, W_2) = \begin{cases} \frac{|B|^{\alpha_1+\alpha_2}}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |W_1|^{\alpha_1-\frac{p+1}{2}} |W_2|^{\alpha_2-\frac{p+1}{2}} e^{-\text{tr}(B(W_1+W_2))} \\ \quad B > O, W_j > O, \Re(\alpha_j) > \frac{p-1}{2}, j = 1, 2, \\ 0 \text{ elsewhere} \end{cases} \quad (8.1.1)$$

Consider the transformations

$$U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}} \text{ and } U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}, \quad (8.1.2)$$

which are matrix-variate counterparts of the changes of variables $u_1 = \frac{w_1}{w_1+w_2}$ and $u_2 = \frac{w_1}{w_2}$ in the real scalar case, that is, for $p = 1$. Note that the square roots in (8.1.2) are symmetric positive definite matrices. Then, we have the following result:

Theorem 8.1.1. *When the real matrices U_1 and U_2 are as defined in (8.1.2), then U_1 is distributed as a real matrix-variate type-1 beta variable with the parameters (α_1, α_2) and U_2 , as a real matrix-variate type-2 beta variable with the parameters (α_1, α_2) . Further, U_1 and $U_3 = W_1 + W_2$ are independently distributed, with U_3 having a real matrix-variate gamma distribution with the parameters $(\alpha_1 + \alpha_2, B)$.*

Proof: Given the joint density of W_1 and W_2 specified in (8.1.1), consider the transformation $(W_1, W_2) \rightarrow (U_3 = W_1 + W_2, U = W_1)$. On observing that its Jacobian is equal to one, the joint density of U_3 and U , denoted by $f_1(U_3, U)$, is obtained as

$$f_1(U_3, U) dU_3 \wedge dU = c |U|^{\alpha_1-\frac{p+1}{2}} |U_3 - U|^{\alpha_2-\frac{p+1}{2}} e^{-\text{tr}(U_3)} dU_3 \wedge dU \quad (i)$$

where

$$c = \frac{|B|^{\alpha_1+\alpha_2}}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)}. \quad (ii)$$

Noting that

$$|U_3 - U| = |U_3| |I - U_3^{-\frac{1}{2}} U U_3^{-\frac{1}{2}}|,$$

we now let $U_1 = U_3^{-\frac{1}{2}} U U_3^{-\frac{1}{2}}$ for fixed U_3 , so that $dU_1 = |U_3|^{-\frac{p+1}{2}} dU$. Accordingly, the joint density of U_3 and U_1 , denoted by $f_2(U_3, U_1)$, is the following, observing that U_1 is as defined in (8.1.2) with $W_1 = U$ and $U_3 = W_1 + W_2$:

$$f_2(U_3, U_1) = c |U_3|^{\alpha_1+\alpha_2-\frac{p+1}{2}} e^{-\text{tr}(U_3)} |U_1|^{\alpha_1-\frac{p+1}{2}} |I - U_1|^{\alpha_2-\frac{p+1}{2}} \quad (iii)$$

for $\Re(\alpha_1) > \frac{p-1}{2}$, $\Re(\alpha_2) > \frac{p-1}{2}$, $U_3 > O$, $O < U_1 < I$, and zero elsewhere. On multiplying and dividing (iii) by $\Gamma_p(\alpha_1 + \alpha_2)$, it is seen that U_1 and $U_3 = W_1 + W_2$ are independently distributed as their joint density factorizes into the product of two densities $g(U_1)$ and $g_1(U_3)$, that is, $f_2(U_1, U_3) = g(U_1)g_1(U_3)$ where

$$g(U_1) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |U_1|^{\alpha_1 - \frac{p+1}{2}} |I - U_1|^{\alpha_2 - \frac{p+1}{2}}, \quad O < U_1 < I, \quad (8.1.3)$$

for $\Re(\alpha_1) > \frac{p-1}{2}$, $\Re(\alpha_2) > \frac{p-1}{2}$, and zero elsewhere, is a real matrix-variate type-1 beta density with the parameters (α_1, α_2) , and

$$g_1(U_3) = \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |U_3|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(B U_3)}, \quad U_3 > O, \quad (8.1.4)$$

for $B > O$, $\Re(\alpha_1 + \alpha_2) > \frac{p-1}{2}$, and zero elsewhere, which is a real matrix-variate gamma density with the parameters $(\alpha_1 + \alpha_2, B)$. Thus, given two independently distributed $p \times p$ real positive definite matrices W_1 and W_2 , where $W_1 \sim \text{gamma}(\alpha_1, B)$, $B > O$, $\Re(\alpha_1) > \frac{p-1}{2}$, and $W_2 \sim \text{gamma}(\alpha_2, B)$, $B > O$, $\Re(\alpha_2) > \frac{p-1}{2}$, one has $U_3 = W_1 + W_2 \sim \text{gamma}(\alpha_1 + \alpha_2, B)$, $B > O$.

In order to determine the distribution of U_2 , we first note that the exponent in (8.1.1) is $\text{tr}(B(W_1 + W_2)) = \text{tr}[B^{\frac{1}{2}}W_1B^{\frac{1}{2}} + B^{\frac{1}{2}}W_2B^{\frac{1}{2}}]$. Letting $V_j = B^{\frac{1}{2}}W_jB^{\frac{1}{2}}$, $dV_j = |B|^{\frac{p+1}{2}}dW_j$, $j = 1, 2$, which eliminates B , the resulting joint density of V_1 and V_2 , denoted by $f_3(V_1, V_2)$, being

$$f_3(V_1, V_2) = \frac{1}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |V_1|^{\alpha_1 - \frac{p+1}{2}} |V_2|^{\alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(V_1 + V_2)}, \quad V_j > O, \quad (8.1.5)$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, and zero elsewhere. Now, noting that $\text{tr}[V_1 + V_2] = \text{tr}[V_2^{\frac{1}{2}}(I + V_2^{-\frac{1}{2}}V_1V_2^{-\frac{1}{2}})V_2^{\frac{1}{2}}]$ and letting $V = V_2^{-\frac{1}{2}}V_1V_2^{-\frac{1}{2}} = U_2$ of (8.1.2) so that $dV = |V_2|^{-\frac{p+1}{2}}dV_1$ for fixed V_2 , the joint density of V and $V_2 = V_3$, denoted by $f_4(V, V_3)$, is obtained as

$$f_4(V, V_3) = \frac{1}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |V|^{\alpha_1 - \frac{p+1}{2}} |V_3|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}((I+V)^{\frac{1}{2}}V_3(I+V)^{\frac{1}{2}})} \quad (8.1.6)$$

where $\text{tr}[V_3^{\frac{1}{2}}(I + V)V_3^{\frac{1}{2}}]$ was replaced by $\text{tr}[(I + V)^{\frac{1}{2}}V_3(I + V)^{\frac{1}{2}}]$. It then suffices to integrate out V_3 from the joint density specified in (8.1.6) by making use of a real matrix-variate gamma integral, to obtain the density of $V = U_2$ that follows:

$$g_2(V) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |V|^{\alpha_1 - \frac{p+1}{2}} |I + V|^{-(\alpha_1 + \alpha_2)}, \quad V > O, \quad (8.1.7)$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, and zero elsewhere, which is a real matrix-variate type-2 beta density whose parameters are (α_1, α_2) . This completes the proof.

8.1a. Matrix-variate Gamma and Beta Densities, Complex Case

Parallel results can be obtained in the complex domain. If \tilde{W}_1 and \tilde{W}_2 are statistically independently distributed $p \times p$ Hermitian positive definite matrices having complex matrix-variate gamma densities with the parameters (α_1, \tilde{B}) and (α_2, \tilde{B}) , $\tilde{B} = \tilde{B}^* > O$, where an asterisk designates a conjugate transpose, then their joint density, denoted by $\tilde{f}(\tilde{W}_1, \tilde{W}_2)$, is given by

$$\tilde{f}(\tilde{W}_1, \tilde{W}_2) = \frac{|\det(B)|^{\alpha_1 + \alpha_2}}{\tilde{\Gamma}_p(\alpha_1)\tilde{\Gamma}_p(\alpha_2)} |\det(\tilde{W}_1)|^{\alpha_1 - p} |\det(\tilde{W}_2)|^{\alpha_2 - p} e^{-\text{tr}(\tilde{B}(\tilde{W}_1 + \tilde{W}_2))} \quad (8.1a.1)$$

for $\tilde{B} > O$, $\tilde{W}_1 > O$, $\tilde{W}_2 > O$, $\Re(\alpha_j) > p - 1$, $j = 1, 2$, and zero elsewhere, with $|\det(\tilde{W}_j)|$ denoting the absolute value or modulus of the determinant of \tilde{W}_j . Since the derivations are similar to those provided in the previous subsection for the real case, the next results will be stated without proof. Note that, in the complex domain, the square roots involved in the transformations are Hermitian positive definite matrices.

Theorem 8.1a.1. *Let the $p \times p$ Hermitian positive definite matrices \tilde{W}_1 and \tilde{W}_2 be independently distributed as complex matrix-variate gamma variables with the parameters (α_1, \tilde{B}) and (α_2, \tilde{B}) , $\tilde{B} = \tilde{B}^* > O$, respectively. Letting $\tilde{U}_3 = \tilde{W}_1 + \tilde{W}_2$,*

$$\tilde{U}_1 = (\tilde{W}_1 + \tilde{W}_2)^{-\frac{1}{2}} \tilde{W}_1 (\tilde{W}_1 + \tilde{W}_2)^{-\frac{1}{2}} = \tilde{U}_3^{-\frac{1}{2}} \tilde{W}_1 \tilde{U}_3^{-\frac{1}{2}} \text{ and } \tilde{U}_2 = \tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}},$$

then (1): \tilde{U}_3 is distributed as a complex matrix-variate gamma with the parameters $(\alpha_1 + \alpha_2, \tilde{B})$, $\tilde{B} = \tilde{B}^* > O$, $\Re(\alpha_1 + \alpha_2) > p - 1$; (2): \tilde{U}_1 and \tilde{U}_3 are independently distributed; (3): \tilde{U}_1 is distributed as a complex matrix-variate type-1 beta random variable with the parameters (α_1, α_2) ; (4): \tilde{U}_2 is distributed as a complex matrix-variate type-2 beta random variable with the parameters (α_1, α_2) .

8.1.2. Real Wishart matrices

Since Wishart matrices are distributed as matrix-variate gamma variables whose parameters are $\alpha_j = \frac{m_j}{2}$, $m_j \geq p$, and $B = \frac{1}{2}\Sigma^{-1}$, $\Sigma > O$, we have the following corollaries in the real and complex cases:

Corollary 8.1.1. *Let the $p \times p$ real positive definite matrices W_1 and W_2 be independently Wishart distributed, $W_j \sim W_p(m_j, \Sigma)$, with $m_j \geq p$, $j = 1, 2$, degrees of freedom, common parameter matrix $\Sigma > O$, and respective densities given by*

$$\phi_j(W_j) = \frac{1}{2^{\frac{pm_j}{2}} \Gamma_p(\frac{m_j}{2}) |\Sigma|^{\frac{m_j}{2}}} |W_j|^{\frac{m_j}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} W_j)}, \quad j = 1, 2, \quad (8.1.8)$$

and zero elsewhere. Then (1): $U_3 = W_1 + W_2$ is Wishart distributed with $m_1 + m_2$ degrees of freedom and parameter matrix $\Sigma > O$, that is, $U_3 \sim W_p(m_1 + m_2, \Sigma)$, $\Sigma > O$; (2): $U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}} = U_3^{-\frac{1}{2}} W_1 U_3^{-\frac{1}{2}}$ is a real matrix-variate type-1 beta random variable with the parameters (α_1, α_2) , that is, $U_1 \sim \text{type-1 beta}(\frac{m_1}{2}, \frac{m_2}{2})$; (3): U_1 and U_3 are independently distributed; (4): $U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ is a real matrix-variate type-2 beta random variable with the parameters (α_1, α_2) , that is, $V \sim \text{type-2 beta}(\frac{m_1}{2}, \frac{m_2}{2})$.

The corresponding results for the complex case are parallel with identical numbers of degrees of freedom, m_1 and m_2 , and parameter matrix $\tilde{\Sigma} = \tilde{\Sigma}^* > O$ (Hermitian positive definite). Properties associated with type-1 and type-2 beta variables hold as well in the complex domain. Consider for instance the following results which are also valid in the complex case. If U is a type-1 beta variable with the parameters (α_1, α_2) , then $I - U$ is a type-1 beta variable with the parameters (α_2, α_1) and $(I - U)^{-\frac{1}{2}} U (I - U)^{-\frac{1}{2}}$ is a type-2 beta variable with the parameters (α_1, α_2) .

8.2. Some Eigenvalues and Eigenvectors, Real Case

Observe that when X , a $p \times p$ real positive definite matrix, has a real matrix-variate gamma density with the parameters (α, B) , $B > O$, $\Re(\alpha) > \frac{p-1}{2}$, then $Z = B^{\frac{1}{2}} X B^{\frac{1}{2}}$ has a real matrix-variate gamma density with the parameters (α, I) where I is the identity matrix. The corresponding result for a Wishart matrix is the following: Let W be a real Wishart matrix having m degrees of freedom and $\Sigma > O$ as its parameter matrix, that is, $W \sim W_p(m, \Sigma)$, $\Sigma > O$, $m \geq p$, then $Z = \Sigma^{-\frac{1}{2}} W \Sigma^{-\frac{1}{2}} \sim W_p(m, I)$, $m \geq p$, that is, Z is a Wishart matrix having m degrees of freedom and I as its parameter matrix. If we are considering the roots of the determinantal equation $|\hat{W}_1 - \lambda \hat{W}_2| = 0$ where $\hat{W}_1 \sim W_p(m_1, \Sigma)$ and $\hat{W}_2 \sim W_p(m_2, \Sigma)$, $\Sigma > O$, $m_j \geq p$, $j = 1, 2$, and if \hat{W}_1 and \hat{W}_2 are independently distributed, so will $W_1 = \Sigma^{-\frac{1}{2}} \hat{W}_1 \Sigma^{-\frac{1}{2}}$ and $W_2 = \Sigma^{-\frac{1}{2}} \hat{W}_2 \Sigma^{-\frac{1}{2}}$ be. Then

$$\begin{aligned} |\Sigma^{-\frac{1}{2}} \hat{W}_1 \Sigma^{-\frac{1}{2}} - \lambda \Sigma^{-\frac{1}{2}} \hat{W}_2 \Sigma^{-\frac{1}{2}}| = 0 &\Rightarrow |\Sigma|^{-\frac{1}{2}} |\hat{W}_1 - \lambda \hat{W}_2| |\Sigma|^{-\frac{1}{2}} = 0 \\ &\Rightarrow |\hat{W}_1 - \lambda \hat{W}_2| = 0. \end{aligned} \quad (8.2.1)$$

Thus, the roots of $|W_1 - \lambda W_2| = 0$ and $|\hat{W}_1 - \lambda \hat{W}_2| = 0$ are identical. Hence, without any loss of generality, one needs only consider the roots of W_j , $W_j \sim W_p(m_j, I)$, $m_j \geq$

p , $j = 1, 2$, when independently distributed Wishart matrices sharing a common matrix parameter are involved. Observe that

$$|W_1 - \lambda W_2| = 0 \Rightarrow |W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}} - \lambda I| = 0 \quad (8.2.2),$$

which means that λ is an eigenvalue of $W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ when $W_j \stackrel{ind}{\sim} W_p(m_j, I)$, $j = 1, 2$. If Y_j is an eigenvector corresponding to the eigenvalue λ_j , it must satisfy the equation

$$(W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}) Y_j = \lambda_j Y_j. \quad (8.2.3)$$

Let the eigenvalues λ_j 's be distinct so that $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Actually, it can be shown that $Pr\{\lambda_i = \lambda_j\} = 0$ almost surely for all $i \neq j$. When the eigenvalues of $W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ are distinct, then the eigenvectors are orthogonal since $W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ is symmetric. Thus, in this case, there exists a set of p linearly independent mutually orthogonal eigenvectors. Let Y_1, \dots, Y_p be a set of normalized mutually orthonormal eigenvectors and let $Y = (Y_1, \dots, Y_p)$ be the $p \times p$ matrix consisting of the normalized eigenvectors. Our aim is to determine the joint density of Y and $\lambda_1, \dots, \lambda_p$, and thereby the marginal densities of Y and $\lambda_1, \dots, \lambda_p$. To this end, we will need the Jacobian provided in the next theorem. For its derivation and connection to other Jacobians, the reader is referred to Mathai (1997).

Theorem 8.2.1. *Let Z be a $p \times p$ real symmetric matrix comprised of distinct real scalar variables as its elements, except for symmetry, and let its distinct nonzero eigenvalues be $\lambda_1 > \lambda_2 > \dots > \lambda_p$, which are real owing to its symmetry. Let $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, $dD = d\lambda_1 \wedge \dots \wedge d\lambda_p$, and P be a unique orthonormal matrix such that $PP' = I$, $P'P = I$, and $Z = PDP'$. Then, after integrating out the differential element of P over the full orthogonal group O_p , we have*

$$dZ = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i=1}^{p-1} \prod_{j=i+1}^p (\lambda_i - \lambda_j) \right\} dD = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i<j} (\lambda_i - \lambda_j) \right\} dD. \quad (8.2.4)$$

Corollary 8.2.1. *Let $g(Z)$ be a symmetric function of the $p \times p$ real symmetric matrix Z —symmetric function in the sense that $g(AB) = g(BA)$ whenever AB and BA are defined, even if $AB \neq BA$. Let the eigenvalues of Z be distinct such that $\lambda_1 > \lambda_2 > \dots > \lambda_p$, $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ and $dD = d\lambda_1 \wedge \dots \wedge d\lambda_p$. Then,*

$$\int_Z g(Z) dZ = \int_D g(D) \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i<j} (\lambda_i - \lambda_j) \right\} dD.$$

Example 8.2.1. Consider a $p \times p$ real matrix X having a matrix-variate gamma density with shape parameter $\alpha = \frac{p+1}{2}$ and scale parameter matrix I , whose density is

$$f(X) = \frac{1}{\Gamma_p(\frac{p+1}{2})} e^{-\text{tr}(X)}, \quad X > O.$$

A variable having this density is also said to follow *the real $p \times p$ matrix-variate exponential distribution*. Let $p = 2$ and denote the eigenvalues of X by $\infty > \lambda_1 > \lambda_2 > 0$. It follows from Corollary 8.2.1 that the joint density of λ_1 and λ_2 , denoted by $f_1(D)$, with $D = \text{diag}(\lambda_1, \lambda_2)$, is given by

$$f_1(D) \, dD = \frac{1}{\Gamma_2(\frac{2+1}{2})} \frac{\pi^{\frac{2}{2}}}{\Gamma_2(\frac{2}{2})} (\lambda_1 - \lambda_2) e^{-\text{tr}(D)} \, dD.$$

Verify that $f_1(D)$ is a density.

Solution 8.2.1. Since $f_1(D)$ is nonnegative, it suffices to show that the total integral is equal to 1. Excluding the constant part, the integral to be evaluated is the following:

$$\begin{aligned} & \int_{\lambda_1=0}^{\infty} \int_{\lambda_2=0}^{\lambda_1} (\lambda_1 - \lambda_2) e^{-(\lambda_1+\lambda_2)} \, d\lambda_1 \wedge \, d\lambda_2 \\ &= \int_{\lambda_1=0}^{\infty} \lambda_1 e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} e^{-\lambda_2} \, d\lambda_2 \right] \, d\lambda_1 - \int_{\lambda_1=0}^{\infty} e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} \lambda_2 e^{-\lambda_2} \, d\lambda_2 \right] \, d\lambda_1 \\ &= \int_{\lambda_1=0}^{\infty} \lambda_1 e^{-\lambda_1} [1 - e^{-\lambda_1}] \, d\lambda_1 \\ &\quad + \int_0^{\infty} \lambda_1 e^{-2\lambda_1} \, d\lambda_1 - \int_0^{\infty} e^{-\lambda_1} \, d\lambda_1 + \int_0^{\infty} e^{-2\lambda_1} \, d\lambda_1 \\ &= \int_0^{\infty} \lambda_1 e^{-\lambda_1} \, d\lambda_1 - \int_0^{\infty} e^{-\lambda_1} \, d\lambda_1 + \int_0^{\infty} e^{-2\lambda_1} \, d\lambda_1 \\ &= 1 - 1 + \frac{1}{2} = \frac{1}{2}. \end{aligned} \tag{i}$$

Let us now compute the constant part:

$$\begin{aligned} \frac{1}{\Gamma_2(\frac{2+1}{2})} \frac{\pi^{\frac{2}{2}}}{\Gamma_2(\frac{2}{2})} &= \frac{1}{\Gamma_2(\frac{3}{2})} \frac{\pi^2}{\Gamma_2(1)} \\ &= \frac{1}{\sqrt{\pi}(\frac{1}{2}\sqrt{\pi})} \frac{\pi^2}{\sqrt{\pi}\sqrt{\pi}} = 2. \end{aligned} \tag{ii}$$

Since the product of (i) and (ii) gives 1, $f_1(D)$ is indeed a density.

Example 8.2.2. Consider a $p \times p$ matrix having a real matrix-variate type-1 beta density with the parameters $\alpha = \frac{p+1}{2}$, $\beta = \frac{p+1}{2}$, whose density, denoted by $f(X)$, is

$$f(X) = \frac{\Gamma_p(p+1)}{\Gamma_p(\frac{p+1}{2})\Gamma_p(\frac{p+1}{2})}, \quad O < X < I,$$

and zero elsewhere. This density $f(X)$ is also referred to as a *real $p \times p$ matrix-variate uniform density*. Let $p = 2$ and the eigenvalues of X be $1 > \lambda_1 > \lambda_2 > 0$. Then, the density of $D = \text{diag}(\lambda_1, \lambda_2)$, denoted by $f_1(D)$, is

$$f_1(D)dD = \frac{\Gamma_2(2+1)}{[\Gamma_2(\frac{2+1}{2})]^2} \frac{\pi^{\frac{2^2}{2}}}{\Gamma_2(\frac{2}{2})} (\lambda_1 - \lambda_2) dD, \quad 1 > \lambda_1 > \lambda_2 > 0,$$

and zero elsewhere. Verify that $f_1(D)$ is a density.

Solution 8.2.2. The constant part simplifies to the one:

$$\begin{aligned} \frac{\Gamma_2(3)}{[\Gamma_2(\frac{3}{2})]^2} \frac{\pi^2}{\Gamma_2(1)} &= \frac{\sqrt{\pi} \Gamma(3) \Gamma(\frac{5}{2})}{[\sqrt{\pi} \Gamma(\frac{3}{2}) \Gamma(\frac{2}{2})]^2} \frac{\pi^2}{\sqrt{\pi} \Gamma(1) \Gamma(\frac{1}{2})} \\ &= \frac{\sqrt{\pi} (2)(\frac{3}{2})(\frac{1}{2}) \sqrt{\pi}}{[\sqrt{\pi}(\frac{1}{2}) \sqrt{\pi}]^2} \frac{\pi^2}{\sqrt{\pi} \sqrt{\pi}} = 6. \end{aligned} \quad (i)$$

Let us now consider the functional part of the integrand:

$$\begin{aligned} \int_{\lambda_1=0}^1 \int_{\lambda_2=0}^{\lambda_1} (\lambda_1 - \lambda_2) d\lambda_1 \wedge d\lambda_2 &= \int_0^1 \lambda_1^2 d\lambda_1 - \int_0^1 \frac{\lambda_1^2}{2} d\lambda_1 \\ &= \frac{1}{2} \int_0^1 \lambda_1^2 d\lambda_1 = \frac{1}{6}. \end{aligned} \quad (ii)$$

As the product of (i) and (ii) equals 1, it is verified that $f_1(D)$ is a density.

Example 8.2.3. Consider a $p \times p$ matrix having a matrix-variate gamma density with shape parameter α and scale parameter matrix I . Let $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $\infty > \lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ are the eigenvalues of that matrix. The joint density of the λ_j 's or the density of D is then available as

$$f_1(D) dD = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2}) \Gamma_p(\alpha)} [\lambda_1 \dots \lambda_p]^{\alpha - \frac{p+1}{2}} e^{-(\lambda_1 + \dots + \lambda_p)} \left[\prod_{i < j} (\lambda_i - \lambda_j) \right] dD.$$

Even when α is specified, the integral representation of the density $f_1(D)$ will generally only be expressible in terms of incomplete gamma functions or confluent hypergeometric series, with simple functional forms being obtainable only for certain values of α and p . Verify that $f_1(D)$ is a density for $\alpha = \frac{7}{2}$ and $p = 2$.

Solution 8.2.3. For those values of p and α , we have

$$[\lambda_1 \lambda_2]^{\frac{7}{2} - \frac{3}{2}} (\lambda_1 - \lambda_2) e^{-(\lambda_1 + \lambda_2)} = [\lambda_1^3 \lambda_2^2 - \lambda_1^2 \lambda_2^3] e^{-(\lambda_1 + \lambda_2)}$$

whose integral over $\infty > \lambda_1 > \lambda_2 > 0$ is the sum of (i) and (ii):

$$\begin{aligned} \int_{\lambda_1=0}^{\infty} \int_{\lambda_2=0}^{\lambda_1} \lambda_1^3 \lambda_2^2 e^{-(\lambda_1 + \lambda_2)} d\lambda_2 \wedge d\lambda_1 &= \int_{\lambda_1=0}^{\infty} \lambda_1^3 e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} \lambda_2^2 e^{-\lambda_2} d\lambda_2 \right] d\lambda_1 \\ &= \int_{\lambda_1=0}^{\infty} [(-\lambda_1^5 - 2\lambda_1^4 - 2\lambda_1^3) e^{-2\lambda_1} + 2\lambda_1^3 e^{-\lambda_1}] d\lambda_1, \end{aligned} \tag{i}$$

$$\begin{aligned} - \int_{\lambda_1=0}^{\infty} \lambda_1^2 e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} \lambda_2^3 e^{-\lambda_2} d\lambda_2 \right] d\lambda_1 \\ = \int_0^{\infty} [(\lambda_1^5 + 3\lambda_1^4 + 6\lambda_1^3 + 6\lambda_1^2) e^{-2\lambda_1} - 6\lambda_1^2 e^{-\lambda_1}] d\lambda_1, \end{aligned} \tag{ii}$$

that is,

$$\begin{aligned} \int_0^{\infty} [(\lambda_1^4 + 4\lambda_1^3 + 6\lambda_1^2) e^{-2\lambda_1} + (2\lambda_1^3 - 6\lambda_1^2) e^{-\lambda_1}] d\lambda_1 \\ = 2^{-5} \Gamma(5) + 4(2^{-4}) \Gamma(4) + 6(2^{-3}) \Gamma(3) + 2\Gamma(4) - 6\Gamma(3) \\ = \frac{4!}{2^5} + \frac{4(3!)}{2^4} + \frac{6(2!)}{2^3} + 2(3!) - 6(2!) = \frac{15}{4}. \end{aligned} \tag{iii}$$

Now, consider the constant part:

$$\begin{aligned} \frac{1}{\Gamma_p(\alpha)} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} &= \frac{1}{\Gamma_2(\frac{7}{2})} \frac{\pi^2}{\Gamma_2(\frac{2}{2})} \\ &= \frac{1}{\sqrt{\pi}(\frac{5}{2})(\frac{3}{2})\frac{1}{2}\sqrt{\pi}(2!)} \frac{\pi^2}{\sqrt{\pi}\sqrt{\pi}} = \frac{4}{15}. \end{aligned} \tag{iv}$$

The product of (iii) and (iv) giving 1, this verifies that $f_1(D)$ is a density when $p = 2$ and $\alpha = \frac{7}{2}$.

8.2a. The Distributions of Eigenvalues in the Complex Case

The complex counterpart of Theorem 8.2.1 is stated next.

Theorem 8.2a.1. *Let \tilde{Z} be a $p \times p$ Hermitian matrix with distinct real nonzero eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. Let \tilde{Q} be a $p \times p$ unique unitary matrix, $\tilde{Q}\tilde{Q}^* = I$, $\tilde{Q}^*\tilde{Q} = I$ such that $\tilde{Z} = \tilde{Q}D\tilde{Q}^*$ where an asterisk designates the conjugate transpose. Then, after integrating out the differential element of \tilde{Q} over the full orthogonal group \tilde{O}_p , we have*

$$d\tilde{Z} = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \left\{ \prod_{i < j} |\lambda_i - \lambda_j|^2 \right\} dD. \quad (8.2a.1)$$

Note 8.2a.1. When the unitary matrix \tilde{Q} has diagonal elements that are real, then the integral of the differential element over the full orthogonal group \tilde{O}_p will be the following:

$$\int_{\tilde{O}_p} \tilde{h}(\tilde{Q}) = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \quad (8.2a.2)$$

where $\tilde{h}(\tilde{Q}) = \wedge[(d\tilde{Q})\tilde{Q}^*]$; the reader may refer to Theorem 4.4 and Corollary 4.3.1 of Mathai (1997) for details. If all the elements comprising \tilde{Q} are complex, then the numerator in (8.2a.2) will be π^{p^2} instead of $\pi^{p(p-1)}$. When unitary transformations are made on Hermitian matrices such as \tilde{Z} in Theorem 8.2a.1, the diagonal elements in the unitary matrix \tilde{Q} are real and hence the numerator in (8.2a.2) remains $\pi^{p(p-1)}$ in this case.

Note 8.2a.2. A corollary parallel to Corollary 8.2.1 also holds in the complex domain.

Example 8.2a.1. Consider a complex $p \times p$ matrix \tilde{X} having a matrix-variate type-1 beta density with the parameters $\alpha = p$ and $\beta = p$, so that its density, denoted by $\tilde{f}(\tilde{X})$, is the following:

$$\tilde{f}(\tilde{X}) = \frac{\tilde{\Gamma}_p(2p)}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(p)}, \quad O < \tilde{X} < I,$$

which is also referred to as *the $p \times p$ complex matrix-variate uniform density*. Let $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $1 > \lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ are the eigenvalues of \tilde{X} . Then, the density of D , denoted by $f_1(D)$, is given by

$$f_1(D) dD = \frac{\tilde{\Gamma}_p(2p)}{[\tilde{\Gamma}_p(p)]^2} \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \left[\prod_{i < j} (\lambda_i - \lambda_j)^2 \right] dD. \quad (i)$$

Verify that (i) is a density for $p = 2$.

Solution 8.2a.1. For Hermitian matrices the eigenvalues are real. Consider the integral over $(\lambda_1 - \lambda_2)^2$:

$$\begin{aligned} & \int_{\lambda_1=0}^1 \left[\int_{\lambda_2=0}^{\lambda_1} (\lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2) d\lambda_2 \right] d\lambda_1 \\ &= \int_{\lambda_1=0}^1 \left[\lambda_1^3 + \frac{\lambda_1^3}{3} - 2\frac{\lambda_1^3}{2} \right] d\lambda_1 \\ &= \int_0^1 \frac{\lambda_1^3}{3} d\lambda_1 = \frac{1}{12}. \end{aligned} \tag{ii}$$

Let us now evaluate the constant part:

$$\begin{aligned} \frac{\tilde{\Gamma}_p(2p) \pi^{p(p-1)}}{[\tilde{\Gamma}_p(p)]^2 \tilde{\Gamma}_p(p)} &= \frac{\tilde{\Gamma}_2(4) \pi^{2(1)}}{[\tilde{\Gamma}_2(2)]^2 \tilde{\Gamma}_2(2)} \\ &= \frac{\pi \Gamma(4)\Gamma(3)}{\pi^2[\Gamma(2)\Gamma(1)]^2} \frac{\pi^2}{\pi \Gamma(2)\Gamma(1)} = 12. \end{aligned} \tag{iii}$$

The product of (ii) and (iii) equalling 1, the solution is complete.

Example 8.2a.2. Consider a $p \times p$ complex matrix \tilde{X} having a matrix-variate gamma density with the parameters $(\alpha = p, \beta = I)$. Let $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, where $\infty > \lambda_1 > \dots > \lambda_p > 0$ are the eigenvalues of \tilde{X} . Denoting the density of D by $f_1(D)$, we have

$$f_1(D) = \frac{1}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(p)} \pi^{p(p-1)} e^{-(\lambda_1 + \dots + \lambda_p)} \left[\prod_{i < j} (\lambda_i - \lambda_j)^2 \right].$$

When $\alpha = p$, this density is the $p \times p$ complex matrix-variate exponential density. Verify that $f_1(D)$ is a density for $p = 2$.

Solution 8.2a.2. The constant part simplifies to the following:

$$\begin{aligned} \frac{1}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(p)} \pi^{p(p-1)} &= \frac{1}{\tilde{\Gamma}_2(2) \tilde{\Gamma}_2(2)} \pi^{2(1)} \\ &= \frac{1}{\pi \Gamma(2)\Gamma(1) \pi \Gamma(2)\Gamma(1)} \pi^2 = 1, \end{aligned} \tag{i}$$

and the integrals over the λ_j 's are evaluated as follows:

$$\begin{aligned}
& \int_{\lambda_1=0}^{\infty} \int_{\lambda_2=0}^{\lambda_1} (\lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2) e^{-(\lambda_1+\lambda_2)} d\lambda_1 \wedge d\lambda_2 \\
&= \int_{\lambda_1=0}^{\infty} \lambda_1^2 e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} e^{-\lambda_2} d\lambda_2 \right] d\lambda_1 + \int_{\lambda_1=0}^{\infty} e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} \lambda_2^2 e^{-\lambda_2} d\lambda_2 \right] d\lambda_1 \\
&\quad - 2 \int_{\lambda_1=0}^{\infty} \lambda_1 e^{-\lambda_1} \left[\int_{\lambda_2=0}^{\lambda_1} \lambda_2 e^{-\lambda_2} d\lambda_2 \right] d\lambda_1 \\
&= \int_0^{\infty} (-\lambda_1^2 - \lambda_1^2 - 2\lambda_1 - 2 + 2\lambda_1^2 + 2\lambda_1) e^{-2\lambda_1} d\lambda_1 \\
&\quad + \int_0^{\infty} (\lambda_1^2 + 2 - 2\lambda_1) e^{-\lambda_1} d\lambda_1 = \int_0^{\infty} (\lambda_1^2 - 2\lambda_1 + 2) e^{-\lambda_1} d\lambda_1 - 2 \int_0^{\infty} e^{-2\lambda_1} d\lambda_1 \\
&= \Gamma(3) - 2\Gamma(2) + 2 - \frac{2}{2} = 1. \tag{ii}
\end{aligned}$$

Now, taking the product of (i) and (ii), we obtain 1, and the result is verified.

8.2.1. Eigenvalues of matrix-variate gamma and Wishart matrices, real case

Let W_1 and W_2 be two $p \times p$ real positive definite matrix-variate random variables that are independently distributed as matrix-variate gamma random variables with the parameters (α_1, B) and (α_2, B) , respectively. When $\alpha_j = \frac{m_j}{2}$, $m_j \geq p$, $j = 1, 2$, with $m_1, m_2 = p, p + 1, \dots$, and $B = \frac{1}{2}I$, W_1 and W_2 are independently Wishart distributed with m_1 and m_2 degrees of freedom, respectively; refer to the earlier discussion about the elimination of the scale parameter matrix $\Sigma > O$ in a matrix-variate Wishart distribution. Consider the determinantal equation

$$|W_1 - \lambda W_2| = 0 \Rightarrow |W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}} - \lambda I| = 0. \tag{8.2.5}$$

Thus, λ is an eigenvalue of $U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$. It has already been established in Theorem 8.1.1 that $U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ is distributed as a real matrix-variate type-2 beta random variable with the parameters α_1 and α_2 whose density is

$$f_u(U_2) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |U_2|^{\alpha_1 - \frac{p+1}{2}} |I + U_2|^{-(\alpha_1 + \alpha_2)}, \tag{8.2.6}$$

for $U_2 > O$, $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, and zero elsewhere. Note that this distribution is free of the scale parameter matrix B . Writing U_2 in terms of its eigenvalues and making use of (8.2.4), we have the following result:

Theorem 8.2.2. Let $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ be the distinct roots of the determinantal equation (8.2.5) or, equivalently, let the λ_j 's be the eigenvalues of $U_2 = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$, as defined in (8.2.5). Then, after integrating out over the full orthogonal group O_p , the joint density of $\lambda_1, \dots, \lambda_p$, denoted by $g_1(D)$ with $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, is obtained as

$$g_1(D) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} \left\{ \prod_{j=1}^p \lambda_j^{\alpha_1 - \frac{p+1}{2}} \right\} \left\{ \prod_{j=1}^p (1 + \lambda_j)^{-(\alpha_1 + \alpha_2)} \right\} \quad (8.2.7)$$

$$\times \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i < j} (\lambda_i - \lambda_j) \right\} dD, \quad dD = d\lambda_1 \wedge \dots \wedge d\lambda_p.$$

Proof: Applying the transformation $U_2 = PDP'$, $PP' = I$, $P'P = I$, where P is a unique orthonormal matrix, to the density of U_2 given in (8.2.6), it follows from Theorem 8.2.1 that

$$dU_2 = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i < j} (\lambda_i - \lambda_j) \right\} dD$$

after integrating out the differential element corresponding to the orthonormal matrix P . On substituting $|U_2| = \lambda_1 \dots \lambda_p$ and $|I + U_2| = (1 + \lambda_1) \dots (1 + \lambda_p)$ in (8.2.6), the result is established.

Note 8.2.1. When $\alpha_1 = \frac{m_1}{2}$, $\alpha_2 = \frac{m_2}{2}$, $m_j \geq p$, with $m_1, m_2 = p, p + 1, \dots$, in Theorem 8.2.2, we have the corresponding result for real Wishart matrices having m_1 and m_2 degrees of freedom and parameter matrix $\frac{1}{2}I_p$.

Example 8.2.4. Let the $p \times p$ real matrix X have a real matrix-variate type-2 beta density with the parameters $\alpha = \frac{p+1}{2}$ and $\beta = \frac{p+1}{2}$. Then, the joint density of its eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$, or that of $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, denoted by $g_1(D)$, is

$$g_1(D) = \frac{\Gamma_p(p + 1)}{[\Gamma_p(\frac{p+1}{2})]^2} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left[\prod_{j=1}^p (1 + \lambda_j)^{-(p+1)} \right] \prod_{i < j} (\lambda_i - \lambda_j).$$

Verify that $g_1(D)$ is a density for $p = 2$.

Solution 8.2.4. Consider the total integral for $p = 2$. The constant part is

$$\begin{aligned} \frac{\Gamma_p(p + 1)}{[\Gamma_p(\frac{p+1}{2})]^2} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} &= \frac{\Gamma_2(3)}{[\Gamma_2(\frac{3}{2})]^2} \frac{\pi^2}{\Gamma_2(\frac{2}{2})} = \frac{\sqrt{\pi} \Gamma(3) \Gamma(\frac{5}{2})}{\pi [\Gamma(\frac{3}{2}) \Gamma(1)]^2} \frac{\pi^2}{\sqrt{\pi} \Gamma(1) \Gamma(\frac{1}{2})} \\ &= \frac{\sqrt{\pi} (2)(\frac{3}{2})(\frac{1}{2}) \sqrt{\pi}}{\pi (\frac{1}{4}) \pi} \frac{\pi^2}{\sqrt{\pi} \sqrt{\pi}} = 6, \end{aligned} \quad (i)$$

and the integral part is obtained as follows:

$$\begin{aligned} & \int_{\lambda_1=0}^{\infty} \int_{\lambda_2=0}^{\lambda_1} \frac{(\lambda_1 - \lambda_2)}{(1 + \lambda_1)^3(1 + \lambda_2)^3} d\lambda_1 \wedge d\lambda_2 \\ &= \int_{\lambda_1=0}^{\infty} \frac{\lambda_1}{(1 + \lambda_1)^3} \left[\int_{\lambda_2=0}^{\lambda_1} \frac{1}{(1 + \lambda_2)^3} d\lambda_2 \right] d\lambda_1 \\ & \quad - \int_{\lambda_1=0}^{\infty} \frac{1}{(1 + \lambda_1)^3} \left[\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2}{(1 + \lambda_2)^3} d\lambda_2 \right] d\lambda_1. \end{aligned} \quad (ii)$$

The first integral over λ_2 in (ii) is

$$\int_{\lambda_2=0}^{\lambda_1} \frac{1}{(1 + \lambda_2)^3} d\lambda_2 = \frac{1}{2} \left[1 - \frac{1}{(1 + \lambda_1)^2} \right];$$

then, integrating with respect to λ_1 yields

$$\begin{aligned} \frac{1}{2} \int_{\lambda_1=0}^{\infty} \frac{\lambda_1}{(1 + \lambda_1)^3} \left[1 - \frac{1}{(1 + \lambda_1)^2} \right] d\lambda_1 &= \frac{1}{2} \left\{ \frac{\Gamma(2)\Gamma(1)}{\Gamma(3)} - \frac{\Gamma(2)\Gamma(3)}{\Gamma(5)} \right\} \\ &= \frac{1}{2} \left\{ \frac{1}{2} - \frac{1}{12} \right\} = \frac{5}{24}. \end{aligned} \quad (iii)$$

Now, after integrating by parts, the second integral over λ_2 in (ii) is the following:

$$\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2}{(1 + \lambda_2)^3} d\lambda_2 = \frac{1}{2} \left[-\frac{\lambda_1}{(1 + \lambda_1)^2} + 1 - \frac{1}{1 + \lambda_1} \right];$$

then, integrating with respect to λ_1 gives

$$\begin{aligned} & -\frac{1}{2} \int_{\lambda_1=0}^{\infty} \frac{1}{(1 + \lambda_1)^3} \left[-\frac{\lambda_1}{(1 + \lambda_1)^2} + 1 - \frac{1}{1 + \lambda_1} \right] d\lambda_1 \\ &= \frac{1}{2} \left[\frac{\Gamma(2)\Gamma(3)}{\Gamma(5)} - \frac{1}{2} + \frac{1}{3} \right] = -\frac{1}{24}. \end{aligned} \quad (iv)$$

Combining (iii) and (iv), the sum is

$$\frac{5}{24} - \frac{1}{24} = \frac{1}{6}. \quad (v)$$

Finally, the product of (i) and (v) is 1, which verifies that $f(D)$ is indeed a density when $p = 2$.

8.2a.1. Eigenvalues of complex matrix-variate gamma and Wishart matrices

A parallel result can be obtained in the complex domain. Let \tilde{W}_1 and \tilde{W}_2 be independently distributed $p \times p$ complex matrix-variate gamma random variables with parameters (α_1, \tilde{B}) and (α_2, \tilde{B}) , $\tilde{B} = \tilde{B}^* > O$, $\Re(\alpha_j) > p - 1$, $j = 1, 2$. Consider the determinantal equation

$$[\det(\tilde{W}_1 - \lambda \tilde{W}_2)] = 0 \Rightarrow [\det(\tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}} - \lambda I)] = 0. \tag{8.2a.3}$$

It follows from Theorem 8.1a.1 that $\tilde{U}_2 = \tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}}$ has a complex matrix-variate type-2 beta distribution with the parameters (α_1, α_2) , whose associated density is

$$\tilde{f}_u(\tilde{U}_2) = \frac{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)}{\tilde{\Gamma}_p(\alpha_1) \tilde{\Gamma}_p(\alpha_2)} |\det(\tilde{U}_2)|^{\alpha_1 - p} |\det(I + \tilde{U}_2)|^{-(\alpha_1 + \alpha_2)} \tag{8.2a.4}$$

for $\tilde{U}_2 = \tilde{U}_2^* > O$, $\Re(\alpha_j) > p - 1$, $j = 1, 2$, and zero elsewhere. Observe that the distribution of \tilde{U}_2 is free of the scale parameter matrix $\tilde{B} > O$ and that \tilde{W}_1 and \tilde{W}_2 are Hermitian positive definite so that their eigenvalues $\lambda_1 > \dots > \lambda_p > 0$, assumed to be distinct, are real and positive. Writing \tilde{U}_2 in terms of its eigenvalues and making use of (8.2a.1), we have the following result:

Theorem 8.2a.2. *Let $\tilde{U}_2 = \tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}}$ and its distinct eigenvalues $\lambda_1 > \dots > \lambda_p > 0$ be as defined in the determinantal equation (8.2a.3). Then, after integrating out the differential element corresponding to the unique unitary matrix \tilde{Q} , $\tilde{Q} \tilde{Q}^* = I$, $\tilde{Q}^* \tilde{Q} = I$, such that $\tilde{U}_2 = \tilde{Q} D \tilde{Q}^*$, with $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, the joint density of $\lambda_1, \dots, \lambda_p$, denoted by $\tilde{g}_1(\tilde{D})$, is obtained as*

$$\begin{aligned} \tilde{g}_1(D) dD &= \frac{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)}{\tilde{\Gamma}_p(\alpha_1) \tilde{\Gamma}_p(\alpha_2)} \left[\prod_{j=1}^p \lambda_j^{\alpha_1 - p} \right] \left[\prod_{j=1}^p (1 + \lambda_j) \right]^{-(\alpha_1 + \alpha_2)} \\ &\times \left[\prod_{i < j} |\lambda_i - \lambda_j|^2 \right] \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} dD, \quad dD = d\lambda_1 \wedge \dots \wedge d\lambda_p, \end{aligned} \tag{8.2a.5}$$

where

$$\tilde{\Gamma}_p(\alpha) = \pi^{\frac{p(p-1)}{2}} \Gamma(\alpha) \Gamma(\alpha - 1) \dots \Gamma(\alpha - p + 1), \quad \Re(\alpha) > p - 1. \tag{8.2a.6}$$

Example 8.2a.3. Let the $p \times p$ matrix \tilde{X} have a complex matrix-variate type-2 beta density with the parameters $(\alpha = p, \beta = p)$. Let its eigenvalues be $\lambda_1 > \dots > \lambda_p > 0$ and their joint density be denoted by $\tilde{g}_1(D)$, $D = \text{diag}(\lambda_1, \dots, \lambda_p)$. Then,

$$\tilde{g}_1(D) = \frac{\tilde{\Gamma}_p(2p) \pi^{p(p-1)}}{[\tilde{\Gamma}_p(p)]^2 \tilde{\Gamma}_p(p)} \left[\prod_{j=1}^p \frac{1}{(1 + \lambda_j)^{2p}} \right] \prod_{i < j} (\lambda_i - \lambda_j)^2.$$

Verify that $\tilde{g}_1(D)$ is a density for $p = 2$.

Solution 8.2a.3. Since the total integral must be unity, let us integrate out the λ_j 's. The constant part is the following:

$$\begin{aligned} \frac{\tilde{\Gamma}_p(2p) \pi^{p(p-1)}}{[\tilde{\Gamma}_p(p)]^2 \tilde{\Gamma}_p(p)} &= \frac{\tilde{\Gamma}_2(4) \pi^2}{[\tilde{\Gamma}_2(2)]^2 \tilde{\Gamma}_2(2)} \\ &= \frac{\pi \Gamma(4) \Gamma(3) \pi^2}{\pi^2 \pi} = 12. \end{aligned} \quad (i)$$

Now, consider the integrals over λ_1 and λ_2 , noting that $(\lambda_1 - \lambda_2)^2 = \lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2$:

$$\begin{aligned} &\int_{\lambda_1=0}^{\infty} \int_{\lambda_2=0}^{\lambda_1} \frac{[\lambda_1^2 + \lambda_2^2 - 2\lambda_1\lambda_2]}{(1 + \lambda_1)^4 (1 + \lambda_2)^4} d\lambda_1 \wedge d\lambda_2 \\ &= \int_{\lambda_1=0}^{\infty} \frac{\lambda_1^2}{(1 + \lambda_1)^4} \left[\int_{\lambda_2=0}^{\lambda_1} \frac{1}{(1 + \lambda_2)^4} d\lambda_2 \right] d\lambda_1 \\ &\quad + \int_{\lambda_1=0}^{\infty} \frac{1}{(1 + \lambda_1)^4} \left[\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2^2}{(1 + \lambda_2)^4} d\lambda_2 \right] d\lambda_1 \\ &\quad - 2 \int_{\lambda_1=0}^{\infty} \frac{\lambda_1}{(1 + \lambda_1)^4} \left[\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2}{(1 + \lambda_2)^4} d\lambda_2 \right] d\lambda_1. \end{aligned} \quad (ii)$$

As they appear in (ii), the integrals over λ_2 are

$$\int_{\lambda_2=0}^{\lambda_1} \frac{1}{(1 + \lambda_2)^4} d\lambda_2 = \frac{1}{3} \left[1 - \frac{1}{(1 + \lambda_1)^3} \right], \quad (iii)$$

$$\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2^2}{(1+\lambda_2)^4} d\lambda_2 = \frac{1}{3} \left[-\frac{\lambda_1^2}{(1+\lambda_1)^3} - \frac{\lambda_1}{(1+\lambda_1)^2} + 1 - \frac{1}{1+\lambda_1} \right], \quad (iv)$$

$$\int_{\lambda_2=0}^{\lambda_1} \frac{\lambda_2}{(1+\lambda_2)^4} d\lambda_2 = \frac{1}{3} \left[-\frac{\lambda_1}{(1+\lambda_1)^3} + \frac{1}{2} - \frac{1}{2(1+\lambda_1)^2} \right]; \quad (v)$$

then, integrating with respect to λ_1 yields

$$\begin{aligned} & \int_{\lambda_1=0}^{\infty} \frac{\lambda_1^2}{(1+\lambda_1)^4} \left[\frac{1}{3} \left(1 - \frac{1}{(1+\lambda_1)^3} \right) \right] d\lambda_1 \\ &= \frac{1}{3} \left[\frac{\Gamma(3)\Gamma(1)}{\Gamma(4)} - \frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} \right], \end{aligned} \quad (vi)$$

$$\begin{aligned} & \int_{\lambda_1=0}^{\infty} \frac{1}{(1+\lambda_1)^4} \left(\frac{1}{3} \right) \left[-\frac{\lambda_1^2}{(1+\lambda_1)^3} - \frac{\lambda_1}{(1+\lambda_1)^2} + 1 - \frac{1}{1+\lambda_1} \right] d\lambda_1 \\ &= -\frac{1}{3} \left[\frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} + \frac{\Gamma(2)\Gamma(4)}{\Gamma(6)} - \frac{\Gamma(1)\Gamma(3)}{\Gamma(4)} + \frac{\Gamma(1)\Gamma(4)}{\Gamma(5)} \right], \end{aligned} \quad (vii)$$

$$\begin{aligned} & -2 \int_{\lambda_1=0}^{\infty} \frac{\lambda_1}{(1+\lambda_1)^4} \left(\frac{1}{3} \right) \left[-\frac{\lambda_1}{(1+\lambda_1)^3} + \frac{1}{2} - \frac{1}{2(1+\lambda_1)^2} \right] d\lambda_1 \\ &= \frac{1}{3} \left[2 \frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} - \frac{\Gamma(2)\Gamma(2)}{\Gamma(4)} + \frac{\Gamma(2)\Gamma(4)}{\Gamma(6)} \right]. \end{aligned} \quad (viii)$$

Summing (vi),(vii) and (viii), we have

$$\frac{1}{3} \left[2 \frac{\Gamma(3)\Gamma(1)}{\Gamma(4)} - \frac{\Gamma(1)\Gamma(4)}{\Gamma(5)} - \frac{\Gamma(2)\Gamma(2)}{\Gamma(4)} \right] = \frac{1}{3} \left[\frac{4}{3!} - \frac{1}{4} - \frac{1}{3!} \right] = \frac{1}{12}. \quad (ix)$$

As the product of (ix) and (i) is 1, the result is established. Note that since $2p$ is a positive integer, the method of integration by parts works for a general p when the first parameter in the type-2 beta density α is equal to p . However, $\prod_{i < j} (\lambda_i - \lambda_j)$ will be difficult to handle for a general p .

Example 8.2a.4. Give an explicit representation of (8.2a.5) for $p = 3$, $\alpha_1 = 4$ and $\alpha_2 = 3$.

Solution 8.2a.4. For $p = 3$, $\alpha_1 - p = 4 - 3 = 1$, $\alpha_1 + \alpha_2 = 4 + 3 = 7$, $p(p - 1) = 6$.

The constant part is the following:

$$\begin{aligned} \frac{\tilde{\Gamma}_p(\alpha_1 + \alpha_2) \pi^{p(p-1)}}{\tilde{\Gamma}_p(\alpha_1)\tilde{\Gamma}_p(\alpha_2) \tilde{\Gamma}_p(p)} &= \frac{\tilde{\Gamma}_3(7) \pi^{3(2)}}{\tilde{\Gamma}_3(4)\tilde{\Gamma}_3(3) \tilde{\Gamma}_3(3)} \\ &= \frac{\Gamma(7)\Gamma(6)\Gamma(5)}{\pi^3[\Gamma(4)\Gamma(3)\Gamma(2)][\Gamma(3)\Gamma(2)\Gamma(1)]} \\ &\quad \times \frac{\pi^6}{\pi^3\Gamma(3)\Gamma(2)\Gamma(1)} = 43200, \end{aligned} \quad (i)$$

the functional part being the product of

$$\left(\prod_{j=1}^p \lambda_j^{\alpha_1-p}\right) \left(\prod_{j=1}^p (1 + \lambda_j)^{-(\alpha_1+\alpha_2)}\right) = (\lambda_1\lambda_2\lambda_3)[(1 + \lambda_1)(1 + \lambda_2)(1 + \lambda_3)]^{-7} \quad (ii)$$

and

$$\prod_{i<j} (\lambda_i - \lambda_j)^2 = (\lambda_1 - \lambda_2)^2(\lambda_1 - \lambda_3)^2(\lambda_2 - \lambda_3)^2. \quad (iii)$$

Multiplying (i), (ii) and (iii) yields the answer.

8.2.2. An alternative procedure in the real case

This section describes an alternative procedure that is presented in Anderson (2003). The real case will first be discussed. Let W_1 and W_2 be independently distributed real $p \times p$ matrix-variate gamma random variables with the parameters (α_1, B) , (α_2, B) , $B > O$, $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$. We are considering the determinantal equation

$$\begin{aligned} |W_1 - \lambda W_2| = 0 &\Rightarrow |W_1 - \mu(W_1 + W_2)| = 0 \\ &\Rightarrow |W_1 - \frac{\mu}{1-\mu} W_2| = 0 \\ &\Rightarrow |(W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}} - \mu I| = 0 \end{aligned} \quad (8.2.8)$$

where $\lambda = \frac{\mu}{1-\mu}$. Thus, μ is an eigenvalue of $U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}}$. It follows from Theorem 8.1.1 that the joint density of W_1 and W_2 , denoted by $f(W_1, W_2)$, can be written as

$$f(W_1, W_2) = g(U_1)g_1(U_3) \quad (8.2.9)$$

where $U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}}$ and $U_3 = W_1 + W_2$ are independently distributed. Further,

$$g(U_1) = \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)} |U_1|^{\alpha_1 - \frac{p+1}{2}} |I - U_1|^{\alpha_2 - \frac{p+1}{2}}, \quad O < U_1 < I, \quad (8.2.10)$$

for $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$, and $g(U_1) = 0$ elsewhere, is a real matrix-variate type-1 beta density, and

$$g_1(U_3) = \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |U_3|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(B U_3)} \tag{8.2.11}$$

for $B > O$, $\Re(\alpha_1 + \alpha_2) > \frac{p-1}{2}$, and $g_1(U_3) = 0$ elsewhere, is a real matrix-variate gamma density with the parameters $(\alpha_1 + \alpha_2, B)$, $B > O$. Now, consider the transformation $U_1 = P D P'$, $P P' = I$, $P' P = I$, where $D = \text{diag}(\mu_1, \dots, \mu_p)$, $\mu_1 > \dots > \mu_p > 0$ being the distinct eigenvalues of the real positive definite matrix U_1 , and the orthonormal matrix P is unique. Given the density of U_1 specified in (8.2.10), the joint density of μ_1, \dots, μ_p , denoted by $g_4(D)$, which is obtained after integrating out the differential element corresponding to P , is

$$\begin{aligned} g_4(D) dD &= \frac{\Gamma_p(\alpha_1 + \alpha_2)}{\Gamma_p(\alpha_1) \Gamma_p(\alpha_2)} \left[\prod_{j=1}^p \mu_j^{\alpha_1 - \frac{p+1}{2}} \right] \left[\prod_{j=1}^p (1 - \mu_j)^{\alpha_2 - \frac{p+1}{2}} \right] \\ &\times \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p\left(\frac{p}{2}\right)} \left[\prod_{i < j} (\mu_i - \mu_j) \right] dD. \end{aligned} \tag{8.2.12}$$

Hence, the following result:

Theorem 8.2.3. *The joint density of the eigenvalues $\mu_1 > \dots > \mu_p > 0$ of the determinantal equation in (8.2.8) is given by the expression appearing in (8.2.12), which is equal to the density specified in (8.2.7).*

Proof: It has already been established in Theorem 8.1.1 that $U_1 = (W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}}$ has the real matrix-variate type-1 beta density given in (8.2.10). Now, make the transformation $U_1 = P D P'$ where $D = \text{diag}(\mu_1, \dots, \mu_p)$ and the orthonormal matrix, P is unique. Then, the first part is established from Theorem 8.2.2. It follows from (8.2.8) that $\lambda = \frac{\mu}{1-\mu}$ or $\mu = \frac{\lambda}{1+\lambda}$ with $d\mu = \frac{1}{(1+\lambda)^2} d\lambda$ and $\mu = 1 - \frac{1}{1+\lambda}$. Observe that $\prod_{i < j} (\mu_i - \mu_j) = \prod_{i < j} \frac{(\lambda_i - \lambda_j)}{(1+\lambda_i)(1+\lambda_j)}$ and that, in this product's denominator, $1 + \lambda_i$ appears $p - 1$ times for $i = 1, \dots, p$. The exponent of $\frac{1}{1+\lambda_i}$ is $\alpha_1 - \frac{p+1}{2} + \alpha_2 - \frac{p+1}{2} + 2 + (p - 1) = \alpha_1 + \alpha_2$. On substituting these values in (8.2.12), a perfect agreement with (8.2.7) is established, which completes the proof.

Example 8.2.5. Provide an explicit representation of (8.2.12) for $p = 3$, $\alpha_1 = 4$ and $\alpha_2 = 3$.

Solution 8.2.5. Note that $\frac{p+1}{2} = \frac{3+1}{2} = 2$, $\alpha_1 - \frac{p+1}{2} = 4 - 2 = 2$ and $\alpha_2 - \frac{p+1}{2} = 3 - 2 = 1$. The constant part is

$$\begin{aligned} \frac{\Gamma_p(\alpha_1 + \alpha_2) \pi^{\frac{p^2}{2}}}{\Gamma_p(\alpha_1) \Gamma_p(\alpha_2) \Gamma_p(\frac{p}{2})} &= \frac{\Gamma_3(7) \pi^{\frac{9}{2}}}{\Gamma_3(4) \Gamma_3(3) \Gamma_3(\frac{3}{2})} \\ &= \frac{(6!)(5!)(\frac{11}{2})(\frac{9}{2})(\frac{7}{2})(\frac{5}{2})(\frac{3}{2})(\frac{1}{2})\sqrt{\pi}}{\pi^{\frac{3}{2}}(3!)(2!)(1!)(\frac{5}{2})(\frac{3}{2})(\frac{1}{2})\sqrt{\pi}(\frac{3}{2})2!(\frac{1}{2})\sqrt{\pi}1!} \\ &\quad \times \frac{\pi^{\frac{9}{2}}}{\pi^{\frac{3}{2}}(\frac{1}{2})\sqrt{\pi}\sqrt{\pi}} = 831600, \end{aligned} \quad (i)$$

the functional part being the product of

$$\left(\prod_{j=1}^p \mu_j^{\alpha_1 - \frac{p+1}{2}} \right) \left(\prod_{j=1}^p (1 - \mu_j)^{\alpha_2 - \frac{p+1}{2}} \right) = (\mu_1 \mu_2 \mu_3)^2 [(1 - \mu_1)(1 - \mu_2)(1 - \mu_3)]^1 \quad (ii)$$

and

$$\prod_{i < j} (\mu_i - \mu_j) = (\mu_1 - \mu_2)(\mu_1 - \mu_3)(\mu_2 - \mu_3). \quad (iii)$$

The product of (i), (ii) and (iii) yields the answer.

8.2.3. The joint density of the eigenvectors in the real case

In order to establish the joint density of the eigenvectors, we will proceed as follows, our starting equation being $|W_1 - \lambda W_2| = 0$. Let λ_j be a root of this equation and let Y_j be the corresponding vector. Then,

$$W_1 Y_j = \lambda_j W_2 Y_j \Rightarrow (W_1 + \lambda_j W_1) Y_j = \lambda_j (W_1 + W_2) Y_j \quad (8.2.13)$$

$$\Rightarrow W_1 Y_j = \frac{\lambda_j}{1 + \lambda_j} (W_1 + W_2) Y_j = \mu_j (W_1 + W_2) Y_j \quad (i)$$

$$\Rightarrow (W_1 + W_2)^{-1} W_1 Y_j = \mu_j Y_j. \quad (ii)$$

This shows that Y_j is the eigenvector corresponding to the eigenvalue μ_j of $(W_1 + W_2)^{-1} W_1$ or, equivalently, of $(W_1 + W_2)^{-\frac{1}{2}} W_1 (W_1 + W_2)^{-\frac{1}{2}} = U_1$ which is a real matrix-variate type-1 beta random variable. Since the μ_j 's are distinct, $\mu_1 > \dots > \mu_p > 0$ and the matrix is symmetric, the eigenvectors Y_1, \dots, Y_p are mutually orthogonal. Consider the equation

$$W_1 Y_j = \mu_j (W_1 + W_2) Y_j. \quad (iii)$$

For $i \neq j$, we also have

$$W_1 Y_i = \mu_i (W_1 + W_2) Y_i. \quad (iv)$$

Premultiplying (iii) by Y_i' and (iv) by Y_j' , and observing that $W_1' = W_1$, it follows that $(Y_i' W_1 Y_j)' = Y_j' W_1 Y_i$. Since both are real 1×1 matrices and one is the transpose of the other, they are equal. Then, on subtracting the resulting right-hand sides, we obtain

$$0 = \mu_j Y_i' (W_1 + W_2) Y_j - \mu_i Y_j' (W_1 + W_2) Y_i = (\mu_j - \mu_i) Y_i' (W_1 + W_2) Y_j.$$

Since $Y_i' (W_1 + W_2) Y_j = Y_j' (W_1 + W_2) Y_i$ by the previous argument and $\mu_i \neq \mu_j$, we must have

$$Y_i' (W_1 + W_2) Y_j = 0 \text{ for all } i \neq j. \quad (v)$$

Let us normalize Y_j as follows:

$$Y_j' (W_1 + W_2) Y_j = 1, \quad j = 1, 2, \dots, p. \quad (vi)$$

Then, combining (v) and (vi), we have

$$Y' (W_1 + W_2) Y = I, \quad Y = (Y_1, \dots, Y_p), \quad (vii)$$

which is the $p \times p$ matrix of the normalized eigenvectors. Thus,

$$\begin{aligned} W_1 + W_2 &= (Y')^{-1} Y^{-1} = Z' Z, \quad Z = Y^{-1} \\ &\Rightarrow dY = |Z|^{-2p} dZ \text{ or } dZ = |Y|^{-2p} dY, \end{aligned}$$

which follows from an application of Theorem 1.6.6. We are seeking the joint density of Y_1, \dots, Y_p or the density of Y . The density of $W_1 + W_2 = U_3$ denoted by $g_1(U_3)$, is available from (8.2.11) as

$$\begin{aligned} g_1(U_3) dU_3 &= \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |U_3|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(B U_3)} dU_3 \\ &= \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |Z' Z|^{\alpha_1 + \alpha_2 - \frac{p+1}{2}} e^{-\text{tr}(B Z Z')} d(Z' Z). \end{aligned} \quad (8.2.14)$$

Letting $U_3 = Z' Z$, ascertain the connection between the differential elements dZ and dU_3 from Theorem 4.2.3 for the case $q = p$. Then,

$$dZ = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} |U_3|^{\frac{p}{2} - \frac{p+1}{2}} dU_3 \Rightarrow dU_3 = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} |Z' Z|^{\frac{1}{2}} dZ.$$

Hence, the density of Z , denoted by $g_z(Z)$, is the following:

$$g_z(Z) dZ = \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} |Z'Z|^{\alpha_1 + \alpha_2 - \frac{p}{2}} e^{-\text{tr}(BZ'Z)} dZ,$$

so that the density of $Y = Z^{-1}$, denoted by $g_5(Y)$, is given by

$$\begin{aligned} g_5(Y) dY &= \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |Y|^{-2p} |YY'|^{-(\alpha_1 + \alpha_2) + \frac{p}{2}} e^{-\text{tr}(B(YY')^{-1})} dY \\ &= \begin{cases} \frac{|B|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |YY'|^{-(\alpha_1 + \alpha_2 + \frac{p}{2})} e^{-\text{tr}(B(YY')^{-1})} dY \\ 0, \text{ elsewhere.} \end{cases} \end{aligned} \quad (8.2.15)$$

Then, we have the following result:

Theorem 8.2.4. *Let W_1 and W_2 be independently distributed $p \times p$ real matrix-variate gamma random variables with parameters (α_1, B) , (α_2, B) , $B > O$, $\Re(\alpha_j) > \frac{p-1}{2}$, $j = 1, 2$. Consider the equation*

$$|W_1 - \lambda W_2| = 0 \Rightarrow W_1 Y_j = \lambda_j W_2 Y_j, \quad j = 1, \dots, p,$$

where Y_j is a vector corresponding to the root λ_j of the determinantal equation. Let $\lambda_1 > \dots > \lambda_p > 0$ be its distinct roots, which are also the eigenvalues of the matrix $W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$. The eigenvalues $\lambda_1, \dots, \lambda_p$ and the linearly independent orthogonal eigenvectors Y_1, \dots, Y_p are independently distributed. The joint density of the eigenvalues $\lambda_1, \dots, \lambda_p$ is available from Theorem 8.2.2 and the joint density of the eigenvectors is given in (8.2.15).

Example 8.2.6. Illustrate the steps to show that the solutions of the determinantal equation $|W_1 - \lambda W_2| = 0$ are also the eigenvalues of $W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$ for the following matrices:

$$W_1 = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Solution 8.2.6. First, let us assess whether W_1 and W_2 are positive definite matrices. Clearly, $W_1 = W_1'$ and $W_2 = W_2'$ are symmetric and their leading minors are positive:

$$|(2)| = 2 > 0, \quad \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 > 0 \Rightarrow W_2 > O; \quad |(3)| > 0, \quad \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix} = 8 > 0 \Rightarrow W_1 > O.$$

Consider the determinantal equation $|W_1 - \lambda W_2| = 0$, that is,

$$\left| \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right| = (3 - 2\lambda)^2 - (1 - \lambda)^2 = 0, \quad (i)$$

whose roots are $\lambda_1 = 2$, $\lambda_2 = \frac{4}{3}$. Let us determine $W_2^{-\frac{1}{2}}$. To this end, let us evaluate the eigenvalues and eigenvectors of W_2 . The eigenvalues of W_2 are the values of ν satisfying the equation $|W_2 - \nu I| = 0 \Rightarrow (2 - \nu)^2 - 1^2 = 0 \Rightarrow \nu_1 = 3$ and $\nu_2 = 1$ are the eigenvalues of W_2 . An eigenvector corresponding to $\nu_1 = 3$ is given by

$$\begin{bmatrix} 2 - \nu_1 & 1 \\ 1 & 2 - \nu_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_1 = 1, x_2 = 1,$$

is one solution, the normalized eigenvector corresponding to $\nu_1 = 3$ being Y_1 whose transpose is $Y_1' = \frac{1}{\sqrt{2}}[1, 1]$. Similarly, an eigenvector associated with $\nu_2 = 1$ is $x_1 = 1$, $x_2 = -1$, which once normalized becomes Y_2 whose transpose is $Y_2' = \frac{1}{\sqrt{2}}[1, -1]$. Let $\Lambda = \text{diag}(3, 1)$ be the diagonal matrix of the eigenvalues of W_2 . Then,

$$W_2[Y_1, Y_2] = [Y_1, Y_2]\Lambda \Rightarrow W_2 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Observe that W_2 , W_2^{-1} , $W_2^{\frac{1}{2}}$ and $W_2^{-\frac{1}{2}}$ share the same eigenvectors Y_1 and Y_2 . Hence,

$$\begin{aligned} W_2^{\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow \\ W_2^{-\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} T &= W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \\ &\quad \times \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \frac{1}{12} \begin{bmatrix} 20 & -4 \\ -4 & 20 \end{bmatrix}. \end{aligned}$$

The eigenvalues of T are $\frac{1}{12}$ times the solutions of $(20 - \delta)^2 - 4^2 = 0 \Rightarrow \delta_1 = 24$ and $\delta_2 = 16$. Thus, the eigenvalues of T are $\frac{24}{12} = 2$ and $\frac{16}{12} = \frac{4}{3}$, which are the solutions of the determinantal equation (i). This verifies the result.

8.2a.2. An alternative procedure in the complex case

Let \tilde{W}_1 and \tilde{W}_2 be independently distributed $p \times p$ complex matrix-variate gamma random variables with the parameters (α_1, \tilde{B}) , (α_2, \tilde{B}) , $\tilde{B} = \tilde{B}^* > O$, $\Re(\alpha_j) > p - 1$, $j = 1, 2$. We are considering the roots λ_j 's and the corresponding eigenvectors Y_j 's of the determinantal equation

$$\begin{aligned} \det(\tilde{W}_1 - \lambda \tilde{W}_2) = 0 &\Rightarrow \tilde{W}_1 \tilde{Y}_j = \lambda_j \tilde{W}_2 \tilde{Y}_j, \quad j = 1, \dots, p \\ &\Rightarrow \det(\tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}} - \lambda I) = 0. \end{aligned} \quad (8.2a.7)$$

Let the eigenvalues $\lambda_1, \dots, \lambda_p$ of $\tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}} > O$ be distinct and such that $\lambda_1 > \dots > \lambda_p > 0$, noting that for Hermitian matrices, the eigenvalues are real. We are interested in the joint distributions of the eigenvalues $\lambda_1, \dots, \lambda_p$ and the eigenvectors $\tilde{Y}_1, \dots, \tilde{Y}_p$. Alternatively, we will consider the equation

$$\det(\tilde{W}_1 - \mu(\tilde{W}_1 + \tilde{W}_2)) = 0 \Rightarrow \det\left(\tilde{W}_1 - \frac{\mu}{1 - \mu} \tilde{W}_2\right) = 0, \quad \lambda = \frac{\mu}{1 - \mu}. \quad (8.2a.8)$$

Proceeding as in the real case, one can observe that the joint density of \tilde{W}_1 and \tilde{W}_2 , denoted by $\tilde{f}(\tilde{W}_1, \tilde{W}_2)$, can be factorized into the product of the density of $\tilde{U}_1 = (\tilde{W}_1 + \tilde{W}_2)^{-\frac{1}{2}} \tilde{W}_1 (\tilde{W}_1 + \tilde{W}_2)^{-\frac{1}{2}}$, denoted by $\tilde{g}(\tilde{U}_1)$, which is a complex matrix-variate type-1 beta random variable with the parameters (α_1, α_2) , and the density of $\tilde{U}_3 = \tilde{W}_1 + \tilde{W}_2$, denoted by $\tilde{g}_1(\tilde{U}_3)$, which is a complex matrix-variate gamma density with the parameters $(\alpha_1 + \alpha_2, \tilde{B})$, $\tilde{B} > O$. That is,

$$\tilde{f}(\tilde{W}_1, \tilde{W}_2) d\tilde{W}_1 \wedge d\tilde{W}_2 = \tilde{g}(\tilde{U}_1) \tilde{g}_1(\tilde{U}_3) d\tilde{U}_1 \wedge d\tilde{U}_3 \quad (8.2a.9)$$

where

$$\tilde{g}(\tilde{U}_1) = \frac{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)}{\tilde{\Gamma}_p(\alpha_1) \tilde{\Gamma}_p(\alpha_2)} |\det(\tilde{U}_1)|^{\alpha_1 - p} |\det(I - \tilde{U}_1)|^{\alpha_2 - p} \quad (8.2a.10)$$

for $O < \tilde{U}_1 < I$, $\Re(\alpha_j) > p - 1$, $j = 1, 2$, and $\tilde{g} = 0$ elsewhere, and

$$\tilde{g}_1(\tilde{U}_3) = \frac{|\det(\tilde{B})|^{\alpha_1 + \alpha_2}}{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)} |\det(\tilde{U}_3)|^{\alpha_1 + \alpha_2 - p} e^{-\text{tr}(\tilde{B} \tilde{U}_3)} \quad (8.2a.11)$$

for $\tilde{U}_3 = \tilde{U}_3^* > O$, $\Re(\alpha_1 + \alpha_2) > p - 1$, and $\tilde{B} = \tilde{B}^* > O$, and zero elsewhere.

Note that by making the transformation $\tilde{U}_1 = QDQ^*$ with $QQ^* = I$, $Q^*Q = I$, and $D = \text{diag}(\mu_1, \dots, \mu_p)$, the joint density of μ_1, \dots, μ_p , as obtained from (8.2a.10), is given by

$$\begin{aligned} \tilde{g}_2(D) dD &= \frac{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)}{\tilde{\Gamma}_p(\alpha_1)\tilde{\Gamma}_p(\alpha_2)} \left[\prod_{j=1}^p \mu_j^{\alpha_1 - p} \right] \left[\prod_{j=1}^p (1 - \mu_j)^{\alpha_2 - p} \right] \\ &\times \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \left[\prod_{i < j} (\mu_i - \mu_j)^2 \right] dD; \end{aligned} \tag{8.2a.12}$$

also refer to Note 8.2a.1. Then, we have the following result on observing that the joint density of $\lambda_1, \dots, \lambda_p$ is available from (8.2a.12) by making the substitution $\lambda = \frac{\mu}{1-\mu}$ or $\mu = \frac{\lambda}{1+\lambda}$. Note that $d\mu_j = \frac{1}{(1+\lambda_j)^2} d\lambda_j$, $1 - \mu_j = \frac{1}{1+\lambda_j}$, and

$$\prod_{i < j} (\mu_i - \mu_j)^2 = \prod_{i < j} \frac{(\lambda_i - \lambda_j)^2}{(1 + \lambda_i)^2(1 + \lambda_j)^2}$$

whose denominator contains $p - 1$ times $(1 + \lambda_j)^2$ for each $j = 1, \dots, p$. Thus, the final exponent of $\frac{1}{1+\lambda_j}$ is $(\alpha_1 - p) + (\alpha_2 - p) + 2 + 2(p - 1) = \alpha_1 + \alpha_2$. Hence, the following result:

Theorem 8.2a.3. *In the complex case, the joint density of the eigenvalues $\mu_1 > \dots > \mu_p > 0$ of the determinantal equation in (8.2a.8) is given by the expression appearing in (8.2a.12), which is equal to the density specified in (8.2a.5).*

We now consider the joint density of the eigenvectors $\tilde{Y}_1, \dots, \tilde{Y}_p$, which will be available from (8.2a.11), thus establishing that the set of eigenvalues $\lambda_1, \dots, \lambda_p$ and the eigenvectors $\tilde{Y}_1, \dots, \tilde{Y}_p$ are independently distributed. For determining the joint density of the eigenvectors, we start with the equation

$$\tilde{W}_1 \tilde{Y}_j = \mu_j (\tilde{W}_1 + \tilde{W}_2) \tilde{Y}_j, \quad j = 1, \dots, p, \tag{i}$$

observing that λ_j and μ_j share the same eigenvector \tilde{Y}_j . That is,

$$\tilde{W}_1 \tilde{Y}_i = \mu_i (\tilde{W}_1 + \tilde{W}_2) \tilde{Y}_i, \quad i = 1, \dots, p. \tag{ii}$$

We continue as in the real case, showing that $\tilde{Y}_i^* (\tilde{W}_1 + \tilde{W}_2) \tilde{Y}_j = 0$ for all $i \neq j$. Then, we normalize \tilde{Y}_j as follows: $\tilde{Y}_j^* (\tilde{W}_1 + \tilde{W}_2) \tilde{Y}_j = 1$, $j = 1, \dots, p$. Letting $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_p)$ be the $p \times p$ matrix of the normalized eigenvectors, we have $\tilde{Y}^* (\tilde{W}_1 + \tilde{W}_2) \tilde{Y} = I \Rightarrow \tilde{U}_1 = \tilde{W}_1 + \tilde{W}_2 = (\tilde{Y}^*)^{-1} (\tilde{Y})^{-1}$. Letting $(\tilde{Y})^{-1} = \tilde{Z}$ so that $\tilde{Z}^* \tilde{Z} = (\tilde{Y} \tilde{Y}^*)^{-1}$ and applying Theorem 4.2a.3, $d(\tilde{Z}^* \tilde{Z}) = \frac{\tilde{\Gamma}_p(p)}{\pi^{p(p-1)}} d\tilde{Z}$. Hence, given the density of $\tilde{W}_1 + \tilde{W}_2$ specified in (8.2a.11), the density of \tilde{Z} , denoted by $\tilde{g}_3(\tilde{Z})$, is obtained as

$$\tilde{g}_3(\tilde{Z}) d\tilde{Z} = \frac{|\det(\tilde{B})|^{\alpha_1 + \alpha_2}}{\tilde{\Gamma}_p(\alpha_1 + \alpha_2)} \frac{\tilde{\Gamma}_p(p)}{\pi^{p(p-1)}} |\det(\tilde{Z}^* \tilde{Z})|^{\alpha_1 + \alpha_2 - p} e^{-\text{tr}(\tilde{B} \tilde{Z}^* \tilde{Z})} d\tilde{Z}. \tag{8.2a.13}$$

Now noting that $\tilde{Z} = \tilde{Y}^{-1} \Rightarrow d\tilde{Z} = |\det(\tilde{Y}^* \tilde{Y})|^{-p} d\tilde{Y}$ from an application of Theorem 1.6a.6, and substituting in (8.2a.13), we obtain the following density of \tilde{Y} , denoted by $\tilde{g}_4(\tilde{Y})$:

$$\tilde{g}_4(\tilde{Y})d\tilde{Y} = \frac{|\det(\tilde{B})|^{\alpha_1 + \alpha_2} \tilde{\Gamma}_p(p)}{\tilde{\Gamma}_p(\alpha_1 + \alpha_2) \pi^{p(p-1)}} |\det(\tilde{Y} \tilde{Y}^*)|^{-\alpha_1 - \alpha_2} e^{-\text{tr}(\tilde{B}(\tilde{Y}^*)^{-1} \tilde{Y}^{-1})} d\tilde{Y} \quad (8.2a.14)$$

for $\tilde{B} = \tilde{B}^* > O$, $\Re(\alpha_1 + \alpha_2) > p - 1$, and zero elsewhere.

Example 8.2a.5. Show that the roots of the determinantal equation $\det(\tilde{W}_1 - \lambda \tilde{W}_2) = 0$ are the same as the eigenvalues of $\tilde{W}_2^{-\frac{1}{2}} \tilde{W}_1 \tilde{W}_2^{-\frac{1}{2}}$ for the following Hermitian positive definite matrices:

$$\tilde{W}_1 = \begin{bmatrix} 3 & 1-i \\ 1+i & 3 \end{bmatrix}, \quad \tilde{W}_2 = \begin{bmatrix} 3 & \sqrt{2}(1+i) \\ \sqrt{2}(1-i) & 3 \end{bmatrix}.$$

Solution 8.2a.5. Let us evaluate the eigenvalues and eigenvectors of \tilde{W}_2 . Consider the equation $\det(\tilde{W}_2 - \mu I) = 0 \Rightarrow (3 - \mu)^2 - 2^2 = 0 \Rightarrow \mu_1 = 5$ and $\mu_2 = 1$ are the eigenvalues of \tilde{W}_2 . An eigenvector corresponding to $\mu_1 = 5$ must satisfy the equation

$$\begin{bmatrix} 3-5 & \sqrt{2}(1+i) \\ \sqrt{2}(1-i) & 3-5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} -2x_1 + \sqrt{2}(1+i)x_2 = 0 \\ \sqrt{2}(1-i)x_1 - 2x_2 = 0 \end{cases}.$$

Since it is a singular system of linear equations, we can solve any one of them for x_1 and x_2 . For $x_2 = 1$, we have $x_1 = \frac{1}{\sqrt{2}}(1+i)$. Thus, one eigenvector is

$$\tilde{X}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}}(1+i) \\ 1 \end{bmatrix} \Rightarrow \tilde{X}_1^* \tilde{X}_1 = 2 \Rightarrow \tilde{Y}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}}(1+i) \\ 1 \end{bmatrix}$$

where \tilde{Y}_1 is the normalized eigenvector obtained from \tilde{X}_1 . Similarly, corresponding to the eigenvalue $\mu_2 = 1$, we have the normalized eigenvector

$$\tilde{Y}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -\frac{1}{\sqrt{2}}(1+i) \\ 1 \end{bmatrix} \quad \text{and} \quad \tilde{W}_2[\tilde{Y}_1, \tilde{Y}_2] = [\tilde{Y}_1, \tilde{Y}_2] \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix},$$

so that

$$\tilde{W}_2 = \frac{1}{2} \begin{bmatrix} \frac{1}{\sqrt{2}}(1+i) & -\frac{1}{\sqrt{2}}(1+i) \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}(1-i) & 1 \\ -\frac{1}{\sqrt{2}}(1-i) & 1 \end{bmatrix},$$

observing that the above format is $\tilde{W}_2 = \tilde{Y}D\tilde{Y}^*$ with $\tilde{Y} = [\tilde{Y}_1, \tilde{Y}_2]$ and $D = \text{diag}(5, 1)$, the diagonal matrix of the eigenvalues of \tilde{W}_2 . Since \tilde{W}_2 , $\tilde{W}_2^{\frac{1}{2}}$, \tilde{W}_2^{-1} and $\tilde{W}_2^{-\frac{1}{2}}$ share the same eigenvectors, we have

$$\begin{aligned} \tilde{W}_2^{\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} \frac{1}{\sqrt{2}}(1+i) & -\frac{1}{\sqrt{2}}(1+i) \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}(1-i) & 1 \\ -\frac{1}{\sqrt{2}}(1-i) & 1 \end{bmatrix} \Rightarrow \\ \tilde{W}_2^{-\frac{1}{2}} &= \frac{1}{2} \begin{bmatrix} \frac{1}{\sqrt{2}}(1+i) & -\frac{1}{\sqrt{2}}(1+i) \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}(1-i) & 1 \\ -\frac{1}{\sqrt{2}}(1-i) & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \frac{1}{\sqrt{5}} + 1 & (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1+i) \\ (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1-i) & \frac{1}{\sqrt{5}} + 1 \end{bmatrix}. \end{aligned} \tag{i}$$

It is easily verified that

$$\begin{aligned} \tilde{W}_2^{-\frac{1}{2}}\tilde{W}_2^{-\frac{1}{2}} &= \frac{1}{4} \begin{bmatrix} \frac{1}{\sqrt{5}} + 1 & (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1+i) \\ (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1-i) & \frac{1}{\sqrt{5}} + 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \frac{1}{\sqrt{5}} + 1 & (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1+i) \\ (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1-i) & \frac{1}{\sqrt{5}} + 1 \end{bmatrix} \\ &= \frac{1}{5} \begin{bmatrix} 3 & -\sqrt{2}(1+i) \\ -\sqrt{2}(1-i) & 3 \end{bmatrix} = \tilde{W}_2^{-1}. \end{aligned}$$

Letting $Q = \tilde{W}_2^{-\frac{1}{2}}\tilde{W}_1\tilde{W}_2^{-\frac{1}{2}}$, we have

$$\begin{aligned} Q &= \frac{1}{4} \begin{bmatrix} \frac{1}{\sqrt{5}} + 1 & (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1+i) \\ (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1-i) & \frac{1}{\sqrt{5}} + 1 \end{bmatrix} \begin{bmatrix} 3 & 1-i \\ 1+i & 3 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \frac{1}{\sqrt{5}} + 1 & (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1+i) \\ (\frac{1}{\sqrt{5}} - 1)\frac{1}{\sqrt{2}}(1-i) & \frac{1}{\sqrt{5}} + 1 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} \frac{6^2}{5} & -\frac{12}{5}\sqrt{2}(1+i) + \frac{4}{\sqrt{5}}(1-i) \\ -\frac{12}{5}\sqrt{2}(1-i) + \frac{4}{\sqrt{5}}(1+i) & \frac{6^2}{5} \end{bmatrix}. \end{aligned} \tag{ii}$$

The eigenvalues of $4Q$ can be determined by solving the equation

$$\begin{aligned} \left(\frac{6^2}{5} - \nu\right)^2 - \left[-\frac{12}{5}\sqrt{2}(1+i) + \frac{4}{\sqrt{5}}(1-i)\right] \left[-\frac{12}{5}\sqrt{2}(1+i) + \frac{4}{\sqrt{5}}(1-i)\right] &= 0 \Rightarrow \\ \left(\frac{6^2}{5} - \nu\right)^2 - \frac{4^2}{5^2}(46) &= 0 \Rightarrow \\ \nu &= \frac{4}{5}(9 \pm \sqrt{46}). \end{aligned} \quad (iii)$$

Thus, the eigenvalues of Q , denoted by δ , are

$$\delta = \frac{1}{5}(9 \pm \sqrt{46}). \quad (iv)$$

Now, let us consider the determinantal equation $\det(\tilde{W}_1 - \lambda\tilde{W}_2) = 0$, that is,

$$\begin{vmatrix} 3(1-\lambda) & (1-i) - \lambda\sqrt{2}(1+i) \\ (1+i) - \lambda\sqrt{2}(1-i) & 3(1-\lambda) \end{vmatrix} = 0,$$

which yields

$$\begin{aligned} 3^2(1-\lambda)^2 - [(1+i) - \lambda\sqrt{2}(1-i)][(1-i) - \lambda\sqrt{2}(1+i)] &= 0 \Rightarrow \\ 3^2(1-\lambda)^2 - [(1-\sqrt{2}\lambda)^2 + (1+\sqrt{2}\lambda)^2] &= 0 \Rightarrow \\ \lambda &= \frac{1}{5}(9 \pm \sqrt{46}). \end{aligned} \quad (v)$$

The eigenvalues obtained in (iv) and (v) being identical, the result is established.

8.3. The Singular Real Case

If a $p \times p$ real matrix-variate gamma distribution with parameters (α, β) and $B > O$ is singular and positive semi-definite, its $p \times p$ -variate density does not exist. When $\alpha = \frac{m}{2}$, $m \geq p$, and $B = \frac{1}{2}\Sigma^{-1}$, $\Sigma > O$, the gamma density is called a Wishart density with m degrees of freedom and parameter matrix $\Sigma > O$. If the rank of the gamma or Wishart matrices is $r < p$, in which case they are positive semi-definite, the resulting distributions are said to be singular. It can be shown that, in this instance, we have in fact nonsingular $r \times r$ -variate gamma or Wishart distributions. In order to establish this, the matrix theory results presented next are required.

Let $A = A' \geq O$ (non-negative definite) be a $p \times p$ real matrix of rank $r < p$, and the elementary matrices E_1, E_2, \dots, E_k be such that by operating on A , one has

$$E_k \cdots E_2 E_1 A E_1' E_2' \cdots E_k' = \begin{bmatrix} I_r & O_1 \\ O_2 & O_3 \end{bmatrix}$$

where O_1, O_2 and O_3 are null matrices, with O_3 being of order $(p - r) \times (p - r)$. Then,

$$A = E_1^{-1} \cdots E_k^{-1} \begin{bmatrix} I_r & O_1 \\ O_2 & O_3 \end{bmatrix} E_k'^{-1} \cdots E_1'^{-1} = Q \begin{bmatrix} I_r & O_1 \\ O_2 & O_3 \end{bmatrix} Q'$$

where Q is a product of inverses of elementary matrices and hence, nonsingular. Letting

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}, \quad Q_{11} \text{ being } r \times r \text{ and } Q_{22}, (p - r) \times (p - r),$$

$$Q \begin{bmatrix} I_r & O_1 \\ O_2 & O_3 \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = \begin{bmatrix} Q_{11} & O \\ Q_{21} & O \end{bmatrix}.$$

Note that

$$Q \begin{bmatrix} I_r & O \\ O & O \end{bmatrix} Q' = \begin{bmatrix} Q_{11} Q'_{11} & Q_{11} Q'_{21} \\ Q_{21} Q'_{11} & Q_{21} Q'_{21} \end{bmatrix} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} [Q'_{11}, Q'_{21}] = A_1 A'_1$$

where $A_1 = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix}$ which is a full rank $p \times r$ matrix, $r < p$, so that the r columns of A_1 are all linearly independent. This result can also be established by appealing to the fact that when $A \geq O$, its eigenvalues are non-negative, and A being symmetric, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that

$$A = P \begin{bmatrix} D & O \\ O & O \end{bmatrix} P' = \lambda_1 P_1 P'_1 + \cdots + \lambda_r P_r P'_r + 0 P_{r+1} P'_{r+1} + \cdots + 0 P_p P'_p = P_{(1)} D P'_{(1)},$$

$$D = \text{diag}(\lambda_1, \dots, \lambda_r), \quad P_{(1)} = [P_1, \dots, P_r], \quad P = [P_1, \dots, P_r, \dots, P_p],$$

where P_1, \dots, P_p are the columns of the orthonormal matrix P , $\lambda_1, \dots, \lambda_r, 0, \dots, 0$ are the eigenvalues of A where $\lambda_j > 0$, $j = 1, \dots, r$, and $P_{(1)}$ contains the first r columns of P . The first r eigenvalues must be positive since A is a non-negative definite matrix of rank r . Now, we can write $A = P_{(1)} D P'_{(1)} = A_1 A'_1$, with $A_1 = P_{(1)} D^{\frac{1}{2}}$. Observe that A_1 is $p \times r$ and of rank $r < p$. Thus, we have the following result:

Theorem 8.3.1. *Let $A = A'$ be a real $p \times p$ positive semi-definite matrix, $A \geq O$, of rank $r < p$. Then, A can be represented in the form $A = A_1 A'_1$ where A_1 is a $p \times r$, $r < p$, matrix of rank r or, equivalently, the r columns of the $p \times r$ matrix A_1 are all linearly independent.*

In the case of Wishart matrices, we can interpret Theorem 8.3.1 as follows: Let the $p \times 1$ vector random variable X_j have a nonsingular Gaussian distribution whose mean value is

the null vector and covariance matrix Σ is positive definite. Let the X_j 's, $j = 1, \dots, n$, be independently distributed, that is, $X_j \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$. Letting the $p \times n$ sample matrix be $\mathbf{X} = [X_1, \dots, X_n]$, the joint density of X_1, \dots, X_n or that of \mathbf{X} , denoted by $f(\mathbf{X})$, is the following:

$$\begin{aligned} f(\mathbf{X})d\mathbf{X} &= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{j=1}^n X_j' \Sigma^{-1} X_j} d\mathbf{X} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{X}\mathbf{X}')} d\mathbf{X}. \end{aligned} \quad (8.3.1)$$

Letting $W = \mathbf{X}\mathbf{X}'$, the $p \times p$ matrix W will be positive definite provided $n \geq p$; otherwise, that is when $n < p$, W will be singular. Let us consider the case $n \geq p$ first. This will also provide a derivation of the real Wishart density which was earlier obtained as a special case of real matrix-variate gamma density. Observe that we can write $d\mathbf{X}$ in terms of dW by applying Theorem 4.2.3, namely,

$$d\mathbf{X} = \frac{\pi^{\frac{np}{2}}}{\Gamma_p(\frac{n}{2})} |W|^{\frac{n-p+1}{2}} dW.$$

Therefore, if the density of W is denoted by $f_1(W)$, then $f_1(W)$ is available from (8.3.1) by expressing $d\mathbf{X}$ in terms of dW . That is,

$$f_1(W) = \frac{1}{2^{\frac{np}{2}} \Gamma_p(\frac{n}{2}) |\Sigma|^{\frac{n}{2}}} |W|^{\frac{n-p+1}{2}} e^{-\frac{1}{2} \text{tr}(\Sigma^{-1} W)} \quad (8.3.2)$$

for $n \geq p$, $W > O$, $\Sigma > O$, and $f_1(W) = 0$ elsewhere. This is the density of a nonsingular Wishart distribution with n degrees of freedom, $n \geq p$, and parameter matrix $\Sigma > O$, which is denoted $W \sim W_p(n, \Sigma)$, $\Sigma > O$, $n \geq p$. It has previously been shown that when $X_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $j = 1, \dots, n$, where $\mu \neq O$ is the common $p \times 1$ mean value vector and Σ is the positive definite covariance matrix,

$$W = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \sim W_p(n-1, \Sigma), \quad \Sigma > O \text{ for } n-1 \geq p,$$

where $\bar{\mathbf{X}} = [\bar{X}_1, \dots, \bar{X}_p]$, $\bar{X}_i = \frac{1}{n}(X_{1i} + \dots + X_{ni})$. Thus, we have the following result:

Theorem 8.3.2. Let $X_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n$. Let $\mathbf{X} = [X_1, \dots, X_n]$ and $W = \mathbf{X}\mathbf{X}'$. Then, when $\mu = O$, the $p \times p$ positive definite matrix $W \sim W_p(n, \Sigma)$, $\Sigma > O$ for $n \geq p$. If $\mu \neq O$, $W = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' \sim W_p(n-1, \Sigma)$, $\Sigma > O$, for $n-1 \geq p$, where $\bar{\mathbf{X}} = [\bar{X}_1, \dots, \bar{X}_p]$, $\bar{X}_i = \frac{1}{n}(X_{1i} + \dots + X_{ni})$.

Now, consider the case $n < p$. Let us denote n as $r < p$ in order to avoid any confusion with n as specified in the nonsingular case. Letting \mathbf{X} be as previously defined, \mathbf{X} is a real matrix of order $p \times r$, $n = r < p$. Let $T_1 = \mathbf{X}\mathbf{X}'$ and $T_2 = \mathbf{X}'\mathbf{X}$ where $\mathbf{X}'\mathbf{X}$ is an $r \times r$ positive definite matrix since \mathbf{X} and \mathbf{X}' are full rank matrices of rank $r < p$. Thus, all the eigenvalues of $T_2 = \mathbf{X}'\mathbf{X}$ are positive and the eigenvalues of T_1 are either positive or equal to zero since T_1 is a real positive semi-definite matrix. Letting λ be a nonzero eigenvalue of T_2 , consider the following determinant, denoted by δ , which is expanded in two ways by making use of certain properties of the determinants of partitioned matrices that are provided in Sect. 1.3:

$$\begin{aligned} \delta &= \begin{vmatrix} \sqrt{\lambda}I_p & \mathbf{X} \\ \mathbf{X}' & \sqrt{\lambda}I_r \end{vmatrix} = |\sqrt{\lambda}I_p| |\sqrt{\lambda}I_r - \mathbf{X}'(\sqrt{\lambda}I_p)^{-1}\mathbf{X}| \\ &= (\sqrt{\lambda})^{p-r} |\lambda I_r - \mathbf{X}'\mathbf{X}|; \\ \delta = 0 &\Rightarrow |\lambda I_r - \mathbf{X}'\mathbf{X}| = 0, \end{aligned} \tag{8.3.3}$$

which shows that λ is an eigenvalue of $\mathbf{X}'\mathbf{X}$. Now expand δ as follows:

$$\begin{aligned} \delta &= |\sqrt{\lambda}I_r| |\sqrt{\lambda}I_p - \mathbf{X}(\sqrt{\lambda}I_r)^{-1}\mathbf{X}'| \\ &= (\sqrt{\lambda})^{-p+r} |\lambda I_p - \mathbf{X}\mathbf{X}'|; \\ \delta = 0 &\Rightarrow |\lambda I_p - \mathbf{X}\mathbf{X}'| = 0, \end{aligned} \tag{8.3.4}$$

so that all the r nonzero eigenvalues of $T_2 = \mathbf{X}'\mathbf{X}$ are also eigenvalues of $T_1 = \mathbf{X}\mathbf{X}'$, the remaining eigenvalues of T_1 being zeros. As well, one has $|I_r - \mathbf{X}'\mathbf{X}| = |I_p - \mathbf{X}\mathbf{X}'|$. These results are next stated as a theorem.

Theorem 8.3.3. *Let \mathbf{X} be a $p \times r$ matrix of full rank $r < p$. Let the real $p \times p$ positive semi-definite matrix $T_1 = \mathbf{X}\mathbf{X}'$ and the $r \times r$ real positive definite matrix $T_2 = \mathbf{X}'\mathbf{X}$. Then, (a) the r positive eigenvalues of T_2 are identical to those of T_1 , the remaining eigenvalues of T_1 being equal to zero; (b) $|I_r - \mathbf{X}'\mathbf{X}| = |I_p - \mathbf{X}\mathbf{X}'|$.*

Additional results relating the p -variate real Gaussian distribution to the real Wishart distribution are needed in connection with the singular case. Let the $p \times 1$ vector X_j have a p -variate real Gaussian distribution whose mean value is the null vector and covariance matrix is positive definite, with $X_j \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, $j = 1, \dots, r$, $r < p$. Let $\mathbf{X} = [X_1, \dots, X_r]$ be a $p \times r$ matrix, which, in this instance, is also the sample matrix. We are seeking the distribution of $T_1 = \mathbf{X}\mathbf{X}'$ when $r < p$, where T_1 corresponds to a singular Wishart matrix. Letting T be an $r \times r$ lower triangular matrix with positive diagonal elements, and G be an $r \times p$, $r < p$, semiorthonormal matrix, that is, $GG' = I_r$, we have the representation $\mathbf{X}' = TG$, so that

$$d\mathbf{X}' = \left\{ \prod_{j=1}^r t_{jj}^{p-j} \right\} dT h(G), \quad p \geq r, \quad (8.3.5)$$

where $h(G)$ is a differential element associated with G . Then, on applying Theorem 4.2.2, we have

$$\int_{V_{r,p}} h(G) = 2^r \frac{\pi^{\frac{pr}{2}}}{\Gamma_r(\frac{p}{2})}, \quad p \geq r, \quad (8.3.6)$$

where $V_{r,p}$ is the Stiefel manifold or the space of semi-orthonormal $r \times p$, $r < p$, matrices. Observe that the density of \mathbf{X}' , denoted by $f_{\mathbf{X}'}(\mathbf{X}')$, is the following:

$$f_{\mathbf{X}'}(\mathbf{X}') d\mathbf{X}' = \frac{e^{-\frac{1}{2} \sum_{j=1}^r X_j' \Sigma^{-1} X_j}}{(2\pi)^{\frac{pr}{2}} |\Sigma|^{\frac{r}{2}}} d\mathbf{X}' = \frac{e^{-\frac{1}{2} \text{tr}(\mathbf{X}' \Sigma^{-1} \mathbf{X})}}{(2\pi)^{\frac{pr}{2}} |\Sigma|^{\frac{r}{2}}} d\mathbf{X}'.$$

Let $T_2 = \mathbf{X}' \Sigma^{-1} \mathbf{X}$ or simply $T_2 = \mathbf{X}' \mathbf{X}$ when $\Sigma = I$; note that Σ will vanish upon letting $\mathbf{Y} = \Sigma^{-\frac{1}{2}} \mathbf{X} \Rightarrow d\mathbf{Y} = |\Sigma|^{-\frac{r}{2}} d\mathbf{X}$. Now, on expressing $d\mathbf{X}'$ in terms of dT_2 by making use of Theorem 4.2.3, the following result is obtained:

Theorem 8.3.4. Let $X_j \stackrel{iid}{\sim} N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \dots, r$, $r < p$. Let $\mathbf{X} = [X_1, \dots, X_r]$, $\bar{X} = \frac{1}{r}(X_1 + \dots + X_r)$ and $\bar{\mathbf{X}} = (\bar{X}, \dots, \bar{X})$. Letting $T_2 = \mathbf{X}' \Sigma^{-1} \mathbf{X}$ or $T_2 = \mathbf{X}' \mathbf{X}$ when $\Sigma = I$, T_2 has the following density, denoted by $f_t(T_2)$, when $\mu = O$:

$$f_t(T_2) = \frac{1}{2^{\frac{pr}{2}} \Gamma_r(\frac{p}{2})} |T_2|^{\frac{p-r+1}{2}} e^{-\frac{1}{2} \text{tr}(T_2)}, \quad T_2 > O, \quad r \leq p, \quad (8.3.7)$$

and zero elsewhere. Note that the $r \times r$ matrix $T_2 = \mathbf{X}' \mathbf{X}$ or $T_2 = \mathbf{X}' \Sigma^{-1} \mathbf{X}$ when $\Sigma \neq I$, has a Wishart distribution with p degrees of freedom and parameter matrix I , that is, $T_2 \sim W_r(p, I)$, $r \leq p$. When $\mu \neq O$, $T_2 = (\mathbf{X} - \bar{\mathbf{X}})' \Sigma^{-1} (\mathbf{X} - \bar{\mathbf{X}})$ has a Wishart distribution with $p-1$ degrees of freedom or, equivalently, $T_2 \sim W_r(p-1, I)$, $r \leq p-1$.

8.3.1. Singular Wishart and matrix-variate gamma distributions, real case

We now consider the case of a singular matrix-variate gamma distribution. Let the $p \times 1$ vector X_j have a p -variate real Gaussian distribution whose mean value is the null vector and covariance matrix is positive definite, with $X_j \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, $j = 1, \dots, r$, $r < p$. For convenience, let $\Sigma = I_p$. Let $\mathbf{X} = [X_1, \dots, X_r]$ be the $p \times r$ sample matrix. Then, for $r \geq p$, $\mathbf{X} \mathbf{X}'$ is distributed as a Wishart matrix with $r \geq p$ degrees of freedom, that is, $\mathbf{X} \mathbf{X}' \sim W_p(r, I)$, $r \geq p$. This result still holds for any positive definite matrix Σ ; it suffices then to replace I by Σ . What about the distribution of $\mathbf{X} \mathbf{X}'$ if $r < p$,

which corresponds to the singular case? In this instance, the real matrix $\mathbf{X}\mathbf{X}' \geq O$ (positive semi-definite) and the density of \mathbf{X} , denoted by $f_1(\mathbf{X})$, is the following:

$$f_1(\mathbf{X}) d\mathbf{X} = \frac{e^{-\frac{1}{2}\text{tr}(\mathbf{X}\mathbf{X}')}}{(2\pi)^{\frac{rp}{2}}} d\mathbf{X}. \quad (8.3.8)$$

Let W_2 be a $p \times p$ nonsingular Wishart matrix with $n \geq p$ degrees of freedom, that is, $W_2 \sim W_p(n, I)$. Let \mathbf{X} and W_2 be independently distributed. Then, the joint density of \mathbf{X} and W_2 , denoted by $f_2(\mathbf{X}, W_2)$, is given by

$$f_2(\mathbf{X}, W_2) = \frac{e^{-\frac{1}{2}\text{tr}(\mathbf{X}\mathbf{X}' + W_2)}}{(2\pi)^{\frac{pr}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})} |W_2|^{\frac{n}{2} - \frac{p+1}{2}}$$

for $W_2 > O$, $\mathbf{X}\mathbf{X}' \geq O$, $n \geq p$, $r < p$. Letting $U = \mathbf{X}\mathbf{X}' + W_2 > O$, and the joint density of U and \mathbf{X} be denoted by $f_3(\mathbf{X}, U)$, we have

$$f_3(\mathbf{X}, U) = \frac{e^{-\frac{1}{2}\text{tr}(U)}}{(2\pi)^{\frac{pr}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})} |U - \mathbf{X}\mathbf{X}'|^{\frac{n}{2} - \frac{p+1}{2}}, \quad n \geq p, \quad r < p,$$

where

$$|U - \mathbf{X}\mathbf{X}'| = |U| |I - U^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'U^{-\frac{1}{2}}|.$$

Letting $V = U^{-\frac{1}{2}}\mathbf{X}$ for fixed U , $dV = |U|^{\frac{r}{2}} d\mathbf{X}$, and the joint density of U and V , denoted by $f_4(U, V)$, is then

$$f_4(U, V) = \frac{|U|^{\frac{n+r}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(U)}}{(2\pi)^{\frac{pr}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})} |I - VV'|^{\frac{n}{2} - \frac{p+1}{2}}. \quad (8.3.9)$$

Note that U and V are independently distributed. By integrating out U with the help of a real matrix-variate gamma integral, we obtain the density of V , denoted by $f_5(V)$, as

$$f_5(V) = c |I - VV'|^{\frac{n}{2} - \frac{p+1}{2}} = c |I - V'V|^{\frac{n}{2} - \frac{p+1}{2}}, \quad (8.3.10)$$

in view of Theorem 8.3.3(b), where V is $p \times r$, $r < p$, c being the normalizing constant. Thus, we have the following result:

Theorem 8.3.5. *Let $\mathbf{X} = [X_1, \dots, X_r]$ and $X_j \stackrel{iid}{\sim} N_p(O, I)$, $j = 1, \dots, r$, $r < p$. Let the $p \times p$ real positive definite matrix W_2 be Wishart distributed with n degrees of freedom, that is, $W_2 \sim W_p(n, I)$. Let $U = \mathbf{X}\mathbf{X}' + W_2 > O$ and let $V = U^{-\frac{1}{2}}\mathbf{X}$ where V*

is a $p \times r$, $r < p$, matrix of full rank r . Observe that $VV' \geq O$ (positive semi-definite). Then, the densities of the $r \times p$ matrix V' and the matrix $S = V'V$, respectively denoted by $f_5(V)$ and $f_6(S)$, are as follows, U and V being independently distributed:

$$f_5(V) = \frac{\Gamma_r(\frac{p}{2})}{\pi^{\frac{pr}{2}}} \frac{\Gamma_r(\frac{n+r}{2})}{\Gamma_r(\frac{p}{2})\Gamma_r(\frac{n-p+r}{2})} |I - V'V|^{\frac{n-p+r}{2}}, \quad (8.3.11)$$

and

$$f_6(S) = \frac{\Gamma_r(\frac{n+r}{2})}{\Gamma_r(\frac{p}{2})\Gamma_r(\frac{n-p+r}{2})} |S|^{\frac{p}{2}-\frac{r+1}{2}} |I - S|^{\frac{n-p+r}{2}-\frac{r+1}{2}}, \quad n \geq p, \quad r < p, \quad S > O. \quad (8.3.12)$$

Proof: In light of Theorem 8.3.3(b), $|I_p - VV'| = |I_r - V'V|$. Observe that VV' is $p \times p$ and positive semi-definite whereas $V'V$ is $r \times r$ and positive definite. As well, $\frac{n}{2} - \frac{p+1}{2} = \frac{n-p+r}{2} - \frac{r+1}{2}$. Letting $S = V'V$ and expressing dV' in terms of dS by applying Theorem 4.2.3, then for $r < p$,

$$dV' = \frac{\pi^{\frac{rp}{2}}}{\Gamma_r(\frac{p}{2})} |S|^{\frac{p}{2}-\frac{r+1}{2}} dS. \quad (i)$$

Now, integrating out S by making use of a type-1 beta integral, we have

$$\int_{S>O} |S|^{\frac{p}{2}-\frac{r+1}{2}} |I - S|^{\frac{n-p+r}{2}-\frac{r+1}{2}} dS = \frac{\Gamma_r(\frac{p}{2})\Gamma_r(\frac{n-p+r}{2})}{\Gamma_r(\frac{n+r}{2})} \quad (ii)$$

for $n \geq p$, $r < p$. The normalizing constants in (8.3.11) and (8.3.12) follow from (i) and (ii). This completes the proofs.

At this juncture, we are considering the singular version of the determinantal equation in (8.2.8). Let W_1 and W_2 be independently distributed $p \times p$ real matrices where $W_1 = \mathbf{X}\mathbf{X}'$, $\mathbf{X} = [X_1, \dots, X_r]$ with $X_j \stackrel{iid}{\sim} N_p(O, I)$, $j = 1, \dots, r$, $r < p$, and the positive definite matrix $W_2 \sim W_p(n, I)$, $n \geq p$. The equation

$$\begin{aligned} |W_1 - \mu(W_1 + W_2)| = O &\Rightarrow |\mathbf{X}\mathbf{X}' - \mu(\mathbf{X}\mathbf{X}' + W_2)| = 0 \\ &\Rightarrow |U^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'U^{-\frac{1}{2}} - \mu I_p| = 0, \quad U = \mathbf{X}\mathbf{X}' + W_2 > O \\ &\Rightarrow |VV' - \mu I_p| = 0, \end{aligned}$$

which, in turn, implies that μ is an eigenvalue of $VV' \geq O$ and all the eigenvalues are positive or zero. However, it follows from (8.3.3) and (8.3.4) that

$$|VV' - \mu I_p| = 0 \Rightarrow |V'V - \mu I_r| = 0.$$

Hence, the following result:

Theorem 8.3.6. Let \mathbf{X} be a $p \times r$ matrix whose columns are iid $N_p(O, I)$ and $r < p$. Let $W_1 = \mathbf{X}\mathbf{X}' \geq O$ which is a $p \times p$ positive semi-definite matrix, $W_2 > O$ be a $p \times p$ Wishart distributed matrix having n degrees of freedom, that is, $W_2 \sim W_p(n, I)$, $n \geq p$, and let W_1 and W_2 be independently distributed. Then, $U = W_1 + W_2 > O$ and $V = U^{-\frac{1}{2}}\mathbf{X}$ are independently distributed. Moreover,

$$|W_1 - \mu(W_1 + W_2)| = 0 \Rightarrow |V'V - \mu I_r| = 0$$

where the roots $\mu_j > 0$, $j = 1, \dots, r$, are the eigenvalues of $V'V > O$, and the eigenvalues of VV' are $\mu_j > 0$, $j = 1, \dots, r$, with the remaining $p - r$ eigenvalues of VV' being equal to zero.

Let $S = PDP'$, $D = \text{diag}(\mu_1, \dots, \mu_r)$, $PP' = I_r$, $P'P = I_r$. Then, on applying Theorem 8.2.2,

$$dS = \frac{\pi^{\frac{r^2}{2}}}{\Gamma_r(\frac{r}{2})} \left[\prod_{i < j} (\mu_i - \mu_j) \right] dD$$

after integrating out over the differential element associated with the orthonormal matrix P . Substituting in $f_6(S)$ of Theorem 8.3.5 yields the following result:

Theorem 8.3.7. Let μ_1, \dots, μ_r be the nonzero roots of the determinantal equation $|W_1 - \mu(W_1 + W_2)| = 0$ where $W_1 = \mathbf{X}\mathbf{X}'$, $\mathbf{X} = [X_1, \dots, X_r]$, $X_j \stackrel{iid}{\sim} N_p(O, I)$, $j = 1, \dots, r < p$, $W_2 \sim W_p(n, I)$, $n \geq p$, and W_1 and W_2 be independently distributed. Letting $\mu_1 > \mu_2 > \dots > \mu_r > 0$, $r < p$, the joint density of the nonzero roots μ_1, \dots, μ_r , denoted by $f_\mu(\mu_1, \dots, \mu_r)$, is given by

$$f_\mu(\mu_1, \dots, \mu_r) = \frac{\pi^{\frac{r^2}{2}}}{\Gamma_r(\frac{r}{2})} \frac{\Gamma_r(\frac{n+r}{2})}{\Gamma_r(\frac{p}{2})\Gamma_r(\frac{n-p+r}{2})} \times \left[\prod_{j=1}^r \mu_j^{\frac{p}{2}-\frac{r+1}{2}} \right] \left[\prod_{j=1}^r (1 - \mu_j)^{\frac{n-p+r}{2}-\frac{r+1}{2}} \right] \left[\prod_{i < j} (\mu_i - \mu_j) \right]. \quad (8.3.13)$$

It can readily be observed from (8.3.9) that $U = \mathbf{X}\mathbf{X}' + W_2$ and $V = U^{-\frac{1}{2}}\mathbf{X}$ are indeed independently distributed.

8.3.2. A direct evaluation as an eigenvalue problem

Consider the singular version of the original determinantal equation

$$|W_1 - \lambda W_2| = 0 \Rightarrow |W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}} - \lambda I| = 0 \quad (8.3.14)$$

where W_1 is singular, W_2 is nonsingular, and W_1 and W_2 are independently distributed. Thus, the roots λ_j 's of the equation $|W_1 - \lambda W_2| = 0$ coincide with the eigenvalues of the matrix $U = W_2^{-\frac{1}{2}} W_1 W_2^{-\frac{1}{2}}$. Let W_2 be a real nonsingular matrix-variate gamma or Wishart distributed matrix. Let $W_1 = \mathbf{X}\mathbf{X}'$, $\mathbf{X} = [X_1, \dots, X_r]$, $X_j \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, $j = 1, \dots, r$, $r < p$, or, equivalently, the $p \times r$, $r < p$, matrix \mathbf{X} is a simple random sample from this p -variate real Gaussian population. We will take $\Sigma = I$ without any loss of generality. In this case, \mathbf{X} is a $p \times r$ full rank matrix with $r < p$. Let $W_2 \sim W_p(n, I)$, $n \geq p$, that is, W_2 is a nonsingular Wishart matrix with n degrees of freedom and parameter matrix I , and $W_1 \geq O$ (positive semi-definite). Then, the joint density of \mathbf{X} and W_2 , denoted by $f_7(\mathbf{X}, W_2)$, is the following:

$$f_7(\mathbf{X}, W_2) = \frac{e^{-\frac{1}{2}\text{tr}(\mathbf{X}\mathbf{X}')} |W_2|^{\frac{n}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(W_2)}}{(2\pi)^{\frac{nr}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})}. \quad (i)$$

Consider the exponent

$$-\frac{1}{2}\text{tr}(\mathbf{X}\mathbf{X}' + W_2) = -\frac{1}{2}\text{tr}[W_2(I + W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}})] = -\frac{1}{2}\text{tr}(W_2(I + VV'))$$

where $V = W_2^{-\frac{1}{2}}\mathbf{X} \Rightarrow d\mathbf{X} = |W_2|^{\frac{r}{2}} dV$ for fixed W_2 . The joint density of W_2 and V , denoted by $f_8(V, W_2)$, is then

$$f_8(V, W_2) = \frac{|W_2|^{\frac{n}{2} + \frac{r}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(W_2(I+VV'))}}{(2\pi)^{\frac{pr}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})}. \quad (ii)$$

Observe that $VV' = W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}} = U$ of (8.3.14). Integrating out W_2 in (ii) by using a real matrix-variate gamma integral, we obtain the marginal density of V , denoted by $f_9(V)$, as

$$f_9(V)dV = \frac{\Gamma_p(\frac{n+r}{2})}{\Gamma_p(\frac{n}{2})\pi^{\frac{pr}{2}}} |I + VV'|^{-\frac{n+r}{2}} dV \quad (iii)$$

where $VV' \geq O$ (positive semi-definite). Note that

$$|I_p + VV'| = |I_r + V'V|, \quad VV' \geq O, \quad V'V > O \text{ (positive definite),}$$

which follows from Theorem 8.3.3(b). This last result can also be established by expanding the following determinant in two ways as was done in (8.3.3) and (8.3.4):

$$\begin{aligned} \begin{vmatrix} I_p & V \\ -V' & I_r \end{vmatrix} &= |I_p| |I_r + V'V| = |I_r| |I_p + VV'| \\ &\Rightarrow |I_r + V'V| = |I_p + VV'|. \end{aligned}$$

Hence, the density of V' must be of the following form where c_1 is the normalizing constant:

$$f_{10}(V')dV' = c_1 |I + V'V|^{-\left(\frac{n+r}{2}\right)} dV', \quad n \geq p, \quad r < p. \quad (iv)$$

On applying Theorem 4.2.3, (iv) can be transformed into a function of $S_1 = V'V > O$, S_1 being of order $r \times r$. Then,

$$dV' = \frac{\pi^{\frac{pr}{2}}}{\Gamma_r\left(\frac{p}{2}\right)} |S_1|^{\frac{p}{2} - \frac{r+1}{2}} dS_1.$$

Substituting the above expression for dV' in $f_{10}(V')$ and then integrating over the $r \times r$ matrix $S_1 > O$, we have

$$\int_{V'} f_{10}(V')dV' = c_1 \frac{\pi^{\frac{pr}{2}}}{\Gamma_r\left(\frac{p}{2}\right)} \frac{\Gamma_r\left(\frac{p}{2}\right)\Gamma_r\left(\frac{n+r-p}{2}\right)}{\Gamma_r\left(\frac{n+r}{2}\right)} = 1.$$

Accordingly, the density of V' is the following:

$$f_{10}(V')dV' = \frac{\Gamma_r\left(\frac{p}{2}\right)}{\pi^{\frac{pr}{2}}} \frac{\Gamma_r\left(\frac{n+r}{2}\right)}{\Gamma_r\left(\frac{p}{2}\right)\Gamma_r\left(\frac{n+r-p}{2}\right)} |I + V'V|^{-\left(\frac{n+r}{2}\right)} dV'. \quad (8.3.15)$$

Note that $\Gamma_r\left(\frac{p}{2}\right)$ cancels out. Then, by re-expressing dV' in terms of dS_1 in (8.3.15), the density of $S_1 = V'V$ is obtained as

$$f_{11}(S_1) = \frac{\Gamma_r\left(\frac{n+r}{2}\right)}{\Gamma_r\left(\frac{p}{2}\right)\Gamma_r\left(\frac{n+r-p}{2}\right)} |S_1|^{\frac{p}{2} - \frac{r+1}{2}} |I + S_1|^{-\left(\frac{n+r}{2}\right)} \quad (8.3.16)$$

for $S_1 = V'V > O$, $n \geq p$, $r < p$, and zero elsewhere, which is a real $r \times r$ matrix-variate type-2 beta density with the parameters $\left(\frac{p}{2}, \frac{n+r-p}{2}\right)$. Thus, the following result:

Theorem 8.3.8. Let $W_1 = \mathbf{X}\mathbf{X}'$, $\mathbf{X} = [X_1, \dots, X_r]$, $X_j \stackrel{iid}{\sim} N_p(O, I)$, $j = 1, \dots, r$, $r < p$, $W_2 \sim W_p(n, I)$, $n \geq p$, and W_1 and W_2 be independently distributed. Let

$V = W_2^{-\frac{1}{2}}X$, $VV' \geq O$ and $S_1 = V'V > O$. Then, $|I_p + VV'| = |I_r + V'V| = |I_r + S_1|$. The density of V' is given in (8.3.15) and that of S_1 , which is specified in (8.3.16), is a real matrix-variate type-2 beta density with the parameters $(\frac{p}{2}, \frac{n+r-p}{2})$. Moreover, the positive semi-definite matrix $S_2 = W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}}$ is distributed, almost surely, as S_1 , which has a nonsingular real matrix-variate type-2 beta distribution with the parameters $(\frac{p}{2}, \frac{n+r-p}{2})$.

Observe that this theorem also holds when $X_j \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$ and $W_2 \sim W_p(n, \Sigma)$, $\Sigma > O$, and the distribution of S_2 will still be free of Σ . Converting (8.3.16) in terms of the eigenvalues of S_1 , which are also the nonzero eigenvalues of $S_2 = W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}}$ of (8.3.14), we have the following density, denoted by $f_{12}(\lambda_1, \dots, \lambda_r)dD$, $D = \text{diag}(\lambda_1, \dots, \lambda_r)$, assuming that the eigenvalues are distinct and such that $\lambda_1 > \lambda_2 > \dots > \lambda_r > 0$:

$$f_{12}(\lambda_1, \dots, \lambda_r)dD = \frac{\Gamma_r(\frac{n+r}{2})}{\Gamma_r(\frac{p}{2})\Gamma_r(\frac{n+r-p}{2})} \frac{\pi^{r^2}}{\Gamma_r(\frac{r}{2})} \\ \times \left[\prod_{j=1}^r \lambda_j^{\frac{p}{2} - \frac{r+1}{2}} \right] \left[\prod_{j=1}^r (1 + \lambda_j)^{-\frac{(n+r)}{2}} \right] \left[\prod_{i < j} (\lambda_i - \lambda_j) \right] dD. \quad (8.3.17)$$

Theorem 8.3.9. Let W_2 and \mathbf{X} be as defined in (8.3.14). Then, the joint density of the nonzero eigenvalues $\lambda_1, \dots, \lambda_r$ of $S_2 = W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}}$, which are assumed to be distinct and such that $\lambda_1 > \dots > \lambda_r > 0$, is given in (8.3.17).

In (8.3.13) we have obtained the joint density of the nonzero roots $\mu_1 > \dots > \mu_r$ of the determinantal equation

$$|\mathbf{X}\mathbf{X}' - \mu(\mathbf{X}\mathbf{X}' + W_2)| = 0 \Rightarrow |\mathbf{X}\mathbf{X}' - \lambda W_2| = 0, \quad \lambda = \frac{\mu}{1 - \mu}, \quad \mu = \frac{\lambda}{1 + \lambda}, \\ \Rightarrow |W_2^{-\frac{1}{2}}\mathbf{X}\mathbf{X}'W_2^{-\frac{1}{2}} - \lambda I| = 0. \quad (8.3.18)$$

Hence, making the substitution $\mu_j = \frac{\lambda_j}{1 + \lambda_j}$ in (8.3.13) should yield the density appearing in (8.3.17). This will be stated as a theorem.

Theorem 8.3.10. When $\mu_j = \frac{\lambda_j}{1 + \lambda_j}$ or $\lambda_j = \frac{\mu_j}{1 - \mu_j}$, the distributions of the μ_j 's or the λ_j 's, as respectively specified in (8.3.13) and (8.3.17), coincide.

8.3a. The Singular Complex Case

The matrix manipulations that were utilized in the real case also apply in the complex domain. The following result parallels Theorem 8.3.1:

Theorem 8.3a.1. *Let the $p \times p$ matrix $A = A^* \geq O$ be a Hermitian positive semi-definite matrix of rank $r < p$, where A^* designate the conjugate transpose of A . Then, A can be represented as $A = A_1 A_1^*$ where A_1 is $p \times r$, $r < p$, of rank r , that is, all the columns of A_1 are linearly independent.*

A derivation of the Wishart matrix in the complex case can also be worked out from a complex Gaussian distribution. In earlier chapters, we have derived the Wishart density as a particular case of the matrix-variate gamma density, whether in the real or complex domain. Let the $p \times 1$ complex vectors \tilde{X}_j , $j = 1, \dots, n$, be independently distributed as p -variate complex Gaussian random variables with the null vector as their mean value and a common Hermitian positive definite covariance matrix, that is, $\tilde{X}_j \stackrel{iid}{\sim} \tilde{N}_p(O, \Sigma)$, $\Sigma = \Sigma^* > O$ for $j = 1, \dots, n$. Let the $p \times n$ matrix $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_n]$ be the simple random sample matrix from this complex Gaussian population. Then, the density of $\tilde{\mathbf{X}}$, denoted by $\tilde{f}(\tilde{\mathbf{X}})$, is given by

$$\tilde{f}(\tilde{\mathbf{X}}) = \frac{e^{-\sum_{j=1}^n \tilde{X}_j^* \Sigma^{-1} \tilde{X}_j}}{\pi^{np} |\det(\Sigma)|^n} = \frac{e^{-\text{tr}(\Sigma^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^*)}}{\pi^{np} |\det(\Sigma)|^n}, \quad (8.3a.1)$$

for $n \geq p$. Let the $p \times p$ Hermitian positive definite matrix $\tilde{\mathbf{X}} \tilde{\mathbf{X}}^* = \tilde{W}$. For $n \geq p$, it follows from Theorem 4.2a.3 that

$$d\tilde{\mathbf{X}} = \frac{\pi^{np}}{\tilde{\Gamma}_p(n)} |\det(\tilde{W})|^{n-p} d\tilde{W} \quad (8.3a.2)$$

where $d\tilde{\mathbf{X}} = dY_1 \wedge dY_2$, $\tilde{\mathbf{X}} = Y_1 + iY_2$, $i = \sqrt{-1}$, Y_1, Y_2 being real $p \times n$ matrices. Given (8.3a.1) and (8.3a.2), the density of \tilde{W} , denoted by $\tilde{f}_1(\tilde{W})$, is obtained as

$$\tilde{f}_1(\tilde{W}) = \frac{|\det(\tilde{W})|^{n-p} e^{-\text{tr}(\Sigma^{-1} \tilde{W})}}{\tilde{\Gamma}_p(n) |\det(\Sigma)|^n}, \quad \tilde{W} = \tilde{W}^* > O, \quad n \geq p, \quad (8.3a.3)$$

and zero elsewhere; this is the complex Wishart density with n degrees of freedom and parameter matrix $\Sigma > O$, which is written as $\tilde{W} \sim \tilde{W}_p(n, \Sigma)$, $\Sigma > O$, $n \geq p$. If $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$, then letting $\tilde{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \dots + \tilde{X}_n)$, $\tilde{\tilde{\mathbf{X}}} = (\tilde{\tilde{X}}, \dots, \tilde{\tilde{X}})$ and

$$\tilde{W} = (\tilde{\mathbf{X}} - \tilde{\tilde{\mathbf{X}}})(\tilde{\mathbf{X}} - \tilde{\tilde{\mathbf{X}}})^*, \quad (8.3a.4)$$

we have $\tilde{W} \sim \tilde{W}_p(n-1, \Sigma)$, $n-1 \geq p$, $\Sigma > O$, or \tilde{W} has a Wishart distribution with $n-1$ instead of n degrees of freedom, $\tilde{\mu} \neq O$ being eliminated by subtracting the sample mean. The remainder of this section is devoted to the distribution of $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$ for $n < p$, that is, in the singular case. Proceeding as in the real case, we have

$$\det(I_p - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*) = \det(I_r - \tilde{\mathbf{X}}^*\tilde{\mathbf{X}}) \quad (8.3a.5)$$

where $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^* \geq O$ is $p \times p$, whereas the $r \times r$, $r < p$, matrix $\tilde{\mathbf{X}}^*\tilde{\mathbf{X}} > O$. Then, we have the following result:

Theorem 8.3a.2. *Let $\tilde{T}_1 = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$ and $\tilde{T}_2 = \tilde{\mathbf{X}}^*\tilde{\mathbf{X}}$. Then, the eigenvalues of \tilde{T}_2 are all real and positive and the nonzero eigenvalues of \tilde{T}_1 are identical to those of \tilde{T}_2 , the remaining ones being equal to zero.*

The complex counterpart of Theorem 8.3.4 that follows can be derived using steps parallel to those utilized in the real case.

Theorem 8.3a.3. *Let the $p \times 1$ complex vectors $\tilde{X}_j \stackrel{iid}{\sim} \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$, $j = 1, \dots, r$. Let $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_r]$ be the simple random sample matrix from this complex p -variate Gaussian population. Let $\tilde{T}_2 = \tilde{\mathbf{X}}^* \Sigma^{-1} \tilde{\mathbf{X}}$ or $\tilde{T}_2 = \tilde{\mathbf{X}}^* \tilde{\mathbf{X}}$ if $\Sigma = I_p$. Then, the density of \tilde{T}_2 , denoted by $\tilde{f}_u(\tilde{T}_2)$, is the following:*

$$\tilde{f}_u(\tilde{T}_2) = \frac{1}{\tilde{\Gamma}_r(p)} |\det(\tilde{T}_2)|^{p-r} e^{-\text{tr}(\tilde{T}_2)}, \quad \tilde{T}_2 > O, \quad r \leq p, \quad (8.3a.6)$$

so that $\tilde{T}_2 \sim \tilde{W}_r(p, I)$, that is, \tilde{T}_2 has a complex Wishart distribution with p degrees of freedom. If $\tilde{\mu} \neq O$, let $\tilde{T}_2 = (\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})^* (\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})$ or $\tilde{T}_2 = (\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})^* \Sigma^{-1} (\tilde{\mathbf{X}} - \tilde{\bar{\mathbf{X}}})$ when $\Sigma \neq I$ with $\tilde{\bar{\mathbf{X}}} = (\tilde{X}, \dots, \tilde{X})$ wherein $\tilde{X} = \frac{1}{r} (\tilde{X}_1 + \dots + \tilde{X}_r)$. Then, $\tilde{T}_2 \sim \tilde{W}_r(p-1, I)$, $r \leq p-1$.

8.3a.1. Singular gamma or singular Gaussian distribution, complex case

Let $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_r]$ where the $p \times 1$ vectors $\tilde{X}_1, \dots, \tilde{X}_r$ are independently distributed as complex Gaussian vectors whose mean value is the null vector and covariance matrix is I , that is, $\tilde{X}_j \stackrel{iid}{\sim} \tilde{N}_p(O, I)$, $j = 1, \dots, r$, $r < p$. Then, the density of $\tilde{\mathbf{X}}$, denoted by $\tilde{f}_1(\tilde{\mathbf{X}})$, is

$$\tilde{f}_1(\tilde{\mathbf{X}}) = \frac{e^{-\text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)}}{\pi^{np}}. \quad (8.3a.7)$$

Let $r < p$ so that $\tilde{\mathbf{X}}\tilde{\mathbf{X}}^* \geq O$ (Hermitian positive semi-definite). Let the $p \times p$ Hermitian positive definite matrix \tilde{W}_2 have a Wishart density with $n \geq p$ degrees of freedom and parameter matrix I , that is, $\tilde{W}_2 \sim \tilde{W}_p(n, I)$, $n \geq p$. Further assume that $\tilde{\mathbf{X}}$ and \tilde{W}_2 are

independently distributed. Then, the joint density of $\tilde{\mathbf{X}}$ and \tilde{W}_2 , denoted by $\tilde{f}_2(\tilde{\mathbf{X}}, \tilde{W}_2)$, is the following:

$$\tilde{f}_2(\tilde{\mathbf{X}}, \tilde{W}_2) = \frac{e^{-\text{tr}(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^* + \tilde{W}_2)} |\det(\tilde{W}_2)|^{n-p}}{\pi^{np} \tilde{\Gamma}_p(n)}, \quad n \geq p, \quad r < p, \quad (8.3a.8)$$

where $\tilde{\mathbf{X}}$ is $p \times r$, $r < p$. Letting the $p \times p$ matrix $\tilde{U} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^* + \tilde{W}_2 > O$, the joint density of \tilde{U} and $\tilde{\mathbf{X}}$, denoted by $\tilde{f}_3(\tilde{\mathbf{X}}, \tilde{U})$, is given by

$$\tilde{f}_3(\tilde{\mathbf{X}}, \tilde{U}) = \frac{e^{-\text{tr}(\tilde{U})} |\det(\tilde{U} - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)|^{n-p}}{\pi^{np} \tilde{\Gamma}_p(n)}, \quad n \geq p, \quad r < p.$$

Letting $\tilde{U} = \tilde{U}^* > O$, one has

$$\begin{aligned} |\det(\tilde{U} - \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*)| &= |\det(\tilde{U})| |\det(I - \tilde{U}^{-\frac{1}{2}} \tilde{\mathbf{X}}\tilde{\mathbf{X}}^* \tilde{U}^{-\frac{1}{2}})| \\ &= |\det(\tilde{U})| |\det(I - \tilde{V}\tilde{V}^*)|, \quad \tilde{V} = \tilde{U}^{-\frac{1}{2}} \tilde{\mathbf{X}}, \end{aligned}$$

where \tilde{V} is a $p \times r$ matrix of rank $r < p$. Since $d\tilde{\mathbf{X}} = |\det(\tilde{U})|^r d\tilde{V}$ for fixed \tilde{U} , the joint density of \tilde{U} and \tilde{V} is as follows:

$$\tilde{f}_4(\tilde{U}, \tilde{V}) = \frac{|\det(\tilde{U})|^{n+r-p} e^{-\text{tr}(\tilde{U})}}{\pi^{np} \tilde{\Gamma}_p(n)} |\det(I_p - \tilde{V}\tilde{V}^*)|^{n-p}. \quad (8.3a.9)$$

As has been previously noted,

$$|\det(I_p - \tilde{V}\tilde{V}^*)| = |\det(I_r - \tilde{V}^*\tilde{V})|$$

where $\tilde{V}^*\tilde{V}$ is an $r \times r$ Hermitian positive definite matrix. Since $\tilde{f}_4(\tilde{U}, \tilde{V})$ can be factorized, \tilde{U} and \tilde{V} are independently distributed, and the marginal density of \tilde{V} , denoted by $\tilde{f}_5(\tilde{V})$, is of the following form, after integrating out \tilde{U} with the help of a complex matrix-variate gamma integral:

$$\tilde{f}_5(\tilde{V}^*)d\tilde{V} = \tilde{c} |\det(I - \tilde{V}\tilde{V}^*)|^{n-p} d\tilde{V} = \tilde{c} |\det(I - \tilde{V}^*\tilde{V})|^{n-p} d\tilde{V}^* \quad (8.3a.10)$$

where \tilde{c} is the normalizing constant. Now, proceeding as in the real case, the following result is obtained:

Theorem 8.3a.4. *Let the $p \times 1$ complex vectors $\tilde{X}_j \stackrel{iid}{\sim} \tilde{N}_p(O, I)$, $j = 1, \dots, r$, and $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_r]$ be the $p \times r$ full rank sample matrix with $r < p$. Let \tilde{W}_2 be a $p \times p$*

Hermitian positive definite matrix having a nonsingular Wishart density with n degrees of freedom and parameter matrix I , that is, $\tilde{W}_2 \sim \tilde{W}_p(n, I)$, $n \geq p$, and assume that \tilde{W}_2 and $\tilde{\mathbf{X}}$ are independently distributed. Let $\tilde{U} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^* + \tilde{W}_2$ be Hermitian positive definite, $\tilde{V} = \tilde{U}^{-\frac{1}{2}}\tilde{\mathbf{X}}$ and $\tilde{S} = \tilde{V}^*\tilde{V}$. Then, \tilde{U} and \tilde{V} are independently distributed and the densities of \tilde{V} and \tilde{S} , respectively denoted by $f_5(\tilde{V})$ and $f_6(\tilde{S})$, are

$$f_5(\tilde{V}) = \frac{\tilde{\Gamma}_r(p)}{\pi^{pr}} \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p)\tilde{\Gamma}_r(n-p+r)} |\det(I - \tilde{V}^*\tilde{V})|^{n-p}, \quad r < p, \quad (8.3a.11)$$

and

$$f_6(\tilde{S}) = \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p)\tilde{\Gamma}_r(n-p+r)} |\det(\tilde{S})|^{p-r} |\det(I - \tilde{S})|^{n-p}, \quad n \geq p, \quad r < p, \quad (8.3a.12)$$

observing that $n - p = (n + r - p) - r$.

Let $\tilde{W}_1 = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$ where the $p \times r$ matrix $\tilde{\mathbf{X}}$ is the previously defined sample matrix arising from a standard complex Gaussian population. Let $\tilde{W}_2 \sim \tilde{W}_p(n, I)$ and assume that \tilde{W}_1 and \tilde{W}_2 are independently distributed. Letting $\tilde{U} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^* + \tilde{W}_2 > O$, consider the determinantal equation

$$\det(\tilde{W}_1 - \mu(\tilde{W}_1 + \tilde{W}_2)) = 0. \quad (i)$$

Then, as in the real case, the following result can be obtained:

Theorem 8.3a.5. *Let \tilde{W}_1 , \tilde{W}_2 , \tilde{V} and \tilde{U} be as previously defined. Then,*

$$\det(\tilde{W}_1 - \mu(\tilde{W}_1 + \tilde{W}_2)) = 0 \Rightarrow \det(\tilde{V}^*\tilde{V} - \mu I_r) = 0. \quad (ii)$$

This establishes that the roots μ_j 's of the determinantal equation (i) coincide with the eigenvalues of $\tilde{V}^*\tilde{V}$, and since $\tilde{V}^*\tilde{V} > O$, the eigenvalues are real and positive. Let the eigenvalues be distinct, in which case $\mu_1 > \dots > \mu_r > 0$. Then, steps parallel to those utilized in the real case will yield the following result:

Theorem 8.3a.6. *Let μ_1, \dots, μ_r be the nonzero roots of the equation*

$$\det(\tilde{W}_1 - \mu(\tilde{W}_1 + \tilde{W}_2)) = 0$$

where \tilde{W}_1 and \tilde{W}_2 are as previously defined. Let $\mu_1 > \dots > \mu_r > 0$, $r < p$, and let $D = \text{diag}(\mu_1, \dots, \mu_r)$. Then, the joint density of the eigenvalues μ_1, \dots, μ_r , denoted by

$\tilde{f}_\mu(\mu_1, \dots, \mu_r)$, which is available from (8.3a.12) and the relationship between $d\tilde{S}$ and dD , is the following:

$$\begin{aligned} \tilde{f}_\mu(\mu_1, \dots, \mu_r)dD &= \frac{\pi^{r(r-1)}}{\tilde{\Gamma}_r(r)} \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p)\tilde{\Gamma}_r(n-p+r)} \\ &\times \left[\prod_{j=1}^r \mu_j^{p-r} \right] \left[\prod_{j=1}^r (1-\mu_j)^{n-p} \right] \left[\prod_{i<j} (\mu_i - \mu_j)^2 \right] dD. \end{aligned} \quad (8.3a.13)$$

8.3a.2. A direct method of evaluation in the complex case

The steps of the derivations being analogous to those utilized in the real case, the corresponding theorems will simply be stated for the complex case. Let $\tilde{W}_1 = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^*$ be a $p \times p$ singular Wishart matrix where the $p \times r$ matrix $\tilde{\mathbf{X}} = [\tilde{X}_1, \dots, \tilde{X}_r]$, $\tilde{X}_j \stackrel{iid}{\sim} \tilde{N}_p(O, I)$, $j = 1, \dots, r$, and $r < p$. That is, $\tilde{\mathbf{X}}$ is a simple random sample matrix from this complex Gaussian population. Let $\tilde{W}_2 > O$ have a complex Wishart distribution with n degrees of freedom and parameter matrix I , that is, $\tilde{W}_2 \sim \tilde{W}_p(n, I)$, $n \geq p$, and assume that \tilde{W}_1 and \tilde{W}_2 be independently distributed. Consider the initial equation

$$\det(\tilde{W}_1 - \lambda\tilde{W}_2) = 0 \Rightarrow \det(\tilde{W}_2^{-\frac{1}{2}}\tilde{W}_1\tilde{W}_2^{-\frac{1}{2}} - \lambda I) = 0, \quad (8.3a.14)$$

whose roots are $\lambda_1, \dots, \lambda_r, 0, \dots, 0$, and the additional equation $\det(\tilde{W}_1 - \mu(\tilde{W}_1 + \tilde{W}_2)) = 0$, whose roots will be denoted by $\mu_1, \dots, \mu_r, 0, \dots, 0$. In the following theorems, the λ_j 's and μ_j 's will refer to these two sets of roots.

Theorem 8.3a.7. Let $\tilde{W}_1 = \mathbf{X}\mathbf{X}^*$, $\mathbf{X} = [X_1, \dots, X_r]$, $X_j \stackrel{iid}{\sim} \tilde{N}_p(O, I)$, $j = 1, \dots, r$, and $r < p$. Let $\tilde{W}_2 \sim \tilde{W}_p(n, I)$, $n \geq p$, be a nonsingular complex Wishart matrix with n degrees of freedom and parameter matrix I . Further assume that \tilde{W}_1 and \tilde{W}_2 are independently distributed. Let $\tilde{V} = \tilde{W}_2^{-\frac{1}{2}}\tilde{\mathbf{X}}$, $\tilde{V}\tilde{V}^* \geq O$, $\tilde{V}^*\tilde{V} > O$, and $\tilde{S}_1 = \tilde{V}^*\tilde{V} > O$. Then, $\det(I_p + \tilde{V}\tilde{V}^*) = \det(I_r + \tilde{V}^*\tilde{V}) = \det(I_r + \tilde{S}_1)$, and the densities of \tilde{V}^* and \tilde{S}_1 are respectively given by

$$\tilde{f}_{10}(\tilde{V}^*)d\tilde{V}^* = \frac{\tilde{\Gamma}_r(p)}{\pi^{p(p-1)}} \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p)\tilde{\Gamma}_r(n+r-p)} |\det(I + \tilde{V}^*\tilde{V})|^{-(n+r)} d\tilde{V}^* \quad (8.3a.15)$$

and

$$\tilde{f}_{11}(\tilde{S}_1) = \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p)\tilde{\Gamma}_r(n+r-p)} |\det(\tilde{S}_1)|^{p-r} |\det(I + \tilde{S}_1)|^{-(n+r)}, \quad (8.3a.16)$$

which is a complex matrix-variate type-2 beta distribution with the parameters $(p, n - p + r)$. Additionally, the positive semi-definite matrix $\tilde{S}_2 = \tilde{W}_2^{-\frac{1}{2}} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^* \tilde{W}_2^{-\frac{1}{2}}$ is distributed, almost surely, as \tilde{S}_1 , which has a nonsingular complex matrix-variate type-2 beta distribution with the parameters $(p, n - p + r)$.

Theorem 8.3a.8. Let \tilde{W}_2 and $\tilde{\mathbf{X}}$ be as defined in Theorem 8.3a.7. Then, the joint density of the nonzero eigenvalues $\lambda_1, \dots, \lambda_r$ of $\tilde{S}_2 = \tilde{W}_2^{-\frac{1}{2}} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^* \tilde{W}_2^{-\frac{1}{2}}$, which are assumed to be distinct and such that $\lambda_1 > \dots > \lambda_r > 0$, is given by

$$\begin{aligned} \tilde{f}_{13}(\lambda_1, \dots, \lambda_r) dD &= \frac{\tilde{\Gamma}_r(n+r)}{\tilde{\Gamma}_r(p) \tilde{\Gamma}_r(n-p+r)} \frac{\pi^{r(r-1)}}{\tilde{\Gamma}_r(r)} \\ &\times \left[\prod_{j=1}^r \lambda_j^{p-r} \right] \left[\prod_{j=1}^r (1 + \lambda_j)^{-(n+r)} \right] \left[\prod_{i < j} (\lambda_i - \lambda_j)^2 \right] dD, \end{aligned} \tag{8.3a.17}$$

where $D = \text{diag}(\lambda_1, \dots, \lambda_r)$.

Theorem 8.3a.9. When $\mu_j = \frac{\lambda_j}{1+\lambda_j}$ or $\lambda_j = \frac{\mu_j}{1-\mu_j}$, the distributions of the μ_j 's and λ_j 's, as respectively defined in (8.3a.13) and (8.3a.17), coincide.

8.4. The Case of One Wishart or Gamma Matrix in the Real Domain

If we only consider a single $p \times p$ gamma matrix W with parameters (α, B) , $B > O$, $\Re(\alpha) > \frac{p-1}{2}$, whose density is

$$f(W) = \frac{|B|^\alpha}{\Gamma_p(\alpha)} |W|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(BW)}, \quad W > O, B > O, \Re(\alpha) > \frac{p-1}{2}, \tag{8.4.1}$$

and zero elsewhere, then it can readily be determined that $Z = B^{\frac{1}{2}} W$ has the density

$$f_1(Z) = \frac{1}{\Gamma_p(\alpha)} |Z|^{\alpha - \frac{p+1}{2}} e^{-\text{tr}(Z)}, \quad Z > O, \Re(\alpha) > \frac{p-1}{2}, \tag{8.4.2}$$

and zero elsewhere. When $\alpha = \frac{m}{2}$ and $B = \frac{1}{2} I$, Z has a Wishart density with $m \geq p$ degrees of freedom and parameter matrix I , its density being given by

$$f_2(Z) = \frac{1}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} |Z|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \text{tr}(Z)} \tag{8.4.3}$$

for $m \geq p$, and zero elsewhere. Since Z is symmetric, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that $Z = PDP'$ with $D = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $\lambda_j > 0$, $j = 1, \dots, p$, are the (assumed distinct) positive eigenvalues of Z , Z being positive definite. Consider the equation $ZQ_j = \lambda_j Q_j$ where the $p \times 1$ vector Q_j is an eigenvector corresponding to the eigenvalue λ_j . Since the eigenvalues are distinct, the eigenvectors are orthogonal to each other. Let Q_j be the normalized eigenvector, $Q_j'Q_j = 1$, $j = 1, \dots, p$, $Q_i'Q_j = 0$, for all $i \neq j$. Letting $Q = [Q_1, \dots, Q_p]$, this $p \times p$ matrix Q is such that

$$ZQ = QD \Rightarrow Z = QDQ' \Rightarrow Q = P, P = [P_1, \dots, P_p]$$

where P_1, \dots, P_p are the columns of the $p \times p$ matrix P . Yet, P need not be unique as $P_j'P_j = 1 \Rightarrow (-P_j)'(-P_j) = 1$. In order to make it unique, let us require that the first nonzero element of each of the vectors P_1, \dots, P_p be positive. Considering the transformation $Z = PDP'$, it follows from Theorem 8.2.1 that, before integrating over the full orthogonal group,

$$dZ = \left\{ \prod_{i < j} (\lambda_i - \lambda_j) \right\} dD h(P) \tag{8.4.4}$$

where $h(P)$ is a differential element associated with the unique matrix of eigenvectors P , as is explained in Mathai (1997). The integral over $h(P)$ gives

$$\int_{O_p} h(P) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \tag{8.4.5}$$

where O_p is the full orthogonal group of $p \times p$ orthonormal matrices. On observing that $|Z| = \prod_{j=1}^p \lambda_j$ and $\text{tr}(Z) = \lambda_1 + \dots + \lambda_p$, it follows from (8.4.3) and (8.4.4) that the joint density of the eigenvalues $\lambda_1, \dots, \lambda_p$ and the matrix of eigenvectors P can be expressed as

$$f_3(D, P) dD \wedge dP = \frac{[\prod_{j=1}^p \lambda_j]^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \sum_{j=1}^p \lambda_j} [\prod_{i < j} (\lambda_i - \lambda_j)]}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} dD h(P). \tag{8.4.6}$$

Thus, the marginal density of $\lambda_1, \dots, \lambda_p$ can be obtained by integrating out P . Denoting this marginal density by $f_4(\lambda_1, \dots, \lambda_p)$, we have

$$f_4(\lambda_1, \dots, \lambda_p) dD = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \frac{[\prod_{j=1}^p \lambda_j]^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2} \sum_{j=1}^p \lambda_j}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \left[\prod_{i < j} |\lambda_i - \lambda_j| \right] dD, \tag{8.4.7}$$

and zero elsewhere. The density of P is then the remainder of the joint density. Denoting it by $f_5(P)$, we have the following:

$$f_5(P) dP = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} h(P), \quad PP' = I, \quad (8.4.8)$$

where $P = [P_1, \dots, P_p]$, the first nonzero element of P_j being positive for $j = 1, \dots, p$, so as to make P unique. Hence, the following result:

Theorem 8.4.1. *Let the $p \times p$ real positive definite matrix Z have a Wishart density with $m \geq p$ degrees of freedom and parameter matrix I . Let $\lambda_1, \dots, \lambda_p$ be the distinct positive eigenvalues of Z in decreasing order and the $p \times p$ orthonormal matrix P be the matrix of normalized eigenvectors corresponding to the λ_j 's. Then, $\{\lambda_1, \dots, \lambda_p\}$ and P are independently distributed, with the densities of $\{\lambda_1, \dots, \lambda_p\}$ and P being respectively given in (8.4.7) and (8.4.8).*

8.4a. The Case of One Wishart or Gamma Matrix, Complex Domain

Let \tilde{W} be a $p \times p$ complex gamma distributed matrix, $\tilde{W} = \tilde{W}^* > O$, whose density is

$$\tilde{f}(\tilde{W}) = \frac{|\det(\tilde{B})|^\alpha}{\tilde{\Gamma}_p(\alpha)} |\det(\tilde{W})|^{\alpha-p} e^{-\text{tr}(\tilde{B}\tilde{W})}, \quad \tilde{W} > O, \quad \tilde{B} > O, \quad \Re(\alpha) > p - 1. \quad (8.4a.1)$$

Letting $\tilde{Z} = \tilde{B}^{\frac{1}{2}} \tilde{W}$, \tilde{Z} has the density

$$\tilde{f}_1(\tilde{Z}) = \frac{1}{\tilde{\Gamma}_p(\alpha)} |\det(\tilde{Z})|^{\alpha-p} e^{-\text{tr}(\tilde{Z})}, \quad \tilde{Z} > O, \quad \Re(\alpha) > p - 1. \quad (8.4a.2)$$

If $\alpha = m$, $m = p, p + 1, \dots$ in (8.4a.2), then we have the following Wishart density having m degrees of freedom in the complex domain:

$$\tilde{f}_2(\tilde{Z}) = \frac{1}{\tilde{\Gamma}_p(m)} |\det(\tilde{Z})|^{m-p} e^{-\text{tr}(\tilde{Z})}. \quad (8.4a.3)$$

Consider a unique unitary matrix \tilde{P} , $\tilde{P}\tilde{P}^* = I$, $\tilde{P}^*\tilde{P} = I$ such that $\tilde{P}^*\tilde{Z}\tilde{P} = \text{diag}(\lambda_1, \dots, \lambda_p)$ where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of \tilde{Z} , which are real and positive since \tilde{Z} is Hermitian positive definite. Letting the eigenvalues be distinct and such that $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$, observe that

$$d\tilde{Z} = \left\{ \prod_{i < j} |\lambda_i - \lambda_j|^2 \right\} dD \tilde{h}(\tilde{P}) \quad (8.4a.4)$$

where $\tilde{h}(\tilde{P})$ is the differential element corresponding to the unique unitary matrix \tilde{P} . Then, as established in Mathai (1997),

$$\int_{\tilde{O}_p} \tilde{h}(\tilde{P}) = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \tag{8.4a.5}$$

where \tilde{O}_p is the full unitary group. Thus, the joint density of the eigenvalues $\lambda_1, \dots, \lambda_p$ and their associated normalized eigenvectors, denoted by $\tilde{f}_3(D, \tilde{P})$, is

$$\tilde{f}_3(D, \tilde{P}) dD \wedge d\tilde{P} = \frac{[\prod_{j=1}^p \lambda_j]^{m-p} e^{-\sum_{j=1}^p \lambda_j}}{\tilde{\Gamma}_p(m)} \left[\prod_{i < j} |\lambda_i - \lambda_j|^2 \right] dD \tilde{h}(\tilde{P}). \tag{8.4a.6}$$

Then, integrating out \tilde{P} with the help of (8.4a.5), the marginal density of the eigenvalues $\lambda_1 > \dots > \lambda_p > 0$, denoted by $\tilde{f}_4(\lambda_1, \dots, \lambda_p)$, is the following:

$$\tilde{f}_4(\lambda_1, \dots, \lambda_p) dD = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \frac{[\prod_{j=1}^p \lambda_j]^{m-p} e^{-\sum_{j=1}^p \lambda_j}}{\tilde{\Gamma}_p(m)} \left[\prod_{i < j} |\lambda_i - \lambda_j|^2 \right] dD. \tag{8.4a.7}$$

Thus, the joint density of the normalized eigenvectors forming \tilde{P} , denoted by $\tilde{f}_5(\tilde{P})$, is given by

$$\tilde{f}_5(\tilde{P}) d\tilde{P} = \frac{\tilde{\Gamma}_p(p)}{\pi^{p(p-1)}} \tilde{h}(\tilde{P}). \tag{8.4a.8}$$

These results are summarized in the following theorem.

Theorem 8.4a.1. *Let \tilde{Z} have the density appearing in (8.4a.3). Then, the joint density of the distinct eigenvalues $\lambda_1 > \dots > \lambda_p > 0$ of \tilde{Z} is as given in (8.4a.7) and the joint density of the associated normalized eigenvectors comprising the unitary matrix P is as specified in (8.4a.8).*

References

T.W. Anderson (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, Wiley Interscience, New York.

A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 9

Principal Component Analysis



9.1. Introduction

We will adopt the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

The requisite theory for the study of *Principal Component Analysis* has already been introduced in Chap. 1, namely, the problem of optimizing a real quadratic form that is subject to a constraint. We shall formulate the problem with respect to a practical situation consisting of selecting the most “relevant” variables in a study. Suppose that a scientist would like to devise a “good health” index in terms of certain indicators. After selecting a random sample of individuals belonging to a population that is homogeneous with respect to a variety of factors, such as age group, racial background and environmental conditions, she managed to secure measurements on $p = 15$ variables, including for instance, x_1 : weight, x_2 : systolic pressure, x_3 : blood sugar level, and x_4 : height. She now

faces a quandary as there is an excessive number of variables and some of them may not be relevant to her investigation. A methodology is thus required for discarding the unimportant ones. As a result, the number of variables will be reduced and the interpretation of the results, possibly facilitated. So, what might be the most pertinent variables in any such study? If all the observations of a particular variable x_j are concentrated around a certain value, μ_j , then that variable is more or less predetermined. As an example, suppose that the height of the individuals comprising a study group is neighboring 1.8 meters and that, on the other hand, it is observed that the weight measurements are comparatively spread out. On account of this, while height is not a particularly consequential variable in connection with this study, weight is. Accordingly, we can utilize the criterion: *the larger the variance of a variable, the more relevant this variable is*. Let the $p \times 1$ vector X , $X' = (x_1, \dots, x_p)$, encompass all the variables on which measurements are available. Let the covariance matrix associated with X be Σ , that is, $\text{Cov}(X) = \Sigma$. Since linear functions also contain individual variables, we may consider linear functions such as $u = a_1x_1 + \dots + a_px_p = A'X = X'A$, a prime designating a transpose, where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \quad \text{and} \quad \Sigma = (\sigma_{ij}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}. \quad (9.1.1)$$

Then,

$$\text{Var}(u) = \text{Var}(A'X) = \text{Var}(X'A) = A'\Sigma A. \quad (9.1.2)$$

9.2. Principal Components

As will be explained further, the central objective in connection with the derivation of principal components consists of maximizing $A'\Sigma A$. Such an exercise would indeed prove meaningless unless some constraint is imposed on A , considering that, for an arbitrary vector A , the minimum of $A'\Sigma A$ occurs at zero and the maximum, at $+\infty$, $\Sigma = E[X - E(X)][X - E(X)]'$ being either positive definite or positive semi-definite. Since Σ is symmetric and non-negative definite, its eigenvalues, denoted by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, are real. Moreover, Σ being symmetric, there exists an orthonormal matrix P , $PP' = I$, $P'P = I$, such that

$$P'\Sigma P = \text{diag}(\lambda_1, \dots, \lambda_p) \equiv \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} \quad (9.2.1)$$

and

$$\Sigma = P \Lambda P' = \lambda_1 P_1 P_1' + \cdots + \lambda_p P_p P_p', \quad (9.2.2)$$

where P_1, \dots, P_p constitute the columns of P , P_i denoting a normalized eigenvector corresponding to λ_i , $i = 1, \dots, p$; this is expounded for instance in Mathai and Haubold (2017a). Note that all real symmetric matrices, including those having repeated eigenvalues, can be diagonalized. Since the optimization problem is pointless when A is arbitrary, the search for an optimum shall be confined to vectors A such that $A'A = 1$, that is, vectors lying on the unit sphere in \Re^p . Without any loss of generality, the coefficients of the linear function can be selected so that the Euclidean norm of the coefficient vector is unity, in which case a minimum and a maximum will both exist. Hence, the problem can be restated as follows:

$$\text{Maximize } A' \Sigma A \text{ subject to } A'A = 1. \quad (i)$$

We will resort to the method of Lagrangian multipliers to optimize $A' \Sigma A$ subject to the constraint $A'A = 1$. Let

$$\phi_1 = A' \Sigma A - \lambda(A'A - 1) \quad (ii)$$

where λ is a Lagrangian multiplier. Differentiating ϕ_1 with respect to A and equating the result to a null vector (vector/matrix derivatives are discussed in Chap. 1, as well as in Mathai 1997), we have the following:

$$\frac{\partial \phi_1}{\partial A} = O \Rightarrow 2 \Sigma A - 2 \lambda A = O \Rightarrow \Sigma A = \lambda A. \quad (iii)$$

On premultiplying (iii) by A' , we have

$$A' \Sigma A = \lambda A'A = \lambda. \quad (9.2.3)$$

In order to obtain a non-null solution for A in (iii), the coefficient matrix $\Sigma - \lambda I$ has to be singular or, equivalently, its determinant has to be zero, that is, $|\Sigma - \lambda I| = 0$, which implies that λ is an eigenvalue of Σ , A being the corresponding eigenvector. Thus, it follows from (9.2.3) that the maximum of the quadratic form $A' \Sigma A$, subject to $A'A = 1$, is the largest eigenvalue of Σ :

$$\max_{A'A=1} [A' \Sigma A] = \lambda_1 = \text{the largest eigenvalue of } \Sigma.$$

Similarly,

$$\min_{A'A=1} [A' \Sigma A] = \lambda_p = \text{the smallest eigenvalue of } \Sigma. \quad (9.2.4)$$

Since $\text{Var}(A'X) = A'\Sigma A = \lambda$, the largest variance associated with a linear combination $A'X$ wherein the vector A is normalized, is equal to λ_1 , the largest eigenvalue of Σ , and letting A_1 be the normalized ($A_1'A_1 = 1$) eigenvector corresponding λ_1 , $u_1 = A_1'X$ will be that linear combination of X having the maximum variance. Thus, u_1 is called the first principal component which is the linear function of X having the maximum variance. Although normalized, the vector A_1 is not unique, as $(-A_1)'(-A_1)$ is also equal to one. In order to ensure the unicity, we will require that the first nonzero element of A_1 —and the other $p - 1$ normalized eigenvectors—be positive. Recall that since Σ is a real symmetric matrix, the λ_j 's are real and so are the corresponding eigenvectors. Consider the second largest eigenvalue λ_2 and determine the associated normalized eigenvector A_2 ; then $u_2 = A_2'X$ will be the second principal component. Since the matrix P that diagonalizes Σ into the diagonal matrix of its eigenvalues is orthonormal, the normalized eigenvectors A_1, A_2, \dots, A_p are necessarily orthogonal to each other, which means that the corresponding principal components $u_1 = A_1'X, u_2 = A_2'X, \dots, u_p = A_p'X$ will be uncorrelated. Let us see whether uncorrelated normalized eigenvectors could be obtained by making use of the above procedure. When constructing A_2 , we can impose an additional condition to the effect that $A'X$ should be uncorrelated with $A_1'X$, $A_1'\Sigma A_1$ being equal to λ_1 , the largest eigenvalue of Σ . The covariance between $A'X$ and $A_1'X$ is

$$\text{Cov}(A'X, A_1'X) = A'\text{Cov}(X)A_1 = A'\Sigma A_1 = A_1'\Sigma A. \quad (9.2.5)$$

Hence, we may require that $A'\Sigma A_1 = A_1'\Sigma A = 0$. However,

$$0 = A'\Sigma A_1 = A'(\Sigma A_1) = A'\lambda_1 A_1 = \lambda_1 A'A_1 \Rightarrow A'A_1 = 0. \quad (iv)$$

Observe that $\lambda_1 > 0$, noting that Σ would be a null matrix if its largest eigenvalue were equal to zero, in which case no optimization problem would remain to be solved. Consider

$$\phi_2 = A'\Sigma A - 2\mu_1(A'\Sigma A_1 - 0) - \mu_2(A'A - 1) \quad (v)$$

where μ_1 and μ_2 are the Lagrangian multipliers. Now, differentiating ϕ_2 with respect to A and equating the result to a null vector, we have the following:

$$\frac{\partial \phi_2}{\partial A} = 0 \Rightarrow 2\Sigma A - 2\mu_1(\Sigma A_1) - 2\mu_2 A = 0. \quad (vi)$$

Premultiplying (vi) by A_1' yields

$$A_1'\Sigma A - \mu_1 A_1'\Sigma A_1 - \mu_2 A_1'A = 0 \Rightarrow 0 - \mu_1 \lambda_1 - 0 \Rightarrow \mu_1 = 0, \quad (vii)$$

which entails that the added condition of uncorrelatedness with $u_1 = A'_1 X$ is superfluous. When determining A_j , we could require that $A'X$ be uncorrelated with the principal components $u_1 = A'_1 X, \dots, u_{j-1} = A'_{j-1} X$; however, as it turns out, these conditions become redundant when optimizing $A'\Sigma A$ subject to $A'A = 1$. Thus, after determining the normalized eigenvector A_j (whose first nonzero element is positive) corresponding the j -th largest eigenvalue of Σ , we form the j -th principal component, $u_j = A'_j X$, which will necessarily be uncorrelated with the preceding principal components. The question that arises at this juncture is whether all of u_1, \dots, u_p are needed or a subset thereof would suffice? For instance, we could interrupt the computations when the variance of the principal component $u_j = A'_j X$, namely λ_j , falls below a predetermined threshold, in which case we can regard the remaining principal components, u_{j+1}, \dots, u_p , as unimportant and omit the associated calculations. In this way, a reduction in the number of variables is achieved as the original number of variables p is reduced to $j < p$ principal components. However, this reduction in the number of variables could be viewed as a compromise since the new variables are linear functions of all the original ones and so, may not be as interpretable in a real-life situation. Other drawbacks will be considered in the next section. Observe that since $\text{Var}(u_j) = \lambda_j$, $j = 1, \dots, p$, the fraction

$$v_1 = \frac{\lambda_1}{\sum_{j=1}^p \lambda_j} = \text{the proportion of the total variation accounted for by } u_1, \quad (9.2.6)$$

and letting $r < p$,

$$v_r = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^p \lambda_j} = \text{the proportion of total variation accounted for by } u_1, \dots, u_r. \quad (9.2.7)$$

If $v_1 = 0.7$, v_1 accounts for 70% of the total variation in the original variables or 70% of the total variation is due to the first principal component. If $r = 3$ and $v_3 = 0.99$, then the sum of the first three principal components accounts for 99% of the total variation. We can also use this percentage of the total variation as a stopping rule for the determination of the principal components. For example when v_r of (9.2.7) is say, greater than or equal to 95%, we may interrupt the determination of the principal components beyond u_r .

Example 9.2.1. Even though reducing the number of variables is a main objective of Principal Component Analysis, for illustrative purposes, we will consider a case involving three variables, that is, $p = 3$. Compute the principal components associated with the following covariance matrix:

$$V = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix}.$$

Solution 9.2.1. Let us verify that V is positive definite. The leading minors of $V = V'$ being

$$\begin{aligned} |(3)| &= 3 > 0, \quad \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 3^2 - (-1)^2 = 8 > 0, \quad \begin{vmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{vmatrix} \\ &= 3[3^2 - (-1)^2] - (-1)[(-1)(3) - 0] = 21 > 0, \end{aligned}$$

$V > O$. Let us compute the eigenvalues of V . Consider the equation $|V - \lambda I| = 0 \Rightarrow (3 - \lambda)[(3 - \lambda)^2 - 1] - (3 - \lambda) = 0 \Rightarrow (3 - \lambda)[(3 - \lambda)^2 - 2] = 0 \Rightarrow (3 - \lambda)(3 - \lambda \pm \sqrt{2}) = 0$. Hence the eigenvalues are $\lambda_1 = 3 + \sqrt{2}$, $\lambda_2 = 3$, $\lambda_3 = 3 - \sqrt{2}$. Let us compute an eigenvector corresponding to $\lambda_1 = 3 + \sqrt{2}$. Consider $(V - \lambda_1 I)X = O$, that is,

$$\begin{bmatrix} 3 - \lambda_1 & -1 & 0 \\ -1 & 3 - \lambda_1 & 1 \\ 0 & 1 & 3 - \lambda_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (i)$$

There are three linear equations involving the x_j 's in (i). Since the matrix $V - \lambda I$ is singular, we need only consider any two of these three linear equations and solve to obtain a solution. Let the first equation be $-\sqrt{2}x_1 - x_2 = 0 \Rightarrow x_2 = -\sqrt{2}x_1$, and the second one be $-x_1 - \sqrt{2}x_2 + x_3 = 0 \Rightarrow -x_1 + 2x_1 + x_3 = 0 \Rightarrow x_3 = -x_1$. Now, one solution is

$$X_1 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix}, \text{ which once normalized is } A_1 = \frac{1}{2} \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix}.$$

Observe that X_1 also satisfies the third equation in (i) and that

$$\begin{aligned} A_1' V A_1 &= \frac{1}{4} [1, -\sqrt{2}, -1] \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -\sqrt{2} \\ -1 \end{bmatrix} \\ &= \frac{1}{4} [3(1)^2 + 3(-\sqrt{2})^2 + 3(-1)^2 - 2(1)(-\sqrt{2}) + 2(-\sqrt{2})(-1)] \\ &= 3 + \sqrt{2} = \lambda_1. \end{aligned}$$

Thus, the first principal component, denoted by u_1 , is

$$u_1 = A_1' X = \frac{1}{2} [x_1 - \sqrt{2}x_2 - x_3]. \quad (ii)$$

Now, consider an eigenvector corresponding to $\lambda_2 = 3$. $(V - \lambda_2 I)X = O$ provides the first equation: $-x_2 = 0 \Rightarrow x_2 = 0$, the third equation giving $-x_1 + x_3 = 0 \Rightarrow x_1 = x_3$. Therefore, one solution for X is

$$X_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ whose normalized form is } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

so that the second principal component is

$$u_2 = \frac{1}{\sqrt{2}}[x_1 + x_3]. \quad (iii)$$

It can readily be checked that $A_2' V A_2 = 3 = \lambda_2$. Let us now obtain an eigenvector corresponding to the eigenvalue $\lambda_3 = 3 - \sqrt{2}$. Consider the linear system $(V - \lambda_3)X = O$. The first equation is $\sqrt{2}x_1 - x_2 = 0 \Rightarrow x_2 = \sqrt{2}x_1$, and the third one gives $x_2 + \sqrt{2}x_3 = 0 \Rightarrow x_2 = -\sqrt{2}x_3$. One solution is

$$X_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ -1 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{2} \begin{bmatrix} 1 \\ \sqrt{2} \\ -1 \end{bmatrix}.$$

The third principal component is then

$$u_3 = \frac{1}{2}[x_1 + \sqrt{2}x_2 - x_3]. \quad (iv)$$

It is easily verified that $A_3' V A_3 = \lambda_3 = 3 - \sqrt{2}$. As well,

$$\begin{aligned} \text{Var}(u_1) &= \frac{1}{4}[\text{Var}(x_1) + 2\text{Var}(x_2) + \text{Var}(x_3) - 2\sqrt{2}\text{Cov}(x_1, x_2) \\ &\quad - 2\text{Cov}(x_1, x_3) + 2\sqrt{2}\text{Cov}(x_2, x_3)] \\ &= \frac{1}{4}[3 + 2 \times 3 + 3 + 2\sqrt{2}(-1) - 2(0) + 2\sqrt{2}] \\ &= \frac{1}{4}[3 \times 4 + 4\sqrt{2}] = 3 + \sqrt{2} = \lambda_1. \end{aligned}$$

Similar calculations will confirm that $\text{Var}(u_2) = 3 = \lambda_2$ and $\text{Var}(u_3) = 2 - \sqrt{2} = \lambda_3$. Now, consider the covariance between u_1 and u_2 :

$$\begin{aligned} \text{Cov}(u_1, u_2) &= \frac{1}{\sqrt{2}}[\text{Var}(x_1) + \text{Cov}(x_1, x_3) - \sqrt{2}\text{Cov}(x_1, x_2) - \sqrt{2}\text{Cov}(x_2, x_3) \\ &\quad - \text{Cov}(x_3, x_1) - \text{Var}(x_3)] \\ &= \frac{1}{4}[3 + 0 + \sqrt{2} - \sqrt{2} - 0 - 0 - 3] = 0. \end{aligned}$$

It can be likewise verified that $\text{Cov}(u_1, u_3) = 0$ and $\text{Cov}(u_2, u_3) = 0$. Note that u_1 , u_2 , and u_3 respectively account for 49.1%, 33.3% and 17.6% of the total variation. As none of these proportions is negligibly small, all three principal components are deemed relevant and, in this instance, it is not indicated to reduce the number of the original variables. Although we still end up with as many variables, the u_j 's, $j = 1, 2, 3$, are uncorrelated, which was not the case for the original variables.

9.3. Issues to Be Mindful of when Constructing Principal Components

Since variances and covariances are expressed in units of measurement, principal components also depend upon the scale on which measurements on the individual variables are made. If we change the units of measurement, the principal components will differ. Suppose that x_j is multiplied by a real scalar constant d_j , $j = 1, \dots, p$, where some of the d_j 's are not equal to 1. This is equivalent to changing the units of measurement of some of the variables. Let

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, Y = \begin{bmatrix} d_1 x_1 \\ \vdots \\ d_p x_p \end{bmatrix} = DX, D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}, \quad (9.3.1)$$

and consider the linear functions $A'X$ and $A'Y$. Then,

$$\text{Var}(A'X) = A'\Sigma A, \quad \text{Var}(A'Y) = A'\text{Var}(DX)A = A'D\Sigma DA. \quad (9.3.2)$$

Since the eigenvalues of Σ and $D\Sigma D$ differ, so will the corresponding principal components, and if the original variables are measured in various units of measurements, it would be advisable to attempt to standardize them. Letting R denote the correlation matrix which is scale-invariant, observe that the covariance matrix $\Sigma = \Sigma_1 R \Sigma_1$ where $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_p)$, σ_j^2 being the variance of x_j , $j = 1, \dots, p$. Thus, if the original variables are scaled by the inverses of their associated standard deviations, that is, $\frac{x_j}{\sigma_j}$ or equivalently, via the transformation $\Sigma_1^{-1}X$, the resulting covariance matrix is $\text{Cov}(\Sigma_1^{-1}X) = R$, the correlation matrix. Accordingly, constructing the principal components by making use of R instead of Σ , will mitigate the issue stemming from the scale of measurement.

If $\lambda_1, \dots, \lambda_p$ are the eigenvalues of Σ , then $\lambda_1^k, \dots, \lambda_p^k$ will be the eigenvalues of Σ^k . Moreover, λ and λ^k will share the same eigenvectors. Note that the collection $(\lambda_1^k, \dots, \lambda_p^k)$ will be well separated compared to the set $(\lambda_1, \dots, \lambda_p)$ when the λ_i 's are distinct and greater than one. Hence, in some instances, it might be preferable to construct principal

components by making use of Σ^k for an appropriate value of k instead of Σ . Observe that, in certain situations, it is not feasible to provide a physical interpretation of the principal components which are a linear function of the original x_1, \dots, x_p . Nonetheless, they can at times be informative by pointing to the average of certain variables (for example, u_2 in the previous numerical example) or by eliciting contrasts between two sets of variables (for example, u_3 in the previous numerical example, which opposes x_3 to x_1 and x_2).

To illustrate how the eigenvalues of a correlation matrix are determined, we revisit Example 9.2.1 wherein the variances of the x_j 's or the diagonal elements of the covariance matrix are all equal to 3. Thus, in this case, the correlation matrix is $R = \frac{1}{3}V$, and the eigenvalues of R will be $\frac{1}{3}$ times the eigenvalues of V . However, the normalized eigenvectors will remain identical to those of V , and therefore the principal components will not change. We now consider an example wherein the diagonal elements of the covariance matrix are different.

Example 9.3.1. Let $X' = (x_1, x_2, x_3)$ where the x_j 's are real scalar random variables. Compute the principal components resulting from R , the correlation matrix of X , where

$$R = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{6}} \\ 0 & 1 & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 1 \end{bmatrix}.$$

Solution 9.3.1. Let us compute the eigenvalues of R . Consider the equation

$$\begin{aligned} |R - \lambda I| = 0 &\Rightarrow (1 - \lambda) \left[(1 - \lambda)^2 - \frac{1}{6} \right] - \frac{1}{\sqrt{6}} \left[-\frac{1}{\sqrt{6}}(1 - \lambda) \right] = 0 \\ &\Rightarrow (1 - \lambda) \left[(1 - \lambda)^2 - \frac{1}{3} \right] = 0. \end{aligned}$$

Thus, the eigenvalues are $\lambda_1 = 1 + \frac{1}{\sqrt{3}}$, $\lambda_2 = 1$, $\lambda_3 = 1 - \frac{1}{\sqrt{3}}$. Let us compute an eigenvector corresponding to $\lambda_1 = 1 + \frac{1}{\sqrt{3}}$. Consider the system $(R - \lambda_1 I)X = O$. The first equation is

$$-\frac{1}{\sqrt{3}}x_1 + \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_1 = \frac{1}{\sqrt{2}}x_3,$$

the second equation being

$$-\frac{1}{\sqrt{3}}x_2 - \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_2 = -\frac{1}{\sqrt{2}}x_3.$$

Let us take $x_3 = \sqrt{2}$. Then, an eigenvector corresponding to λ_1 is

$$X_1 = \begin{bmatrix} 1 \\ -1 \\ \sqrt{2} \end{bmatrix}, \text{ and its normalized form is } A_1 = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ \sqrt{2} \end{bmatrix}.$$

Let us verify that the third equation is also satisfied by X_1 or A_1 . This is the case since $\frac{1}{\sqrt{6}} + \frac{1}{\sqrt{6}} - \frac{\sqrt{2}}{\sqrt{3}} = 0$, and the first principal component is indeed $u_1 = \frac{1}{2}[x_1 - x_2 + \sqrt{2}x_3]$. As well, the variance of u_1 is equal to λ_1 :

$$\begin{aligned} \text{Var}(u_1) &= \frac{1}{4}[\text{Var}(x_1) + \text{Var}(x_2) + 2\text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad + 2\sqrt{2}\text{Cov}(x_1, x_3) - 2\sqrt{2}\text{Cov}(x_2, x_3)] \\ &= \frac{1}{4}\left[1 + 1 + 2 + \frac{2\sqrt{2}}{\sqrt{6}} + \frac{2\sqrt{2}}{\sqrt{6}}\right] = 1 + \frac{1}{\sqrt{3}} = \lambda_1. \end{aligned}$$

Let us compute an eigenvector corresponding to the eigenvalue $\lambda_2 = 1$. Consider the linear system $(R - \lambda_2 I)X = O$. The first equation, $\frac{1}{\sqrt{6}}x_3 = 0$, gives $x_3 = 0$ and the second one also yields $x_3 = 0$; as for the third one, $\frac{x_1}{\sqrt{6}} - \frac{x_2}{\sqrt{6}} = 0 \Rightarrow x_1 = x_2$. Letting $x_1 = 1$, an eigenvector is

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ its normalized form being given by } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Thus, the second principal component is $u_2 = \frac{1}{\sqrt{2}}[x_1 + x_2]$. In the case of λ_3 , we consider the system $(R - \lambda_3 I)X = O$ whose first equation is

$$\frac{1}{\sqrt{3}}x_1 + \frac{1}{\sqrt{6}}x_3 = 0 \Rightarrow x_1 = -\frac{1}{\sqrt{2}}x_3,$$

the second one being

$$\frac{x_2}{\sqrt{3}} + \frac{x_3}{\sqrt{6}} = 0 \Rightarrow x_2 = \frac{1}{\sqrt{2}}x_3.$$

Letting $x_3 = \sqrt{2}$, an eigenvector corresponding to λ_3 is

$$X_3 = \begin{bmatrix} -1 \\ 1 \\ \sqrt{2} \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ \sqrt{2} \end{bmatrix}.$$

For the matrix $[A_1, A_2, A_3]$ to be uniquely determined, we may multiply A_3 by -1 so that the first nonzero element is positive. Hence, the third principal component is $u_3 = \frac{1}{2}[x_1 - x_2 - \sqrt{2}x_3]$. As was shown in the case of u_1 , it can also be verified that $\text{Var}(u_2) = 1 = \lambda_2$, $\text{Var}(u_3) = \lambda_3 = 1 - \frac{1}{\sqrt{3}}$, $\text{Cov}(u_1, u_2) = 0$, $\text{Cov}(u_1, u_3) = 0$ and $\text{Cov}(u_2, u_3) = 0$. We note that

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{1 + 1/\sqrt{3}}{3} \approx 0.525,$$

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{2 + 1/\sqrt{3}}{3} \approx 0.859.$$

Thus, almost 53% of the total variation is accounted for by the first principal component and nearly 86% of the total variation is due to the first two principal components.

9.4. The Vector of Principal Components

Observe that the determinant of a matrix is the product of its eigenvalues and that its trace is the sum of its eigenvalues, that is, $|\Sigma| = \lambda_1 \cdots \lambda_p$ and $\text{tr}(\Sigma) = \lambda_1 + \cdots + \lambda_p$. As previously pointed out, the determinant of a covariance matrix corresponds to Wilks' concept of *generalized variance*. Let us consider the vector of principal components. The principal components are $u_j = A'_j X$, with $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j$, $j = 1, \dots, p$, and $\text{Cov}(u_i, u_j) = 0$ for all $i \neq j$. Thus,

$$U = \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} = \begin{bmatrix} A'_1 X \\ \vdots \\ A'_p X \end{bmatrix} \Rightarrow \text{Cov}(U) = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}. \quad (9.4.1)$$

The determinant of the covariance matrix is the product $\lambda_1 \cdots \lambda_p$ and its trace, the sum $\lambda_1 + \cdots + \lambda_p$. Hence the following result:

Theorem 9.4.1. *Let X be a $p \times 1$ real vector whose associated covariance matrix is Σ . Let the principal components of X be denoted by $u_j = A'_j X$ with $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j = j$ -th largest eigenvalue of Σ , and $U' = (u_1, \dots, u_p)$ with $\text{Cov}(U) \equiv \Sigma_u$. Then, $|\Sigma_u| = |\Sigma| =$ product of the eigenvalues, $\lambda_1 \cdots \lambda_p$, and $\text{tr}(\Sigma_u) = \text{tr}(\Sigma) =$ sum of the eigenvalues, $\lambda_1 + \cdots + \lambda_p$. Observe that the determinant as well as the eigenvalues and the trace are invariant with respect to orthonormal transformations or rotations of the coordinate axes.*

Let $A = [A_1, A_2, \dots, A_p]$. Then

$$\Sigma A = A\Lambda, \quad A'A = I, \quad A'\Sigma A = \Lambda. \quad (9.4.2)$$

Note that $U' = (u_1, \dots, u_p)$ where u_1, \dots, u_p are linearly independent. We have not assumed any distribution for the $p \times 1$ vector X so far. If the $p \times 1$ vector X has a p -variate nonsingular Gaussian distribution, that is, $X \sim N_p(O, \Sigma)$, $\Sigma > O$, then

$$u_j \sim N_1(0, \lambda_j), \quad U \sim N_p(O, \Lambda). \quad (9.4.3)$$

Since *Principal Components* involve variances or covariances, which are free of any location parameter, we may take the mean value vector to be a null vector without any loss of generality. Then, $E[u_j] = 0$ by assumption, $\text{Var}(u_j) = A'_j \Sigma A_j = \lambda_j$, $j = 1, \dots, p$ and $\text{Cov}(U) = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$. Accordingly, $\frac{\lambda_1}{\lambda_1 + \dots + \lambda_p}$ is the proportion of the total variance accounted for by the largest eigenvalue λ_1 , where λ_1 is equal to the variance of the first principal component. Similarly, $\frac{\lambda_1 + \dots + \lambda_r}{\lambda_1 + \dots + \lambda_p}$ is the proportion of the total variance due to the first r principal components, $r \leq p$.

Example 9.4.1. Let $X \sim N_3(\mu, \Sigma)$, $\Sigma > O$, where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix}.$$

Derive the densities of the principal components of X .

Solution 9.4.1. Let us determine the eigenvalues of Σ . Consider the equation

$$\begin{aligned} |\Sigma - \lambda I| = 0 &\Rightarrow \begin{vmatrix} 2 - \lambda & 0 & 1 \\ 0 & 2 - \lambda & -1 \\ 1 & -1 & 3 - \lambda \end{vmatrix} = 0 \\ &\Rightarrow (2 - \lambda)[(2 - \lambda)(3 - \lambda) - 1] - (2 - \lambda) = 0 \\ &\Rightarrow (2 - \lambda)[\lambda^2 - 5\lambda + 4] = 0 \Rightarrow \lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 1. \end{aligned}$$

Let us compute an eigenvector corresponding to $\lambda_1 = 4$. Consider the linear system $(\Sigma - \lambda_1 I)X = O$, whose first equation gives $-2x_1 + x_3 = 0$ or $x_1 = \frac{1}{2}x_3$, the second equation, $-2x_2 - x_3 = 0$, yielding $x_2 = -\frac{1}{2}x_3$. Letting $x_3 = 2$, one solution is

$$X_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \quad \text{and its normalized form is } A_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Thus, the first principal component is $u_1 = A'_1 X = \frac{1}{\sqrt{6}}[x_1 - x_2 + 2x_3]$ with $E[u_1] = \frac{1}{\sqrt{6}}[1 - 0 - 2] = -\frac{1}{\sqrt{6}}$ and

$$\begin{aligned}\text{Var}(u_1) &= \frac{1}{6}[\text{Var}(x_1) + \text{Var}(x_2) + 4\text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad + 4\text{Cov}(x_1, x_3) - 4\text{Cov}(x_2, x_3)] \\ &= \frac{1}{6}[2 + 2 + 4 \times 3 - 2(0) + 4(1) - 4(-1)] = 4 = \lambda_1.\end{aligned}$$

Since u_1 is a linear function of normal variables, u_1 has the following Gaussian distribution:

$$u_1 \sim N_1\left(-\frac{1}{\sqrt{6}}, 4\right).$$

Let us compute an eigenvector corresponding to $\lambda_2 = 2$. Consider the system $(\Sigma - \lambda_2 I)X = O$, whose first and second equations give $x_3 = 0$, the third equation $x_1 - x_2 + x_3 = 0$ yielding $x_1 = x_2$. Hence, one solution is

$$X_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ its normalized form being } A_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

and the second principal component is $u_2 = A'_2 X = \frac{1}{\sqrt{2}}[x_1 + x_2]$ with $E[u_2] = \frac{1}{\sqrt{2}}[1 + 0] = \frac{1}{\sqrt{2}}$. Let us verify that the variance of u_2 is $\lambda_2 = 2$:

$$\text{Var}(u_2) = \frac{1}{2}[\text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2)] = \frac{1}{2}[2 + 2 + 2(0)] = 2 = \lambda_2.$$

Hence, u_2 has the following real univariate normal distribution:

$$u_2 \sim N_1\left(\frac{1}{\sqrt{2}}, 2\right).$$

We finally construct an eigenvector associated with $\lambda_3 = 1$. Consider the linear system $(\Sigma - \lambda_3 I)X = O$. In this case, the first equation is $x_1 + x_3 = 0 \Rightarrow x_1 = -x_3$ and the second one is $x_2 - x_3 = 0$ or $x_2 = x_3$. Let $x_3 = 1$. One eigenvector is

$$X_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

In order to ensure the uniqueness of the matrix $[A_1, A_2, A_3]$, we may multiply A_3 by -1 to ensure that the first nonzero element is positive. Then, the third principal component is $u_3 = \frac{1}{\sqrt{3}}[x_1 - x_2 - x_3]$ with $E[u_3] = \frac{1}{\sqrt{3}}[1 - 0 + 1] = \frac{2}{\sqrt{3}}$. The variance of u_3 is indeed equal to λ_3 as

$$\begin{aligned}\text{Var}(u_3) &= \frac{1}{3}[\text{Var}(x_1) + \text{Var}(x_2) + \text{Var}(x_3) - 2\text{Cov}(x_1, x_2) \\ &\quad - 2\text{Cov}(x_1, x_3) + 2\text{Cov}(x_2, x_3)] \\ &= \frac{1}{3}[2 + 2 + 3 - 2(0) - 2(1) + 2(-1)] = 1.\end{aligned}$$

Thus,

$$u_3 \sim N_1\left(\frac{2}{\sqrt{3}}, 1\right).$$

It can easily be verified that $\text{Cov}(u_1, u_2) = 0$, $\text{Cov}(u_1, u_3) = 0$ and $\text{Cov}(u_2, u_3) = 0$. Accordingly, letting

$$U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} A_1' X \\ A_2' X \\ A_3' X \end{bmatrix},$$

$$\text{Cov}(U) = \begin{bmatrix} A_1' \Sigma A_1 & A_1' \Sigma A_2 & A_1' \Sigma A_3 \\ A_2' \Sigma A_1 & A_2' \Sigma A_2 & A_2' \Sigma A_3 \\ A_3' \Sigma A_1 & A_3' \Sigma A_2 & A_3' \Sigma A_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

As well, $\Sigma A_j = \lambda_j A_j$, $j = 1, 2, 3$, that is,

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This completes the computations.

9.4.1. Principal components viewed from differing perspectives

Let X be a $p \times 1$ real vector whose covariance matrix is $\text{Cov}(X) = \Sigma > O$. Assume that $E(X) = O$. Then, $X' \Sigma^{-1} X = c > 0$ is commonly referred to as an ellipsoid of concentration, centered at the origin of the coordinate system, with $X' X$ being the square of the distance between the origin and a point X on the surface of this ellipsoid. A principal axis of this ellipsoid is defined when this squared distance has a stationary point.

The stationary points are determined by optimizing $X'X$ subject to $X'\Sigma^{-1}X = c > 0$. Consider

$$\begin{aligned} w = X'X - \lambda(X'\Sigma^{-1}X - c), \quad \frac{\partial w}{\partial X} = O &\Rightarrow 2X - 2\lambda\Sigma^{-1}X = O \\ \Rightarrow \Sigma^{-1}X = \frac{1}{\lambda}X &\Rightarrow \Sigma X = \lambda X \end{aligned} \quad (i)$$

where λ is a Lagrangian multiplier. It is seen from (i) that λ is an eigenvalue of Σ and X is the corresponding eigenvector. Letting $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ be the eigenvalues and A_1, \dots, A_p be the corresponding eigenvectors, A_1, \dots, A_p give the principal axes of this ellipsoid of concentration. It follows from (i) that

$$c = A'_j \Sigma^{-1} A_j = \frac{1}{\lambda_j} A'_j A_j \Rightarrow A'_j A_j = \lambda_j c.$$

Thus, the length of the j -th principal axis is $2\sqrt{A'_j A_j} = 2\sqrt{\lambda_j c}$.

As another approach, consider a plane passing through the origin. The equation of this plane will be $\beta'X = 0$ where β is a $p \times 1$ constant vector and X is the $p \times 1$ vector of the coordinate system. Without any loss of generality, let $\beta'\beta = 1$. Let the $p \times 1$ vector Y be a point in the Euclidean space. The distance between this point and the plane is then $\beta'Y$. Letting Y be a random point such that $E(Y) = O$ and $\text{Cov}(Y) = \Sigma > O$, the expected squared distance from this point to the plane is $E[\beta'Y]^2 = E[\beta'YY'\beta] = \beta'E(YY')\beta = \beta'\Sigma\beta$. Accordingly, the two-dimensional planar manifold of closest fit to the point Y is that plane whose coefficient vector β is such that $\beta'\Sigma\beta$ is minimized subject to $\beta'\beta = 1$. This, once again, leads to the eigenvalue problem encountered in principal component analysis.

9.5. Sample Principal Components

When $X \sim N_p(O, \Sigma)$, $\Sigma > O$, the maximum likelihood estimator of Σ is $\hat{\Sigma} = \frac{1}{n}S$ where $n > p$ is the sample size and S is the sample sum of products matrix, as defined for instance in Mathai and Haubold (2017b). In general, whether Σ is nonsingular or singular and X is p -variate Gaussian or not, we may take $\hat{\Sigma}$ as an estimate of Σ . The sample eigenvalues and eigenvectors are then available from $\hat{\Sigma}\beta = k\beta$ where $\beta \neq O$ is a $p \times 1$ eigenvector corresponding to the eigenvalue k of $\hat{\Sigma} = \frac{1}{n}S$. For a non-null β , $(\hat{\Sigma} - kI)\beta = O \Rightarrow \hat{\Sigma} - kI$ is singular or $|\hat{\Sigma} - kI| = 0$ and k is a solution of

$$|\hat{\Sigma} - kI| = 0 \Rightarrow (\hat{\Sigma} - k_j I)B_j = O, \quad j = 1, \dots, p. \quad (9.5.1)$$

We may only consider the normalized eigenvectors B_j such that $B'_j B_j = 1$, $j = 1, \dots, p$. If the eigenvalues of Σ are distinct, it can be shown that the eigenvalues of $\hat{\Sigma}$ are also

distinct, $k_1 > k_2 > \dots > k_p$ almost surely. Note that even though $B_j' B_j = 1$, B_j is not unique as one also has $(-B_j)'(-B_j) = 1$. Thus, we require that the first nonzero element of B_j be positive to ensure uniqueness. Since $\hat{\Sigma}$ is a real symmetric matrix, its eigenvalues k_1, \dots, k_p are real, and so are the corresponding normalized eigenvectors B_1, \dots, B_p . As well, there exist a full set of orthonormal eigenvectors B_1, \dots, B_p , $B_j' B_j = 1$, $B_j' B_i = 0$, $i \neq j$, $j = 1, \dots, p$, such that for $B = (B_1, \dots, B_p)$,

$$B' \hat{\Sigma} B = K = \text{diag}(k_1, \dots, k_p), \quad \hat{\Sigma} B = B K \quad \text{and} \quad \hat{\Sigma} = k_1 B_1 B_1' + \dots + k_p B_p B_p'. \quad (9.5.2)$$

Also, observe that $\frac{k_1}{k_1 + \dots + k_p}$ is the proportion of the total variation in the data which is accounted for by the first principal component. Similarly, $\frac{k_1 + \dots + k_r}{k_1 + \dots + k_p}$ is the proportion of the total variation due to the first r principal components, $r \leq p$.

Example 9.5.1. Let X be a 3×1 vector of real scalar random variables, $X' = [x_1, x_2, x_3]$, with $E[X] = \mu$ and covariance matrix $\text{Cov}(X) = \Sigma > O$ where both μ and Σ are unknown. Let the following observation vectors be a simple random sample of size 5 from this population:

$$X_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \quad X_4 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \quad X_5 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

Compute the principal components of X from an estimate of Σ that is based on those observations.

Solution 9.5.1. An estimate of Σ is $\hat{\Sigma} = \frac{1}{n} S$ where n is the sample size and S is the sample sum of products matrix. In this case, $n = 5$. We first compute S . To this end, we determine the sample average vector, the sample matrix, the deviation matrix and finally S . The sample average is $\bar{X} = \frac{1}{n}[X_1 + X_2 + X_3 + X_4 + X_5]$ and the sample matrix is $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5]$. The matrix of sample means is $\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \bar{X}, \bar{X}, \bar{X}]$. The deviation matrix is $\mathbf{X}_d = \mathbf{X} - \bar{\mathbf{X}}$ and the sample sum of products matrix $S = \mathbf{X}_d \mathbf{X}_d' = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]'$. Based on the given random sample, these quantities are

$$\begin{aligned}\bar{X} &= \frac{1}{5} \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \right\} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \\ \mathbf{X} &= [X_1, X_2, X_3, X_4, X_5] = \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}, \\ \mathbf{X}_d &= [X_1 - \bar{X}, X_2 - \bar{X}, X_3 - \bar{X}, X_4 - \bar{X}, X_5 - \bar{X}] \\ &= \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}, \\ S &= \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 & 0 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}' \\ &= \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix}.\end{aligned}$$

An estimate of Σ is $\hat{\Sigma} = \frac{1}{n}S = \frac{1}{5}S$. However, since the eigenvalues of $\hat{\Sigma}$ are those of S multiplied by $\frac{1}{5}$ and the normalized eigenvectors of $\hat{\Sigma}$ and S will then be identical, we will work with S . The eigenvalues of S are available from the equation $|S - \lambda I| = 0$. That is, $(4 - \lambda)[(4 - \lambda)(6 - \lambda) - 4] = 0 \Rightarrow (4 - \lambda)[\lambda^2 - 10\lambda + 20] = 0 \Rightarrow \lambda_1 = 5 + \sqrt{5}$, $\lambda_2 = 4$, $\lambda_3 = 5 - \sqrt{5}$. An eigenvector corresponding to $\lambda_1 = 5 + \sqrt{5}$ can be determined from the system $(S - \lambda_1 I)X = O$ wherein first equation gives $x_1 = 0$ and the third one yields $2x_2 + (1 - \sqrt{5})x_3 = 0$. Taking $x_3 = 2$, $x_2 = -(1 - \sqrt{5})$, and it is easily verified that these values also satisfy the second equation. Thus, an eigenvector, denoted by Y_1 , is the following:

$$Y_1 = \begin{bmatrix} 0 \\ \sqrt{5} - 1 \\ 2 \end{bmatrix} \text{ with its normalized form being } A_1 = \frac{1}{\sqrt{10 - 2\sqrt{5}}} \begin{bmatrix} 0 \\ \sqrt{5} - 1 \\ 2 \end{bmatrix},$$

so that the first principal component is $u_1 = \frac{1}{\sqrt{10 - 2\sqrt{5}}} [(\sqrt{5} - 1)x_2 + 2x_3]$. Let us verify that the variance of u_1 equals λ_1 :

$$\begin{aligned}
\text{Var}(u_1) &= \frac{1}{10 - 2\sqrt{5}} [(\sqrt{5} - 1)^2 \text{Var}(x_2) + 4\text{Var}(x_3) + 2(\sqrt{5} - 1)\text{Cov}(x_2, x_3)] \\
&= \frac{1}{10 - 2\sqrt{5}} [(6 - 2\sqrt{5})(4) + 4(6) + 2(\sqrt{5} - 1)(4)] = \frac{20(5 + \sqrt{5})}{20} \\
&= 5 + \sqrt{5} = \lambda_1.
\end{aligned}$$

An eigenvector corresponding to $\lambda_2 = 4$ is available from $(S - \lambda_2 I)X = O$. The first equation shows that all variables are arbitrary. The second and third equations yield $x_3 = 0$ and $x_2 = 0$. Taking $x_1 = 1$, an eigenvector corresponding to $\lambda_2 = 4$, denoted by Y_2 , is

$$Y_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ which is already normalized so that } A_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Thus, the second principal component is $u_2 = x_1$ and its variance is $\text{Var}(x_1) = 4$. An eigenvector corresponding to $\lambda_3 = 5 - \sqrt{5}$ can be determined from the linear system $(S - \lambda_3 I)X = O$ whose first equation yields $x_1 = 0$, the third one giving $2x_2 + (1 + \sqrt{5})x_3 = 0$. Taking $x_3 = 2$, $x_2 = -(1 + \sqrt{5})$, and it is readily verified that these values satisfy the second equation as well. Then, an eigenvector, denoted by Y_3 , is

$$Y_3 = \begin{bmatrix} 0 \\ -(\sqrt{5} + 1) \\ 2 \end{bmatrix}, \text{ its normalized form being } A_3 = \frac{1}{\sqrt{10 + 2\sqrt{5}}} \begin{bmatrix} 0 \\ -(\sqrt{5} + 1) \\ 2 \end{bmatrix}.$$

In order to select the matrix $[A_1, A_2, A_3]$ uniquely, we may multiply A_3 by -1 so that the first nonzero element is positive. Thus, the third principal component is $u_3 = \frac{1}{\sqrt{10 + 2\sqrt{5}}} [(1 + \sqrt{5})x_2 - 2x_3]$ and, as the following calculations corroborate, its variance is indeed equal to λ_3 :

$$\begin{aligned}
\text{Var}(u_3) &= \frac{1}{\sqrt{10 + 2\sqrt{5}}} [(1 + \sqrt{5})^2 \text{Var}(x_2) + 4\text{Var}(x_3) - 4(1 + \sqrt{5})\text{Cov}(x_2, x_3)] \\
&= \frac{1}{2(5 + \sqrt{5})} [4(6 + 2\sqrt{5}) + 4(6) - 4(1 + \sqrt{5})(2)] = \frac{20}{5 + \sqrt{5}} = 5 - \sqrt{5} = \lambda_3.
\end{aligned}$$

Additionally,

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{5 + \sqrt{5}}{14} \approx 0.52,$$

that is, approximately 52% of the total variance is due to u_1 or nearly 52% of the total variation in the data is accounted for by the first principal component. Also observe that

$$S[A_1, A_2, A_3] = [A_1, A_2, A_3]D, \quad D = \text{diag}(5 + \sqrt{5}, 4, 5 - \sqrt{5}) \text{ or}$$

$$S = [A_1, A_2, A_3]D \begin{bmatrix} A'_1 \\ A'_2 \\ A'_3 \end{bmatrix}.$$

That is,

$$\begin{aligned} S &= \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{\sqrt{5}-1}{\sqrt{10-2\sqrt{5}}} & 0 & -\frac{(1+\sqrt{5})}{\sqrt{10+2\sqrt{5}}} \\ \frac{2}{\sqrt{10-2\sqrt{5}}} & 0 & \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix} \begin{bmatrix} 5 + \sqrt{5} & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 - \sqrt{5} \end{bmatrix} \begin{bmatrix} 0 & \frac{\sqrt{5}-1}{\sqrt{10-2\sqrt{5}}} & \frac{2}{\sqrt{10-2\sqrt{5}}} \\ 1 & 0 & 0 \\ 0 & -\frac{(1+\sqrt{5})}{\sqrt{10+2\sqrt{5}}} & \frac{2}{\sqrt{10+2\sqrt{5}}} \end{bmatrix}, \end{aligned}$$

which completes the computations.

9.5.1. Estimation and evaluation of the principal components

If $X \sim N_p(O, \Sigma)$, $\Sigma > O$, then the maximum likelihood estimator of Σ , denoted by $\hat{\Sigma} = \frac{1}{n}S$ where n is the sample size and S is the sample sum of products matrix. How can one evaluate the eigenvalues and eigenvectors in order to construct the principal components? One method consists of solving the polynomial equation $|\Sigma - \lambda I| = 0$ for the population eigenvalues $\lambda_1, \dots, \lambda_p$, or $|\hat{\Sigma} - kI| = 0$ for obtaining the sample eigenvalues $k_1 \geq k_2 \geq \dots \geq k_p$. Direct evaluation of k by solving the determinantal equation is not difficult when p is small. However, for large values of p , one has to resort to mathematical software or some iterative process. We will illustrate such an iterative process for the population values $\lambda_1 > \lambda_2 > \dots > \lambda_p$ and the corresponding normalized eigenvectors A_1, \dots, A_p (using our notations). Let us assume that the eigenvalues are distinct. Consider the following equation for determining the eigenvalues and eigenvectors:

$$\Sigma A_j = \lambda_j A_j, \quad j = 1, \dots, k. \quad (9.5.3)$$

Take any initial p -vector W_o that is not orthogonal to A_1 , the eigenvector corresponding to the largest eigenvalue λ_1 . If W_o is orthogonal to A_1 , then, in the iterative process, W_j will not reach A_1 . Let $Y_o = \frac{1}{\sqrt{W'_o W_o}} W_o$ be the normalized W_o . Consider the equation

$$\Sigma Y_{j-1} = W_j, \quad \text{that is, } \Sigma Y_0 = W_1, \quad \Sigma Y_1 = W_2, \dots, \quad Y_j = \frac{1}{\sqrt{(W'_j W_j)}} W_j, \quad j = 1, \dots. \quad (9.5.4)$$

Halt the iteration when W_j approximately agrees with W_{j-1} , that is, W_j converges to some W which will then be $\lambda_1 A_1$ or Y_j converges to Y_1 . At each stage, compute the quadratic form $\delta_j = Y_j' \Sigma Y_j$ and make sure that δ_j is increasing. Suppose that W_j converges to some W for certain values of j or as $j \rightarrow \infty$. At that stage, the equation is $\Sigma Y = W$ where $Y = \frac{1}{\sqrt{W'W}} W$, the normalized W . Then, the equation is $\Sigma Y = \sqrt{W'W} Y$. In other words, $\sqrt{W'W} = \lambda_1$ and $Y = A_1$, which are the largest eigenvalue of Σ and the corresponding eigenvector A_1 . That is,

$$\lim_{j \rightarrow \infty} \sqrt{W_j' W_j} = \lambda_1$$

$$\lim_{j \rightarrow \infty} \left[\frac{1}{\sqrt{(W_j' W_j)}} W_j \right] = \lim_{j \rightarrow \infty} Y_j = A_1.$$

The rate of convergence in (9.5.4) depends upon the ratio $\frac{\lambda_2}{\lambda_1}$. If λ_2 is close to λ_1 , then the convergence will be very slow. Hence, the larger the difference between λ_1 and λ_2 , the faster the convergence. It is thus indicated to raise Σ to a suitable power m and initiate the iteration on this Σ^m so that the difference between the resulting λ_1 and λ_2 be magnified accordingly, λ_j^m , $j = 1, \dots, p$, being the eigenvalues of Σ^m . If an eigenvalue is equal to one, then Σ must first be multiplied by a constant so that all the resulting eigenvalues of Σ are well separated, which will not affect the eigenvectors. As well, observe that Σ and Σ^m , $m = 1, 2, \dots$, share the same eigenvectors even though the eigenvalues are λ_j and λ_j^m , $j = 1, \dots, p$, respectively; in other words, the normalized eigenvectors remain the same. After obtaining the largest eigenvalue λ_1 and the corresponding normalized eigenvector A_1 , consider

$$\Sigma_2 = \Sigma - \lambda_1 A_1 A_1', \quad \Sigma = \lambda_1 A_1 A_1' + \lambda_2 A_2 A_2' + \dots + \lambda_p A_p A_p',$$

where A_j is the column eigenvector corresponding to λ_j so that $A_j A_j'$ is a $p \times p$ matrix for $j = 1, 2, \dots, p$. Now, carry out the iteration on Σ_2 , as was previously done on Σ . This will produce λ_2 and the corresponding normalized eigenvector A_2 . Note that λ_2 is the largest eigenvalue of Σ_2 . Next, consider

$$\Sigma_3 = \Sigma_2 - \lambda_2 A_2 A_2'$$

and continue this process until all the required eigenvectors are obtained. Similarly, for small p , the sample eigenvalues $k_1 \geq k_2 \geq \dots \geq k_p$ are available by solving the equation $|\hat{\Sigma} - kI| = 0$; otherwise, the sample eigenvalues and eigenvectors can be computed by applying the iterative process described above with Σ replaced by $\hat{\Sigma}$, λ_j interchanged with k_j and B_j substituted to the eigenvector A_j , $j = 1, \dots, p$.

Example 9.5.2. Principal component analysis is especially called for when the number of variables involved is large. For the sake of illustration, we will take $p = 2$. Consider the following 2×2 covariance matrix

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \text{Cov}(X), \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Construct the principal components associated with this matrix while working with the symbolic representations of the eigenvalues λ_1 and λ_2 rather than their actual values.

Solution 9.5.2. Let us first evaluate the eigenvalues of Σ in the customary fashion. Consider

$$\begin{aligned} |\Sigma - \lambda I| = 0 &\Rightarrow (2 - \lambda)(1 - \lambda) - 1 = 0 \Rightarrow \lambda^2 - 3\lambda + 1 = 0 & (i) \\ &\Rightarrow \lambda_1 = \frac{1}{2}[3 + \sqrt{5}], \quad \lambda_2 = \frac{1}{2}[3 - \sqrt{5}]. \end{aligned}$$

Let us compute an eigenvector corresponding to the largest eigenvalue λ_1 . Then,

$$\begin{aligned} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \lambda_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow \\ (2 - \lambda_1)x_1 - x_2 &= 0 \\ -x_1 + (1 - \lambda_1)x_2 &= 0. & (ii) \end{aligned}$$

Since the system is singular, we need only solve one of these equations. Letting $x_2 = 1$ in (ii), $x_1 = (1 - \lambda_1)$. For illustrative purposes, we will complete the remaining steps with the general parameters λ_1 and λ_2 rather than their numerical values. Hence, one eigenvector is

$$C = \begin{bmatrix} 1 - \lambda_1 \\ 1 \end{bmatrix}, \quad \|C\| = \sqrt{1 + (1 - \lambda_1)^2} \Rightarrow A_1 = \frac{1}{\|C\|}C = \frac{1}{\sqrt{1 + (1 - \lambda_1)^2}} \begin{bmatrix} 1 - \lambda_1 \\ 1 \end{bmatrix}.$$

Thus, the principal components are the following:

$$\begin{aligned} u_1 &= \frac{1}{\|C\|}C'X = \frac{1}{\sqrt{1 + (1 - \lambda_1)^2}}\{(1 - \lambda_1)x_1 + x_2\} \\ u_2 &= \frac{1}{\sqrt{1 + (1 - \lambda_2)^2}}\{(1 - \lambda_2)x_1 + x_2\}. \end{aligned}$$

Let us verify that $\text{Var}(u_1) = \lambda_1$ and $\text{Var}(u_2) = \lambda_2$. Note that

$$\|C\|^2 = 1 + (1 - \lambda_1)^2 = \lambda_1^2 - 2\lambda_1 + 2 = (\lambda_1^2 - 3\lambda_1 + 1) + \lambda_1 + 1 = \lambda_1 + 1, \quad (iii)$$

given the characteristic equation (i). However,

$$\begin{aligned}
 \text{Var}(C'X) &= \text{Var}((1 - \lambda_1)x_1 + x_2) \\
 &= (1 - \lambda_1)^2 \text{Var}(x_1) + \text{Var}(x_2) + 2(1 - \lambda_1)\text{Cov}(x_1, x_2) \\
 &= 2(1 - \lambda_1)^2 + 1 - 2(1 - \lambda_1) \\
 &= 2\lambda_1^2 - 2\lambda_1 + 1 = 2(\lambda_1^2 - 3\lambda_1 + 1) + 4\lambda_1 - 1 = 4\lambda_1 - 1, \quad (iv)
 \end{aligned}$$

given (i), the characteristic (or starting) equation. We now have $\text{Var}(C'X) = \text{Var}((1 - \lambda_1)x_1 + x_2) = 4\lambda_1 - 1$ from (iv) and $\|C\| = \lambda_1 + 1$ from (iii). Hence, we must have $\text{Var}(C'X) = \|C\|^2\lambda_1 = [1 + (1 - \lambda_1)^2]\lambda_1 = (\lambda_1 + 1)\lambda_1$ from (iii). Since

$$\lambda_1(\lambda_1 + 1) = \lambda_1^2 + \lambda_1 = (\lambda_1^2 - 3\lambda_1 + 1) + 4\lambda_1 - 1 = 4\lambda_1 - 1,$$

agrees with (iv), the result is verified for u_1 . Moreover, on replacing λ_1 by λ_2 , we have $\text{Var}(u_2) = \lambda_2$.

9.5.2. L1 and L2 norm principal components

For any $p \times 1$ vector Y , $Y' = (y_1, \dots, y_p)$, where y_1, \dots, y_p are real quantities, the L2 and L1 norms are respectively defined as follows:

$$\|Y\|_2 = (y_1^2 + \dots + y_p^2)^{\frac{1}{2}} = [Y'Y]^{\frac{1}{2}} \Rightarrow \|Y\|_2^2 = Y'Y \quad (9.5.5)$$

$$\|Y\|_1 = |y_1| + |y_2| + \dots + |y_p| = \sum_{j=1}^p |y_j|. \quad (9.5.6)$$

In Sect. 9.5, we set up the principal component analysis on the basis of the sample sum of products matrix, assuming a p -variate real population, not necessarily Gaussian, having μ as its mean value vector and $\Sigma > O$ as its covariance matrix. Then, we considered X_j , $j = 1, \dots, n$, iid vector random variables, that is, a simple random sample of size n from this population such that $E[X_j] = \mu$ and $\text{Cov}(X_j) = \Sigma > O$, $j = 1, \dots, n$. We denoted the sample matrix by $\mathbf{X} = [X_1, \dots, X_n]$, the sample average by $\bar{X} = \frac{1}{n}[X_1 + \dots + X_n]$, the matrix of sample means by $\bar{\mathbf{X}} = [\bar{X}, \dots, \bar{X}]$ and the sample sum of products matrix by $S = [\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]'$. If the population mean value vector is the null vector, that is, $\mu = O$, then we can take $S = \mathbf{X}\mathbf{X}'$. For convenience, we will consider this case. For determining the sample principal components, we then maximized

$$A'\mathbf{X}\mathbf{X}'A \text{ subject to } A'A = 1$$

where A is an arbitrary constant vector that will result in the coefficient vector of the principal components. This can also be stated in terms of maximizing the square of an $L2$ norm subject to $A'A = 1$, that is,

$$\max_{A'A=1} \|A'X\|_2^2. \quad (9.5.7)$$

When carrying out statistical analyses, it turns out that the $L2$ norm is more sensitive to outliers than the $L1$ norm. If we utilize the $L1$ norm, the problem corresponding to (9.5.7) is

$$\max_{A'A=1} \|A'X\|_1. \quad (9.5.8)$$

Observe that when $\mu = O$, the sum of products matrix can be expressed as

$$S = \mathbf{X}\mathbf{X}' = \sum_{j=1}^p X_j X_j', \quad \mathbf{X} = [X_1, \dots, X_n],$$

where X_j is the j -th column of \mathbf{X} or j -th sample vector. Then, the initial optimization problem can be restated as follows:

$$\max_{A'A=1} \|A'X\|_2^2 = \max_{A'A=1} \sum_{j=1}^n A'X_j X_j' A = \max_{A'A=1} \sum_{j=1}^n \|A'X_j\|_2^2, \quad (9.5.9)$$

and the corresponding $L1$ norm optimization problem can be formulated as follows:

$$\max_{A'A=1} \|A'X\|_1 = \max_{A'A=1} \sum_{j=1}^n \|A'X_j\|_1. \quad (9.5.10)$$

We have obtained exact analytical solutions for the coefficient vector A in (9.5.9); however, this is not possible when attempting to optimize (9.5.10) subject to $A'A = 1$. Thankfully, iterative procedures are available in this case.

Let us consider a generalization of the basic principal component analysis. Let W be a $p \times m$, $m < p$, matrix of full rank m . Then, the general problem in $L2$ norm optimization is the following:

$$\max_{W'W=I} \text{tr}(W'SW) = \max_{W'W=I} \sum_{j=1}^n \|W'X_j\|_2^2 \quad (9.5.11)$$

where I is the $m \times m$ identity matrix, the corresponding $L1$ norm optimization problem being

$$\max_{W'W=I} \sum_{j=1}^n \|W'X_j\|_1. \quad (9.5.12)$$

In Sect. 9.4.1, we have considered a dual problem of minimization for constructing principal components. The dual problems of minimization corresponding to (9.5.11) and (9.5.12) can be stated as follows:

$$\min_{W'W=I} \sum_{j=1}^n \|X_j - WW'X_j\|_2^2 \quad (9.5.13)$$

with respect to the $L2$ norm, and

$$\min_{W'W=I} \sum_{j=1}^n \|X_j - WW'X_j\|_1 \quad (9.5.14)$$

with respect to the $L1$ norm. The form appearing in (9.5.13) suggests that the construction of principal components by making use of the $L2$ norm is also connected to general model building, Factor Analysis and related topics. The general mathematical problem pertaining to the basic structure in (9.5.13) is referred to as low-rank matrix factorization. Readers interested in such statistical or mathematical problems may refer to the survey article, Shi et al. (2017). $L1$ norm optimization problems are for instance discussed in Kwak (2008) and Nie and Huang (2016).

9.6. Distributional Aspects of Eigenvalues and Eigenvectors

Let us examine the distributions of the variances of the sample principal components and the coefficient vectors in the sample principal components. Let A be a $p \times p$ orthonormal matrix whose columns are denoted by A_1, \dots, A_p , so that $A_j'A_j = 1$, $A_i'A_j = 0$, $i \neq j$, or $AA' = I$, $A'A = I$. Let $u_j = A_j'X$ be the sample principal components for $j = 1, \dots, p$. Let the $p \times 1$ vector X have a nonsingular normal density with the null vector as its mean value vector, that is, $X \sim N_p(O, \Sigma)$, $\Sigma > O$. We can assume that the Gaussian distribution has a null mean value vector without any loss of generality since we are dealing with variances and covariances, which are free of any location parameter. Consider a simple random sample of size n from this normal population. Letting S denote the sample sum of products matrix, the maximum likelihood estimate of Σ is $\hat{\Sigma} = \frac{1}{n}S$. Given that S has a Wishart distribution having $n - 1$ degrees of freedom, the density of $\hat{\Sigma}$, denoted by $f(\hat{\Sigma})$, is given by

$$f(\hat{\Sigma}) = \frac{n^{\frac{mp}{2}}}{|2\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} |\hat{\Sigma}|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{n}{2}\text{tr}(\Sigma^{-1}\hat{\Sigma})},$$

where Σ is the population covariance matrix and $m = n - 1$, n being the sample size. Let k_1, \dots, k_p be the eigenvalues of $\hat{\Sigma}$; it was shown that $k_1 \geq k_2 \geq \dots \geq k_p > 0$ are

actually the variances of the sample principal components. Letting B_j , $j = 1, \dots, p$, denote the coefficient vectors of the sample principal components and $B = (B_1, \dots, B_p)$, it was established that $B' \hat{\Sigma} B = \text{diag}(k_1, \dots, k_p) \equiv D$. The first nonzero component of B_j is required to be positive for $j = 1, \dots, p$, so that B be uniquely determined. Then, the joint density of B and D is available by transforming $\hat{\Sigma}$ in terms of B and D . Let the joint density of B and D and the marginal densities of D and B be respectively denoted by $g(D, B)$, $g_1(D)$ and $g_2(B)$. In light of the procedures and results presented in Chap. 8 or in Mathai (1997), we have the following joint density:

$$g(D, B)dD \wedge dB = \frac{n^{\frac{mp}{2}}}{|2\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} \left\{ \prod_{i < j} (k_i - k_j) \right\} \\ \times e^{-\frac{n}{2} \text{tr}((B' \Sigma^{-1} B) D)} h(B) dD \quad (9.6.1)$$

where $h(B)$ is the differential element corresponding to B , which is given in Chap. 8 and in Mathai (1997). The marginal densities of D and B are not explicitly available in a convenient form for a general Σ . In that case, they can only be expressed in terms of hypergeometric functions of matrix argument and zonal polynomials. See, for example, Mathai et al. (1995) for a discussion of zonal polynomials. However, when $\Sigma = I$, the joint and marginal densities are available in closed forms. They are

$$g(D, B)dD \wedge dB = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} \left\{ \prod_{i < j} (k_i - k_j) \right\} e^{-\frac{n}{2}(k_1 + \dots + k_p)} h(B) dD, \quad (9.6.2)$$

$$g_1(D) = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p k_j \right\}^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{n}{2}(k_1 + \dots + k_p)} \left\{ \prod_{i < j} (k_i - k_j) \right\}, \quad (9.6.3)$$

$$g_2(B)dB = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} h(B), \quad BB' = I = B'B, \quad k_1 > k_2 > \dots > k_p > 0. \quad (9.6.4)$$

Given the representation of the joint density $g(D, B)$ appearing in (9.6.2), it is readily seen that D and B are independently distributed. Similar results are obtainable for $\Sigma = \sigma^2 I$ where $\sigma^2 > 0$ is a real scalar quantity and I is the identity matrix.

Example 9.6.1. Show that the function $g_1(D)$ given in (9.6.3) is a density for $p = 2$, $n = 6$, $m = 5$, and derive the density of k_1 for this special case.

Solution 9.6.1. We have $\frac{p+1}{2} = \frac{3}{2}$, $\frac{m}{2} - \frac{3}{2} = 1$. As $g_1(D)$ is manifestly nonnegative, it suffices to show that the total integral equals 1. When $p = 2$ and $n = 6$, the constant part of (9.6.3) is the following:

$$\begin{aligned} \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} &= \frac{6^5}{2^5 \Gamma_p(\frac{5}{2})} \frac{\pi^2}{\Gamma_2(1)} \\ &= \frac{3^5}{\sqrt{\pi} \Gamma(\frac{5}{2}) \Gamma(\frac{2}{2})} \frac{\pi^2}{\sqrt{\pi} \Gamma(1) \sqrt{\pi}} \\ &= \frac{3^5}{\sqrt{\pi} (\frac{3}{2}) (\frac{1}{2}) \sqrt{\pi}} \frac{\pi^2}{\sqrt{\pi} \sqrt{\pi}} = 4(3^4). \end{aligned} \quad (i)$$

As for the integral part, we have

$$\begin{aligned} &\int_{k_1=0}^{\infty} \int_{k_2=0}^{k_1} k_1 k_2 (k_1 - k_2) e^{-3(k_1+k_2)} dk_1 \wedge dk_2 \\ &= \int_{k_1=0}^{\infty} k_1^2 e^{-3k_1} \left[\int_{k_2=0}^{k_1} k_2 e^{-3k_2} dk_2 \right] dk_1 - \int_{k_1=0}^{\infty} k_1 e^{-3k_1} \left[\int_{k_2=0}^{k_1} k_2^2 e^{-3k_2} dk_2 \right] dk_1 \\ &= \int_{k_1=0}^{\infty} \left(k_1^2 e^{-3k_1} \left[-\frac{k_1}{3} e^{-3k_1} + \frac{1}{3^2} (1 - e^{-3k_1}) \right] \right. \\ &\quad \left. - k_1 e^{-3k_1} \left[-\frac{k_1^2}{3} e^{-3k_1} - \frac{2k_1}{3^2} e^{-3k_1} + \frac{2}{3^3} (1 - e^{-3k_1}) \right] \right) dk_1 \quad (ii) \\ &= -\frac{1}{3} [\Gamma(4)6^{-4}] + \frac{1}{9} [\Gamma(3)3^{-3}] - \frac{1}{3^2} [\Gamma(3)6^{-3}] \\ &\quad + \frac{1}{3} [\Gamma(4)6^{-4}] + \frac{2}{3^2} [\Gamma(3)6^{-3}] - \frac{2}{3^3} [\Gamma(2)3^{-2}] + \frac{2}{3^2} [\Gamma(2)6^{-2}] \\ &= \frac{1}{3^2} [\Gamma(3)6^{-3}] + \frac{2}{3^3} [\Gamma(2)6^{-2}] + \frac{1}{3^2} [\Gamma(3)3^{-3}] - \frac{2}{3^3} [\Gamma(2)3^{-2}] \\ &= \left[\frac{1}{4} + \frac{2}{4} + 2 - 2 \right] \frac{1}{3^5} = \frac{1}{4(3^4)}. \quad (iii) \end{aligned}$$

The product of (i) and (iii) being equal to 1, this establishes that $g_1(D)$ is indeed a density function for $p = 2$ and $n = 6$. On integrating out k_2 , the resulting integrand appearing in (ii) yields the marginal density of k_1 . Denoting the marginal density of k_1 by $g_{11}(k_1)$, after some simplifications, we have

$$g_{11}(k_1) = 4(3^4) \left\{ \left(\frac{k_1^2}{3^2} + \frac{2k_1}{3^3} \right) e^{-6k_1} + \left(\frac{k_1^2}{3^2} - \frac{2k_1}{3^3} \right) e^{-3k_1} \right\}, \quad 0 < k_1 < \infty,$$

and zero elsewhere. Similarly, given $g_1(D)$, the marginal density of k_2 is obtained by integrating out k_1 from k_2 to ∞ . This completes the computations.

9.6.1. The distributions of the largest and smallest eigenvalues

One can determine the distribution of the j -th largest eigenvalue k_j and thereby, the distribution of the largest one, k_1 , as well as that of the smallest one, k_p . Actually, many authors have been investigating the problem of deriving the distributions of the eigenvalues of a central Wishart matrix and matrix-variate type-1 beta and type-2 beta distributions. Some of the test statistics discussed in Chap. 6 can also be treated as eigenvalue problems involving real and complex type-1 beta matrices. Let $k_1 > k_2 > \cdots > k_p > 0$ be the eigenvalues of the real sampling distribution of a Wishart matrix generated from a p -variate real Gaussian sample. Then, as given in (9.6.3), the joint density of k_1, \dots, k_p , denoted by $g_1(D)$, is the following:

$$g_1(D)dD = \frac{n^{\frac{mp}{2}}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p k_j^{\frac{m}{2} - \frac{p+1}{2}} \right\} e^{-\frac{n}{2}(k_1 + \cdots + k_p)} \left\{ \prod_{i < j} (k_i - k_j) \right\} dD \quad (9.6.5)$$

where $m = n - 1$ is the number of degrees of freedom, n being the sample size. More generally, we may consider a $p \times p$ real matrix-variate gamma distribution having α as its shape parameter and aI , $a > 0$, as its scale parameter matrix, I denoting the identity matrix. Noting that the eigenvalues of aS are equal to the eigenvalues of S multiplied by the scalar quantity a , we may take $a = 1$ without any loss of generality. Let $g(S)dS$ denote the density of the resulting gamma matrix and $g(D)$ be $g(S)$ expressed in terms of $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$, the eigenvalues of S . Then, given the Jacobian of the transformation $S \rightarrow D$ specified in (9.6.7), the joint density of $\lambda_1, \dots, \lambda_p$, denoted by $g_1(D)$ is obtained as

$$g(D)dS = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\alpha) \Gamma_p(\frac{p}{2})} \left\{ \prod_{j=1}^p \lambda_j^{\alpha - \frac{p+1}{2}} \right\} e^{-(\lambda_1 + \cdots + \lambda_p)} \left\{ \prod_{i < j} (\lambda_i - \lambda_j) \right\} dD \equiv g_1(D)dD \quad (9.6.6)$$

where $dD = d\lambda_1 \wedge \dots \wedge d\lambda_p$. The corresponding $p \times p$ complex matrix-variate gamma distribution will have real eigenvalues, also denoted by $\lambda_1 > \cdots > \lambda_p > 0$, the matrix \tilde{S} being Hermitian positive definite in this case. Then, in light of the relationship (9.6a.2) between the differential elements of \tilde{S} and D , the joint density of $\lambda_1, \dots, \lambda_p$, denoted by $\tilde{g}_1(D)$, is given by

$$\tilde{g}(D)d\tilde{S} = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(\alpha)\tilde{\Gamma}_p(p)} \left\{ \prod_{j=1}^p \lambda_j^{\alpha-p} \right\} e^{-(\lambda_1+\dots+\lambda_p)} \left\{ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right\} dD \equiv \tilde{g}_1(D)dD \quad (9.6a.1)$$

where $\tilde{g}(D)$ is $\tilde{g}(\tilde{S})$ expressed in terms of D , $\tilde{g}(\tilde{S})$ denoting the density function of a matrix-variate gamma random variable whose shape parameter and scale parameter matrix are α and I , respectively.

As explained for instance in Mathai (1997), when S and \tilde{S} are the $p \times p$ gamma matrices in the real and complex domains, the integration over the Stiefel manifold yields

$$dS = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left\{ \prod_{i<j} (\lambda_i - \lambda_j) \right\} dD \quad (9.6.7)$$

and

$$d\tilde{S} = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)} \left\{ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right\} dD, \quad (9.6a.2)$$

respectively. When endeavoring to derive the marginal density of λ_j for any fixed j , the difficulty arises from the factor $\prod_{i<j} (\lambda_i - \lambda_j)$. So, let us first attempt to simplify this factor.

9.6.2. Simplification of the factor $\prod_{i<j} (\lambda_i - \lambda_j)$

It is well known that one can write $\prod_{i<j} (\lambda_i - \lambda_j)$ as a Vandermonde determinant which, incidentally, has been utilized in connection with nonlinear transformations in Mathai (1997). That is,

$$\prod_{i<j} (\lambda_i - \lambda_j) = \begin{vmatrix} \lambda_1^{p-1} & \lambda_1^{p-2} & \dots & \lambda_1 & 1 \\ \lambda_2^{p-1} & \lambda_2^{p-2} & \dots & \lambda_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_p^{p-1} & \lambda_p^{p-2} & \dots & \lambda_p & 1 \end{vmatrix} \equiv |A| = |(a_{ij})|, \quad (9.6.8)$$

where the (i, j) -th element $a_{ij} = \lambda_i^{p-j}$ for all i and j . We could consider a cofactor expansion of the determinant, $|A|$, consisting of expanding it in terms of the elements and their cofactors along any row. In this case, it would be advantageous to do so along the i -th row as the cofactors would then be free of λ_i and the coefficients of the cofactors would only be powers of λ_i . However, we would then lose the symmetry with respect to the elements of the cofactors in this instance. Since symmetry is required for the procedure to be discussed, we will consider the general expansion of a determinant, that is,

$$|A| = \sum_K (-1)^{\rho_K} a_{1k_1} a_{2k_2} \cdots a_{pk_p} = \sum_K (-1)^{\rho_K} \lambda_1^{p-k_1} \lambda_2^{p-k_2} \cdots \lambda_p^{p-k_p} \quad (9.6.9)$$

where $K = (k_1, \dots, k_p)$ and (k_1, \dots, k_p) is a given permutation of the numbers $(1, 2, \dots, p)$. Defining ρ_K as the number of transpositions needed to bring (k_1, \dots, k_p) into the natural order $(1, 2, \dots, p)$, $(-1)^{\rho_K}$ will correspond to a $-$ sign for the corresponding term if ρ_K is odd, the sign being otherwise positive. For example, for $p = 3$, the possible permutations are $(1, 2, 3)$, $(1, 3, 2)$, $(2, 3, 1)$, $(2, 1, 3)$, $(3, 1, 2)$, $(3, 2, 1)$, so that there are $3! = 6$ terms. For the sequence $(1, 2, 3)$, $k_1 = 1$, $k_2 = 2$ and $k_3 = 3$; for the sequence $(1, 3, 2)$, $k_1 = 1$, $k_2 = 3$ and $k_3 = 2$, and so on, \sum_K representing the sum of all such terms multiplied by $(-1)^{\rho_K}$. For $(1, 2, 3)$, $\rho_K = 0$ corresponding to a plus sign; for $(1, 3, 2)$, $\rho_K = 1$ corresponding to a minus sign, and so on. Other types of expansions of $|A|$ could also be utilized. As it turns out, the general expansion given in (9.6.9) is the most convenient one for deriving the marginal densities.

In the complex case,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |A|^2 = |AA'| = |A'A|$$

where

$$A' = \begin{vmatrix} \lambda_1^{p-1} & \lambda_2^{p-1} & \cdots & \lambda_p^{p-1} \\ \lambda_1^{p-2} & \lambda_2^{p-2} & \cdots & \lambda_p^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{vmatrix} = [\alpha_1, \alpha_2, \dots, \alpha_p], \quad \alpha_j = \begin{bmatrix} \lambda_j^{p-1} \\ \lambda_j^{p-2} \\ \vdots \\ 1 \end{bmatrix}. \quad (i)$$

Observe that α_j only contains λ_j and that $A'A = \alpha_1 \alpha_1' + \cdots + \alpha_p \alpha_p'$, so that for instance the (i, j) -th element in $\alpha_1 \alpha_1'$ is $\lambda_1^{2p-(i+j)}$. Accordingly, the (i, j) -th element of $\alpha_1 \alpha_1' + \cdots + \alpha_p \alpha_p'$ is $\sum_{r=1}^p \lambda_r^{2p-(i+j)}$. Thus, letting $B = A'A = (b_{ij})$, $b_{ij} = \sum_{r=1}^p \lambda_r^{2p-(i+j)}$. Now, consider the expansion (9.6.9) of $|B|$, that is,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |B| = |A'A| = \sum_K (-1)^{\rho_K} b_{1k_1} b_{2k_2} \cdots b_{pk_p} \quad (ii)$$

where $K = (k_1, \dots, k_p)$ and (k_1, \dots, k_p) is a permutation of the sequence $(1, 2, \dots, p)$. Note that

$$\begin{aligned}
b_{1k_1} &= \lambda_1^{2p-(1+k_1)} + \lambda_2^{2p-(1+k_1)} + \dots + \lambda_p^{2p-(1+k_1)} \\
b_{2k_2} &= \lambda_1^{2p-(2+k_2)} + \lambda_2^{2p-(2+k_2)} + \dots + \lambda_p^{2p-(2+k_2)} \\
&\vdots \\
b_{pk_p} &= \lambda_1^{2p-(p+k_p)} + \lambda_2^{2p-(p+k_p)} + \dots + \lambda_p^{2p-(p+k_p)}. \tag{iii}
\end{aligned}$$

Let us write

$$b_{1k_1} b_{2k_2} \cdots b_{pk_p} = \sum_{r_1, \dots, r_p} \lambda_1^{r_1} \cdots \lambda_p^{r_p}. \tag{iv}$$

Then,

$$\prod_{i < j} (\lambda_i - \lambda_j)^2 = |B| = \sum_K (-1)^{\rho_K} \left[\sum_{r_1, \dots, r_p} \lambda_1^{r_1} \cdots \lambda_p^{r_p} \right] \tag{9.6a.3}$$

where the r_j 's, $j = 1, \dots, p$, are nonnegative integers. We may now express the joint density of the eigenvalues in a systematic way.

9.6.3. The distributions of the eigenvalues

The joint density of $\lambda_1, \dots, \lambda_p$ as specified in (9.6.6) and (9.6a.1) can be expressed as follows. In the real case,

$$\begin{aligned}
f(D)dD &= \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \left(\prod_{j=1}^p \lambda_j^{\alpha - \frac{p+1}{2}} \right) e^{-(\lambda_1 + \dots + \lambda_p)} \left(\sum_K (-1)^{\rho_K} \lambda_1^{p-k_1} \cdots \lambda_p^{p-k_p} \right) dD \\
&= \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} (\lambda_1^{m_1} \cdots \lambda_p^{m_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \tag{9.6.10}
\end{aligned}$$

with

$$m_j = \alpha - \frac{p+1}{2} + p - k_j. \tag{i}$$

In the complex case, the joint density is

$$\begin{aligned}
\tilde{f}(D)dD &= \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} (\lambda_1^{\alpha-p+r_1} \cdots \lambda_p^{\alpha-p+r_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \\
&= \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} (\lambda_1^{m_1} \cdots \lambda_p^{m_p}) e^{-(\lambda_1 + \dots + \lambda_p)} dD \tag{9.6a.4}
\end{aligned}$$

with $m_j = \alpha - p + r_j$ and r_j as defined in (9.6a.3). For convenience, we will use the same symbol m_j for both the real and complex cases; however, in the real case, $m_j = \alpha - \frac{p+1}{2} + p - k_j$ and, in the complex case, $m_j = \alpha - p + r_j$.

The distributions of the largest, smallest and j -th largest eigenvalues were considered by many authors. Earlier works mostly dealt with eigenvalue problems associated with testing hypotheses on the parameters of one or more real Gaussian populations. In such situations, a one-to-one function of the likelihood ratio statistics could be explored in terms of the eigenvalues of a real type-1 beta distributed matrix-variate random variable. In a series of papers, Pillai constructed the distributions of the seven largest eigenvalues in the type-1 beta distribution and produced percentage points as well; the reader may refer for example to Pillai (1964) and the references therein. In a series of papers including Khatri (1964), Khatri addressed the distributions of eigenvalues in the real and complex domains. In a series of papers, Krishnaiah and his co-authors dealt with various distributional aspects of eigenvalues, see for instance Krishnaiah et al. (1973). Clemm et al. (1973) computed upper percentage points of the distribution of the eigenvalues of the Wishart matrix. James (1964) considered the eigenvalue problem of different types of matrix-variate random variables and determined their distributions in terms of functions of matrix arguments and zonal polynomials. In a series of papers Davis, dealt with the distributions of eigenvalues by creating and solving systems of differential equations, see for example Davis (1972). Edelman (1991) discussed the distributions and moments of the smallest eigenvalue of Wishart type matrices. Johnstone (2001) examined the distribution of the largest eigenvalue in Principal Component Analysis. Recently, Chiani (2014) and James and Lee (2021) discussed the distributions of the eigenvalues of Wishart matrices. The methods employed in these papers lead to representations of the distributions of eigenvalues in terms of Pfaffians of skew symmetric matrices, incomplete gamma functions, multiple integrals, functions of matrix argument and zonal polynomials, ratios of determinants, solutions of differential equations, and so on. None of those methods yield tractable forms for the distribution or density functions of eigenvalues.

In the next subsections, we provide, in explicit forms, the exact distributions of any of the j -th largest eigenvalue of a general real or complex matrix-variate gamma type matrix, either as finite sums when a certain quantity is a positive integer or as a product of infinite series in the general non-integer case. These include, for instance, the distributions of the largest and smallest eigenvalue as well as the joint distributions of several of the largest or smallest eigenvalues, and readily apply to the real and complex Wishart distributions.

9.6.4. Density of the smallest eigenvalue λ_p in the real matrix-variate gamma case

We will initially examine the situation where $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is an integer, so that m_j in the real matrix-variate gamma case is a positive integer. We will integrate out $\lambda_1, \dots, \lambda_{p-1}$ to obtain the marginal density of λ_p . Since m_j is a positive integer, we can integrate by parts. For instance,

$$\begin{aligned} \int_{\lambda_1=\lambda_2}^{\infty} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 &= [-\lambda_1^{m_1} e^{-\lambda_1}]_{\lambda_2}^{\infty} + [-m_1 \lambda_1^{m_1-1} e^{-\lambda_1}]_{\lambda_2}^{\infty} + \dots + [-e^{-\lambda_1}]_{\lambda_2}^{\infty} \\ &= \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \lambda_2^{m_1 - \mu_1} e^{-\lambda_2}, \end{aligned} \quad (i)$$

and integrating $\lambda_1, \dots, \lambda_{j-1}$ gives the following:

$$\begin{aligned} &\int_{\lambda_1} \int_{\lambda_2} \dots \int_{\lambda_{j-1}} \lambda_1^{m_1} \dots \lambda_{j-1}^{m_{j-1}} e^{-(\lambda_1 + \dots + \lambda_{j-1})} d\lambda_1 \wedge \dots \wedge d\lambda_{j-1} \\ &= \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \dots \\ &\quad \times \sum_{\mu_{j-1}=0}^{m_1 - \mu_1 + \dots + m_{j-1}} \frac{(m_1 - \mu_1 + \dots + m_{j-1})!}{(j-1)^{\mu_{j-1}+1} (m_1 - \mu_1 + \dots + m_{j-1} - \mu_{j-1})!} \lambda_j^{m_1 - \mu_1 + \dots + m_{j-1} - \mu_{j-1}} e^{-j\lambda_j} \\ &\equiv \phi_{j-1}(\lambda_j). \end{aligned} \quad (ii)$$

Hence, the following result:

Theorem 9.6.1. *When $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is a positive integer, where m_j is defined in (9.6.10), the marginal density of the smallest eigenvalue λ_p of the $p \times p$ real gamma distributed matrix with parameters (α, I) , denoted by $f_{1p}(\lambda_p)$, is the following:*

$$\begin{aligned} &f_{1p}(\lambda_p) d\lambda_p \\ &= c_K \phi_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} \\ &= c_K \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \dots \sum_{\mu_{p-1}=0}^{m_1 - \mu_1 + \dots + m_{p-1}} \\ &\quad \times \frac{(m_1 - \mu_1 + \dots + m_{p-1})!}{(p-1)^{\mu_{p-1}+1} (m_1 - \mu_1 + \dots + m_{p-1} - \mu_{p-1})!} \lambda_p^{m_1 - \mu_1 + \dots + m_{p-1} - \mu_{p-1} + m_p} e^{-p\lambda_p} d\lambda_p \end{aligned} \quad (9.6.11)$$

for $0 < \lambda_p < \infty$, where

$$c_K = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K},$$

$\phi_{j-1}(\lambda_j)$ is specified in (ii) and $K = (k_1, \dots, k_p)$ is defined in (9.6.9).

In the complex $p \times p$ matrix-variate gamma case, r_j is as defined in (9.6a.3) and the expression for $\phi_{j-1}(\lambda_j)$ given in (ii) remains the same with the exception that m_j in the complex case is $m_j = \alpha - p + r_j$. Then, in the complex domain, the density of λ_p , denoted by $\tilde{f}_{1p}(\lambda_p)$, is the following:

Theorem 9.6a.1. When $m_j = \alpha - p + r_j$ is a positive integer, where r_j is defined in (9.6a.3), the density of the smallest eigenvalue of the complex matrix-variate gamma distribution is the following:

$$\begin{aligned} & \tilde{f}_{1p}(\lambda_p) d\lambda_p \\ &= \tilde{c}_K \phi_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-p\lambda_p} \\ &= \tilde{c}_K \sum_{\mu_1=0}^{m_1} \frac{m_1!}{(m_1 - \mu_1)!} \sum_{\mu_2=0}^{m_1 - \mu_1 + m_2} \frac{(m_1 - \mu_1 + m_2)!}{2^{\mu_2+1} (m_1 - \mu_1 + m_2 - \mu_2)!} \cdots \sum_{\mu_{p-1}=0}^{m_1 - \mu_1 + \cdots + m_{p-1}} \\ & \times \frac{(m_1 - \mu_1 + \cdots + m_{p-1})!}{(p-1)^{\mu_{p-1}+1} (m_1 - \mu_1 + \cdots + m_{p-1} - \mu_{p-1})!} \lambda_p^{m_1 - \mu_1 + \cdots + m_{p-1} - \mu_{p-1} + m_p} e^{-p\lambda_p} d\lambda_p \end{aligned} \quad (9.6a.5)$$

for $0 < \lambda_p < \infty$, where

$$\tilde{c}_K = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_k}.$$

Note 9.6.1. In the complex Wishart case, $\alpha = m$ where m is the number of degrees of freedom, which is a positive integer. Hence, Theorem 9.6a.1 gives the final result in the general case for that distribution. One can obtain the joint density of the $p - j + 1$ smallest eigenvalues from $\phi_{j-1}(\lambda_j)$ as defined in (ii), both in the real and complex cases. If the scale parameter matrix of the gamma distribution is of the form aI where $a > 0$ is a real scalar and I is the identity matrix, then the distributions of the eigenvalues can also be obtained from the proposed procedure since for any square matrix B , the eigenvalues of aB are $a v_j$'s where the v_j 's are the eigenvalues of B . In the case of real Wishart distributions originating from a sample of size n from a p -variate Gaussian population

whose covariance matrix is the identity matrix, the λ_j 's are multiplied by the constant $a = \frac{n}{2}$. If the Wishart matrix is not an estimator obtained from the sample values, they should only be multiplied by $a = \frac{1}{2}$ in the real case, no multiplicative constant being necessary in the complex Wishart case.

9.6.5. Density of the largest eigenvalue λ_1 in the real matrix-variate gamma case

Consider the case of $m_j = \alpha - \frac{p+1}{2} + p - k_j$ being an integer first. Then, in the real case, m_j is as defined in (9.6.10). One has to integrate out $\lambda_p, \dots, \lambda_2$ in order to obtain the marginal density of λ_1 . As the initial step, consider the integration of λ_p , that is,

$$\begin{aligned} \text{Step 1 integral : } & \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p \\ & = [-\lambda_p^{m_p} e^{-\lambda_p}]_0^{\lambda_{p-1}} + \dots + [-m_p! e^{-\lambda_p}]_0^{\lambda_{p-1}} \\ & = m_p! - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \lambda_{p-1}^{m_p - v_p} e^{-\lambda_{p-1}}. \end{aligned} \quad (i)$$

Now, multiplying each term by $\lambda_{p-1}^{m_{p-1}} e^{-\lambda_{p-1}}$ and integrating by parts, we have the second step integral:

Step 2 integral

$$\begin{aligned} & = m_p! \int_{\lambda_{p-1}=0}^{\lambda_{p-2}} \lambda_{p-1}^{m_{p-1}} e^{-\lambda_{p-1}} d\lambda_{p-1} - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \int_{\lambda_{p-1}=0}^{\lambda_{p-2}} \lambda_{p-1}^{m_p - v_p + m_{p-1}} e^{-2\lambda_{p-1}} d\lambda_{p-1} \\ & = m_p! m_{p-1}! - m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{m_{p-1}!}{(m_{p-1} - v_{p-1})!} \lambda_{p-2}^{m_{p-1} - v_{p-1}} e^{-\lambda_{p-2}} \\ & \quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p - v_p + m_{p-1}}} \\ & \quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{2^{v_{p-1}+1} (m_p - v_p + m_{p-1} - v_{p-1})!} \lambda_{p-2}^{m_p - v_p + m_{p-1} - v_{p-1}} e^{-2\lambda_{p-2}}. \end{aligned} \quad (ii)$$

At the j -th step of integration, there will be 2^j terms of which $\frac{2^j}{2} = 2^{j-1}$ will be positive and 2^{j-1} will be negative. All the terms at the j -th step can be generated by 2^j sequences

of zeros and ones. Each sequence (each row) consists of j positions with a zero or one occupying each position. Depending on whether the number of ones in a sequence is odd or even, the corresponding term will start with a minus or a plus sign, respectively. The following are the sequences, where the last column indicates the sign of the term:

$$\begin{array}{l}
 \text{Step 1: } \begin{array}{c} 0 \quad + \\ 1 \quad - \end{array}, \quad \text{Step 2: } \begin{array}{c} 0 \quad 0 \quad + \\ 0 \quad 1 \quad - \\ 1 \quad 0 \quad - \\ 1 \quad 1 \quad + \end{array}, \quad \text{Step 3: } \begin{array}{c} 0 \quad 0 \quad 0 \quad + \\ 0 \quad 0 \quad 1 \quad - \\ 0 \quad 1 \quad 0 \quad - \\ 0 \quad 1 \quad 1 \quad + \end{array}, \quad \text{Step 3: } \begin{array}{c} 1 \quad 0 \quad 0 \quad - \\ 1 \quad 0 \quad 1 \quad + \\ 1 \quad 1 \quad 0 \quad + \\ 1 \quad 1 \quad 1 \quad - \end{array} \\
 \\
 \text{Step 4: } \begin{array}{c} 0 \quad 0 \quad 0 \quad 0 \quad + \\ 0 \quad 0 \quad 0 \quad 1 \quad - \\ 0 \quad 0 \quad 1 \quad 0 \quad - \\ 0 \quad 0 \quad 1 \quad 1 \quad + \end{array}, \quad \begin{array}{c} 0 \quad 1 \quad 0 \quad 0 \quad - \\ 0 \quad 1 \quad 0 \quad 1 \quad + \\ 0 \quad 1 \quad 1 \quad 0 \quad + \\ 0 \quad 1 \quad 1 \quad 1 \quad - \end{array}, \quad \begin{array}{c} 1 \quad 0 \quad 0 \quad 0 \quad - \\ 1 \quad 0 \quad 0 \quad 1 \quad + \\ 1 \quad 0 \quad 1 \quad 0 \quad + \\ 1 \quad 0 \quad 1 \quad 1 \quad - \end{array}, \quad \begin{array}{c} 1 \quad 1 \quad 0 \quad 0 \quad + \\ 1 \quad 1 \quad 0 \quad 1 \quad - \\ 1 \quad 1 \quad 1 \quad 0 \quad - \\ 1 \quad 1 \quad 1 \quad 1 \quad + \end{array}
 \end{array}$$

All the terms at the j -th step can be written down by using the following rules:

- (1): If the first entry in a sequence (row) is zero, then the corresponding factor in the term is $m_p!$;
- (2): If the first entry in a sequence is 1, then the corresponding factor in the term is $\sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p-v_p)!}$ or this sum multiplied by $\lambda_{p-1}^{m_p-v_p} e^{-\lambda_{p-1}}$ if this 1 is the last entry in the sequence;
- (3): If the r -th and $(r-1)$ -th entries in the sequence both equal zero, then the corresponding factor in the term is $m_{p-r+1}!$;
- (4): If the r -th entry in the sequence is zero and the $(r-1)$ -th entry is 1, then the corresponding factor in the term is $\frac{(n_{r-1}+m_{p-r+1})!}{(\eta_{r-1}+1)^{n_{r-1}+m_{p-r+1}+1}}$, where n_{r-1} is the argument of the denominator factorial and $(\eta_{r-1})^{v_{p-r+2}+1}$ is the factor in the denominator corresponding to the $(r-1)$ -th entry;
- (5): If the r -th entry in the sequence is 1 and the $(r-1)$ -th entry is zero, then the corresponding factor in the term is $\sum_{v_{p-r+1}=0}^{m_{p-r+1}} \frac{m_{p-r+1}!}{(m_{p-r+1}-v_{p-r+1})!}$ or this sum multiplied by $\lambda_{p-r}^{m_{p-r+1}-v_{p-r+1}} e^{-\lambda_{p-r}}$ if this 1 is the last entry in the sequence;
- (6): If the r -th and $(r-1)$ -th entries in the sequence are both equal to 1, then the corresponding factor in the term is $\sum_{v_{p-r+1}=0}^{n_{r-1}+m_{p-r+1}} \frac{(n_{r-1}+m_{p-r+1})!}{(\eta_{r-1}+1)^{v_{p-r+1}} (n_{r-1}+m_{p-r+1}-v_{p-r+1})!}$ or this sum multiplied by $\lambda_{p-r}^{n_{r-1}+m_{p-r+1}-v_{p-r+1}} e^{-(\eta_{r-1}+1)\lambda_{p-r}}$ if this 1 is the last entry in the sequence, where n_{r-1} and η_{r-1} are defined in rule (4). By applying the above rules, let us write down the terms at step 3, that is, $j = 3$. The sequences are then the following:

$$\text{Step 3 sequences: } \begin{array}{cccc} 0 & 0 & 0 & + \\ 0 & 0 & 1 & - \\ 0 & 1 & 0 & - \\ 0 & 1 & 1 & + \end{array}, \begin{array}{cccc} 1 & 0 & 0 & - \\ 1 & 0 & 1 & + \\ 1 & 1 & 0 & + \\ 1 & 1 & 1 & - \end{array}. \quad (iii)$$

The corresponding terms in the order of the sequences are the following:

Step 3 integral

$$\begin{aligned} &= m_p! m_{p-1}! m_{p-2}! - m_p! m_{p-1}! \sum_{v_{p-2}=0}^{m_{p-2}} \frac{m_{p-2}!}{(m_{p-2} - v_{p-2})!} \lambda_{p-3}^{m_{p-2}-v_{p-2}} e^{-\lambda_{p-3}} \\ &\quad - m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{m_{p-1}!}{(m_{p-1} - v_{p-1})!} \frac{(m_{p-1} - v_{p-1} + m_{p-2})!}{2^{m_{p-1}-v_{p-1}+m_{p-2}+1}} \\ &\quad + m_p! \sum_{v_{p-1}=0}^{m_{p-1}} \frac{(m_{p-1})!}{(m_{p-1} - v_{p-1})!} \sum_{v_{p-2}=0}^{m_{p-1}-v_{p-1}+m_{p-2}} \frac{(m_{p-1} - v_{p-1} + m_{p-2})!}{2^{v_{p-2}+1}(m_{p-1} - v_{p-1} + m_{p-2} - v_{p-2})!} \\ &\quad \quad \quad \times \lambda_{p-3}^{m_{p-1}-v_{p-1}+m_{p-2}-v_{p-2}} e^{-2\lambda_{p-3}} \\ &\quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p-v_p+m_{p-1}}} m_{p-2}! \\ &\quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \frac{(m_p - v_p + m_{p-1})!}{2^{m_p-v_p+m_{p-1}+1}} \sum_{v_{p-2}=0}^{m_{p-2}} \frac{m_{p-2}!}{(m_{p-2} - v_{p-2})!} \lambda_{p-3}^{m_{p-2}-v_{p-2}} e^{-\lambda_{p-3}} \\ &\quad + \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_p-v_p+m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{(m_p - v_p + m_{p-1} - v_{p-1})!} \\ &\quad \quad \quad \times \frac{(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2})!}{3^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}}} \\ &\quad - \sum_{v_p=0}^{m_p} \frac{m_p!}{(m_p - v_p)!} \sum_{v_{p-1}=0}^{m_p-v_p+m_{p-1}} \frac{(m_p - v_p + m_{p-1})!}{2^{v_{p-1}+1}(m_p - v_p + m_{p-1} - v_{p-1})!} \\ &\quad \quad \quad \times \sum_{v_{p-2}=0}^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}} \frac{(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2})!}{3^{v_{p-2}+1}(m_p - v_p + m_{p-1} - v_{p-1} + m_{p-2} - v_{p-2})!} \\ &\quad \quad \quad \times \lambda_{p-3}^{m_p-v_p+m_{p-1}-v_{p-1}+m_{p-2}-v_{p-2}} e^{-3\lambda_{p-3}}. \quad (iv) \end{aligned}$$

The terms in (iv) are to be multiplied by $\lambda_{p-3}^{m_{p-3}} e^{-\lambda_{p-3}}$ to obtain the final result if we are stopping, that is, if $p = 4$. The terms in (iv) can be verified by multiplying the step 2

integral by $\lambda_{p-2}^{m_{p-2}} e^{-\lambda_{p-2}}$ and then integrating (ii), term by term. Denoting the sum of the terms at the j -th step by $\psi_j(\lambda_{p-j})$, the density of the largest eigenvalue λ_1 , denoted by $f_{11}(\lambda_1)$, is the following:

Theorem 9.6.2. *When $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is a positive integer and $\psi_j(\lambda_{p-j})$ is as defined in the preceding paragraph, where k_j is given in (9.6.10), the density of λ_1 in the real matrix-variate gamma case, denoted by $f_{11}(\lambda_1)$, is the following:*

$$f_{11}(\lambda_1) d\lambda_1 = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\alpha) \Gamma_p(\frac{p}{2})} \sum_K (-1)^{\rho_k} \psi_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \quad (9.6.12)$$

where it is assumed that m_j is a positive integer.

In the corresponding complex case, the procedure is parallel and the expression for $\psi_j(\lambda_{p-j})$ remains the same, except that m_j will then be equal to $\alpha - p + r_j$, where r_j is defined in (9.6a.3). Assuming that m_j is a positive integer and letting the density in the complex case be denoted by $\tilde{f}_{11}(\lambda_1)$, we have the following result:

Theorem 9.6a.2. *Letting $m_j = \alpha - p + r_j$ be a positive integer and $\psi_j(\lambda_{p-j})$ have the same representation as in the real case except that $m_j = \alpha - p + r_j$, in the complex case, the density of λ_1 , denoted by $\tilde{f}_{11}(\lambda_1)$, is the following:*

$$\tilde{f}_{11}(\lambda_1) d\lambda_1 = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_k} \psi_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty. \quad (9.6a.6)$$

Note 9.6.2. One can also compute the density of the j -th eigenvalue λ_j from Theorems 9.6.1 and 9.6.2. For obtaining the density of λ_j , one has to integrate out $\lambda_1, \dots, \lambda_{j-1}$ and $\lambda_p, \lambda_{p-1}, \dots, \lambda_{j+1}$, the resulting expressions being available from the $(j-1)$ -th step when integrating $\lambda_1, \dots, \lambda_{j-1}$ and from the $(p-j)$ -th step when integrating $\lambda_p, \lambda_{p-1}, \dots, \lambda_{j+1}$.

9.6.6. Density of the largest eigenvalue λ_1 in the general real case

By general case, it is meant that $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is not a positive integer. In the real Wishart case, m_j will then be a half-integer; however in the general gamma case α can be any real number greater than $\frac{p-1}{2}$. In this general case, we will expand the exponential part and then integrate term by term. That is,

$$\begin{aligned}
\text{Step 1 integral: } \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p &= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \int_{\lambda_p=0}^{\lambda_{p-1}} \lambda_p^{m_p+v_p} d\lambda_p \\
&= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \frac{1}{(m_p + v_p + 1)} \lambda_{p-1}^{m_p+v_p+1}. \tag{i}
\end{aligned}$$

Continuing this process, we have

$$\begin{aligned}
\text{Step } j \text{ integral} &= \sum_{v_p=0}^{\infty} \frac{(-1)^{v_p}}{v_p!} \frac{1}{m_p + v_p + 1} \sum_{v_{p-1}=0}^{\infty} \frac{(-1)^{v_{p-1}}}{v_{p-1}!} \frac{1}{m_p + v_p + m_{p-1} + v_{p-1} + 2} \\
&\quad \cdots \sum_{v_{p-j+1}=0}^{\infty} \frac{(-1)^{v_{p-j+1}}}{v_{p-j+1}!} \frac{1}{m_p + v_p + \cdots + m_{p-j+1} + v_{p-j+1} + j} \\
&\equiv \Delta_j(\lambda_{p-j}). \tag{ii}
\end{aligned}$$

Then, in the general real case, the density of λ_1 , denoted by $f_{21}(\lambda_1)$, is the following:

Theorem 9.6.3. *When $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is not a positive integer, where k_j is as specified in (9.6.10), the density of the largest eigenvalue λ_1 in the general real matrix-variate gamma case, denoted by $f_{21}(\lambda_1)$, is given by*

$$f_{21}(\lambda_1)d\lambda_1 = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})\Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} \Delta_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \tag{9.6.13}$$

where $\Delta_j(\lambda_{p-j})$ is defined in (ii).

The corresponding density of λ_1 in the general situation of the complex matrix-variate gamma distribution is given in the next theorem. Observe that in the complex Wishart case, m_j is an integer and hence there is no general case to consider.

Theorem 9.6a.3. *When $m_j = \alpha - p + r_j$ is not a positive integer, where r_j is as defined in (9.6a.3), the density of λ_1 in the complex case, denoted by $\tilde{f}_{21}(\lambda_1)$, is given by*

$$\tilde{f}_{21}(\lambda_1)d\lambda_1 = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p)\tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} \Delta_{p-1}(\lambda_1) \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1, \quad 0 < \lambda_1 < \infty, \tag{9.6a.7}$$

where the $\Delta_j(\lambda_{p-j})$ has the representation specified in (ii) except that $m_j = \alpha - p + r_j$.

9.6.7. Density of the smallest eigenvalue λ_p in the general real case

Once again, ‘general case’ is understood to mean that $m_j = \alpha - \frac{p+1}{2} + p - k_j$ is not a positive integer, where k_j is defined in (9.6.10). For the real Wishart distribution, ‘general case’ corresponds to m_j being a half-integer. In order to determine the density of the smallest eigenvalue, we will integrate out $\lambda_1, \dots, \lambda_{p-1}$. We initially evaluate the following integral:

$$\begin{aligned} \text{Step 1 integral: } \int_{\lambda_1=\lambda_2}^{\infty} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 &= \Gamma(m_1 + 1) - \int_{\lambda_1=0}^{\lambda_2} \lambda_1^{m_1} e^{-\lambda_1} d\lambda_1 \\ &= \Gamma(m_1 + 1) - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \lambda_2^{m_1 + \mu_1 + 1}. \end{aligned} \quad (i)$$

The second step consists of integrating out λ_2 from the expression obtained in (i) multiplied by $\lambda_2^{m_2} e^{-\lambda_2}$:

Step 2 integral:

$$\begin{aligned} \Gamma(m_1 + 1) \int_{\lambda_2=\lambda_3}^{\infty} \lambda_2^{m_2} e^{-\lambda_2} d\lambda_2 - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \int_{\lambda_2=\lambda_3}^{\infty} \lambda_2^{m_1 + \mu_1 + m_2 + 1} e^{-\lambda_2} d\lambda_2 \\ = \Gamma(m_1 + 1) \Gamma(m_2 + 1) - \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{1}{m_2 + \mu_2 + 1} \lambda_3^{m_2 + \mu_2 + 1} \\ - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1! (m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2) \\ + \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{\lambda_3^{m_1 + \mu_1 + m_2 + \mu_2 + 2}}{m_1 + \mu_1 + m_2 + \mu_2 + 2}. \end{aligned} \quad (ii)$$

A pattern is now seen to emerge. At step j , there will be 2^j terms, of which 2^{j-1} will start with a plus sign and 2^{j-1} will start with a minus sign. All the terms at the j -th step are available from the 2^j sequences of zeros and ones provided in Sect. 9.6.5. The terms can be written down by utilizing the following rules:

- (1): If the sequence starts with a zero, then the corresponding factor in the term is $\Gamma(m_1 + 1)$;
- (2): If the sequence starts with a 1, then the corresponding factor in the term is $\sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!} \frac{1}{m_1 + \mu_1 + 1}$ or this series multiplied by $\lambda_2^{m_1 + \mu_1 + 1}$ if this 1 is the last entry in the sequence;
- (3): If the r -th entry in the sequence is a zero and the $(r - 1)$ -th entry in the sequence is also zero, then the corresponding factor in the term is $\Gamma(m_r + 1)$;

(4): If the r -th entry in the sequence is a zero and the $(r - 1)$ -th entry in the sequence is a 1, then the corresponding factor in the term is $\Gamma(n_{r-1} + m_r + 1)$ where n_{r-1} is the denominator factor in the $(r - 1)$ -th factor excluding the factorial;

(5): If the r -th entry in the sequence is 1 and the $(r - 1)$ -th entry is zero, then the corresponding factor in the term is $\sum_{\mu_r=0}^{\infty} \frac{(-1)^{\mu_r}}{\mu_r!} \frac{1}{m_r + \mu_r + 1}$ or this series multiplied by $\lambda_{r+1}^{m_r + \mu_r + 1}$ if this 1 happens to be the last entry in the sequence;

(6): If the r -th entry in the sequence is 1 and the $(r - 1)$ -th entry is also 1, then the corresponding factor in the term is $\sum_{\mu_r=0}^{\infty} \frac{(-1)^{\mu_r}}{\mu_r!} \frac{1}{n_{r-1} + m_r + \mu_r + 1}$ or this series multiplied by $\lambda_{r+1}^{n_{r-1} + m_r + \mu_r + 1}$ if this 1 is the last entry in the sequence, where n_{r-1} is the factor appearing in the denominator of the $(r - 1)$ -th factor excluding the factorial.

These rules enable one to write down all the terms at any step. For example for $j = 3$, that is, at the third step, the terms are available from the following step 3 sequences:

$$\begin{array}{cccc} 0 & 0 & 0 & + \\ 0 & 0 & 1 & - \\ 0 & 1 & 0 & - \\ 0 & 1 & 1 & + \end{array} , \quad \begin{array}{cccc} 1 & 0 & 0 & - \\ 1 & 0 & 1 & + \\ 1 & 1 & 0 & + \\ 1 & 1 & 1 & - \end{array} .$$

The terms corresponding to the sequences in the order are the following:

Step 3 integral

$$\begin{aligned} &= \Gamma(m_1 + 1)\Gamma(m_2 + 1)\Gamma(m_3 + 1) \\ &- \Gamma(m_1 + 1)\Gamma(m_2 + 1) \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{1}{m_3 + \mu_3 + 1} \lambda_4^{m_3 + \mu_3 + 1} \\ &- \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_2 + \mu_2 + 1)} \Gamma(m_2 + \mu_2 + m_3 + 2) \\ &+ \Gamma(m_1 + 1) \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!} \frac{1}{m_2 + \mu_2 + 1} \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_2 + \mu_2 + m_3 + \mu_3 + 2}}{m_2 + \mu_2 + m_3 + \mu_3 + 2} \\ &- \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2)\Gamma(m_3 + 1) \\ &+ \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \Gamma(m_1 + \mu_1 + m_2 + 2) \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_3 + \mu_3 + 1}}{(m_3 + \mu_3 + 1)} \end{aligned}$$

$$\begin{aligned}
& + \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_1 + \mu_1 + m_2 + \mu_2 + 2)} \\
& \quad \times \Gamma(m_1 + \mu_1 + m_2 + \mu_2 + m_3 + 3) \\
& - \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \sum_{\mu_2=0}^{\infty} \frac{(-1)^{\mu_2}}{\mu_2!(m_1 + \mu_1 + m_2 + \mu_2 + 2)} \\
& \quad \times \sum_{\mu_3=0}^{\infty} \frac{(-1)^{\mu_3}}{\mu_3!} \frac{\lambda_4^{m_1 + \mu_1 + m_2 + \mu_2 + m_3 + \mu_3 + 3}}{(m_1 + \mu_1 + m_2 + \mu_2 + m_3 + \mu_3 + 3)}. \tag{iii}
\end{aligned}$$

Then, the step j will be the following, denoted by $w_j(\lambda_{j+1})$:

$$\begin{aligned}
w_j(\lambda_{j+1}) & = \Gamma(m_1 + 1) \cdots \Gamma(m_j + 1) - \cdots + (-1)^j \sum_{\mu_1=0}^{\infty} \frac{(-1)^{\mu_1}}{\mu_1!(m_1 + \mu_1 + 1)} \cdots \\
& \quad \times \sum_{\mu_j=0}^{\infty} \frac{(-1)^{\mu_j}}{\mu_j!} \frac{\lambda_{j+1}^{m_1 + \mu_1 + \cdots + m_j + \mu_j + j}}{(m_1 + \mu_1 + \cdots + m_j + \mu_j + j)}. \tag{iv}
\end{aligned}$$

Theorem 9.6.4. *The density of λ_p for the general real matrix-variate gamma distribution, denoted by $f_{2p}(\lambda_p)$, is the following:*

$$f_{2p}(\lambda_p) d\lambda_p = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2}) \Gamma_p(\alpha)} \sum_K (-1)^{\rho_K} w_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p, \quad 0 < \lambda_p < \infty, \tag{9.6.14}$$

where the $w_j(\lambda_{j+1})$ is defined in (iv).

The corresponding distribution of λ_p for a general complex matrix-variate gamma distribution, denoted by $\tilde{f}_{2p}(\lambda_p)$, is the following:

Theorem 9.6a.4. *In the general complex case, in which instance $m_j = \alpha - p + r_j$ is not a positive integer, r_j being as defined in (9.6a.3), the density of the smallest eigenvalue λ_p , denoted by $\tilde{f}_{2p}(\lambda_p)$, is given by*

$$\tilde{f}_{2p}(\lambda_p) d\lambda_p = \frac{\pi^{p(p-1)}}{\tilde{\Gamma}_p(p) \tilde{\Gamma}_p(\alpha)} \sum_K (-1)^{\rho_K} \sum_{r_1, \dots, r_p} w_{p-1}(\lambda_p) \lambda_p^{m_p} e^{-\lambda_p} d\lambda_p, \quad 0 < \lambda_p < \infty, \tag{9.6a.8}$$

where $w_j(\lambda_{j+1})$ has the representation given in (iv) above for the real case, except that $m_j = \alpha - p + r_j$.

Note 9.6.3. In the complex Wishart case, $\alpha = m$ where $m > p - 1$ is the number of degrees of freedom, which is a positive integer. Hence, in this instance, one would simply apply Theorem 9.6a.1. It should also be observed that one can integrate out $\lambda_1, \dots, \lambda_{j-1}$ by using the procedure described in Theorem 9.6.4 and integrate out $\lambda_p, \dots, \lambda_{j+1}$ by employing the procedure provided in Theorem 9.6.3, and thus derive the density of λ_j or the joint density of any set of successive λ_j 's. In a similar manner, one can obtain the density of λ_j or the joint density of any set of successive λ_j 's in the complex domain by making use of the procedures outlined in Theorems 9.6a.3 and 9.6a.4.

References

- M. Chiani (2014): Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution, *Journal of Multivariate Analysis*, **129**, 69–81.
- D. S. Clemm, A. K. Chattopadhyay and P. R. Krishnaiah (1973): Upper percentage points of the individual roots of the Wishart matrix, *Sankhya, Series B*, **35**(3), 325–338.
- A. W. Davis (1972): On the marginal distributions of the latent roots of the multivariate beta matrix, *The Annals of Mathematical Statistics*, **43**(5), 1664–1670.
- A. Edelman (1991): The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type, *Linear Algebra and its Applications*, **159**, 55–80.
- A. T. James (1964): Distributions of matrix variates and latent roots derived from normal samples, *The Annals of Mathematical Statistics*, **35**, 475–501.
- O. James and H.-N. Lee (2021): Concise probability distributions of eigenvalues of real-valued Wishart matrices, <https://arxiv.org/ftp/arxiv/paper/1402.6757.pdf>
- I. M. Johnstone (2001): On the distribution of the largest eigenvalue in Principal Components Analysis, *The Annals of Statistics*, **29**(2), 295–327.
- C. G. Khatri (1964): Distribution of the largest or smallest characteristic root under null hypothesis concerning complex multivariate normal populations. *The Annals of Mathematical Statistics*, **35**, 1807–1810
- P. R. Krishnaiah, F.J. Schuurmannan and V.B. Waikar (1973): Upper percentage points of the intermediate roots of the manova matrix, *Sankhya, Ser.B*, **35**(3), 339–358
- N. Kwak (2008): Principal component analysis based on L1-norm maximization, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **30**(9), 1672–1680.
- A. M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A. M. Mathai and H. J. Haubold (2017a): *Linear Algebra for Physicists and Engineers*, De Gruyter, Germany.

A. M. Mathai and H. J. Haubold (2017b): *Probability and Statistics for Physicists and Engineers*, De Gruyter, Germany.

A. M. Mathai, S. B. Provost and T. Hayakawa (1995): *Bilinear Forms and Zonal Polynomials*, Springer Lecture Notes, New York.

F. Nie and H. Huang (2016): Non-greedy L₂₁-norm maximization for principal component analysis, *arXiv:1603.08293v1[cs,LG]* 28 March 2016.

K. C. S. Pillai (1964): On the distribution of the largest seven roots of a matrix in multivariate analysis, *Biometrika*, **51**(1/2), 270–275.

J. Shi, X. Zheng and W. Yang (2017): Survey on probabilistic models of low-rank matrix factorization, *Entropy*, **19**, 424, doi:10.3390/e19080424.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 10

Canonical Correlation Analysis



10.1. Introduction

We will keep utilizing the same notations in this chapter. More specifically, lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed above letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

The necessary theory for the study of *Canonical Correlation Analysis* has already been introduced in Chap. 1, including the problem of optimizing a real bilinear form subject to two quadratic form constraints. This topic happens to be connected to the prediction problem. In regression analysis, the objective consists of seeking the best prediction function of a real scalar variable y based on a collection of preassigned real scalar variables x_1, \dots, x_k . It was previously determined that the regression of y on x_1, \dots, x_k , or the *best* predictor of y at preassigned values of x_1, \dots, x_k , is the conditional expectation of y at the specified values of x_1, \dots, x_k , that is, $E[y|x_1, \dots, x_k]$ where E denotes the expected value. In this case, *best* is understood to mean 'in the minimum mean square' sense. Now,

consider the following generalization of this problem. Suppose that we wish to determine the best prediction function for a set of real scalar variables y_1, \dots, y_q , on the basis of a collection of real scalar variables x_1, \dots, x_p , where p needs not be equal to q . Since individual variables are available from linear functions of those variables, we will convert the problem into one of predicting a linear function of y_1, \dots, y_q from an arbitrary linear function of x_1, \dots, x_p , and vice versa if we are interested in determining the association between two sets of variables. Let the linear functions be $u = \alpha_1 x_1 + \dots + \alpha_p x_p = \alpha' X$ with $\alpha' = (\alpha_1, \dots, \alpha_p)$ and $X' = (x_1, \dots, x_p)$ and $v = \beta_1 y_1 + \dots + \beta_q y_q = \beta' Y$ with $\beta' = (\beta_1, \dots, \beta_q)$ and $Y' = (y_1, \dots, y_q)$, where the coefficient vectors α and β are arbitrary. Let us provide an interpretation of *best predictor* in the case of two linear functions. As a criterion, we may make use of the maximum joint scatter, that is, the joint variation in u and v as measured by the covariance between u and v or, equivalently, the maximum scale-free covariance, namely, the correlation between u and v , and optimize this joint variation. Given the properties of linear functions of real scalar variables, we obtain the variances of linear functions and covariance between linear functions as follows: $\text{Var}(u) = \alpha' \Sigma_{11} \alpha$, $\text{Var}(v) = \beta' \Sigma_{22} \beta$, $\text{Cov}(u, v) = \alpha' \Sigma_{12} \beta = \beta' \Sigma_{21} \alpha$, $\Sigma'_{12} = \Sigma_{21}$, where $\Sigma_{11} > O$ and $\Sigma_{22} > O$ are the variance-covariance matrices of X and Y , respectively, and $\Sigma_{12} = \Sigma'_{21}$ accounts for the covariance between X and Y . Letting the augmented vector $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ and its associated covariance matrix be Σ , we have

$$\Sigma = \text{Cov} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \text{Cov}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y) \end{bmatrix} \equiv \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Our aim is to maximize $\alpha' \Sigma_{12} \beta = \beta' \Sigma_{21} \alpha$. When the coefficient vectors α and β are unrestricted, the optimization of $\alpha' \Sigma_{12} \beta$ proves meaningless since the quantity $\alpha' \Sigma_{12} \beta$ can vary from $-\infty$ to ∞ . Consequently, we impose the constraints, $\alpha' \Sigma_{11} \alpha = 1$ and $\beta' \Sigma_{22} \beta = 1$, to the coefficient vectors α and β . Accordingly, the mathematical problem consists of optimizing $\alpha' \Sigma_{12} \beta$ subject to $\alpha' \Sigma_{11} \alpha = 1$ and $\beta' \Sigma_{22} \beta = 1$.

Letting

$$w = \alpha' \Sigma_{12} \beta - \frac{\rho_1}{2} (\alpha' \Sigma_{11} \alpha - 1) - \frac{\rho_2}{2} (\beta' \Sigma_{22} \beta - 1) \quad (i)$$

where ρ_1 and ρ_2 are the Lagrangian multipliers, we differentiate w with respect to α and β and equate the resulting functions to null vectors. When differentiating with respect to β , we may utilize the equivalent form $\beta' \Sigma_{21} \alpha = \alpha' \Sigma_{12} \beta$. We then obtain the following equations:

$$\frac{\partial}{\partial \alpha} w = O \Rightarrow \Sigma_{12}\beta - \rho_1 \Sigma_{11}\alpha = O \quad (ii)$$

$$\frac{\partial}{\partial \beta} w = O \Rightarrow \Sigma_{21}\alpha - \rho_2 \Sigma_{22}\beta = O. \quad (iii)$$

On pre-multiplying (ii) by α' and (iii), by β' , and using the fact that $\alpha' \Sigma_{11}\alpha = 1$ and $\beta' \Sigma_{22}\beta = 1$, one has $\rho_1 = \rho_2 \equiv \rho$ and $\alpha' \Sigma_{12}\beta = \rho$. Thus,

$$\begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} O \\ O \end{bmatrix} \quad (10.1.1)$$

and

$$\text{Cov}(\alpha'X, \beta'Y) = \alpha' \Sigma_{12}\beta = \beta' \Sigma_{21}\alpha = \rho. \quad (10.1.2)$$

Hence, the maximum value of $\text{Cov}(\alpha'X, \beta'Y)$ yields the largest ρ . It follows from (ii) that $\alpha = \frac{1}{\rho} \Sigma_{11}^{-1} \Sigma_{12}\beta$ which, once substituted in (iii) yields

$$[\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 \Sigma_{22}]\beta = O \Rightarrow [\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \rho^2 I]\beta = O.$$

This entails that $\rho^2 = \lambda$, an eigenvalue of $B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ or its symmetrized form $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$, and that β is a corresponding eigenvector. Similarly, by obtaining a representation of β from (iii), substituting it in (ii) and proceeding as above, it is seen that $\rho^2 = \lambda$ is an eigenvalue of $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ or its symmetrized form $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$, and that α is a corresponding eigenvector. Hence manifestly, all the nonzero eigenvalues of A coincide with those of B . If $p \leq q$ and Σ_{12} is of full rank p , then $A > O$ (real positive definite) and $B \geq O$ (real positive semi-definite), whereas if $q \leq p$ and Σ_{21} is of full rank p , then $A \geq O$ (real positive semi-definite) and $B > O$ (real positive definite). If $p = q$ and Σ_{12} is of full rank p , then A and B are both positive definite. If $p \leq q$ and Σ_{12} is of full rank p , then one should start with A and compute all the p nonzero eigenvalues of A since A will be of lower order; on the other hand, if $q \leq p$ and Σ_{21} is of full rank q , then one ought to begin with B and determine all the nonzero eigenvalues of B . Thus, one can obtain the common nonzero eigenvalues of A and B or their symmetrized forms by making use of one of these sets of steps. Let us denote the largest value of these common eigenvalues $\lambda = \rho^2$ by $\lambda_{(1)}$ and the corresponding eigenvectors with respect to A and B , by $\alpha_{(1)}$ and $\beta_{(1)}$, where the eigenvectors are normalized via the constraints $\alpha'_{(1)} \Sigma_{11}\alpha_{(1)} = 1$ and $\beta'_{(1)} \Sigma_{22}\beta_{(1)} = 1$. Then, $(u_1, v_1) \equiv (\alpha'_{(1)}X, \beta'_{(1)}Y)$ is the first pair of canonical variables in the sense that u_1 is the best predictor of v_1 and v_1 is the best predictor of u_1 . Similarly, letting $\rho^2_{(i)} = \lambda_{(i)}$ be the i -th largest common

eigenvalue of A and B and the corresponding eigenvectors such that $\alpha'_{(i)} \Sigma_{11} \alpha_{(i)} = 1$ and $\beta'_{(i)} \Sigma_{22} \beta_{(i)} = 1$, be denoted by $\alpha_{(i)}$ and $\beta_{(i)}$, the i -th largest correlation between $u = \alpha'X$ and $v = \beta'Y$ will be equal to $\alpha'_{(i)} \Sigma_{12} \beta_{(i)} = \rho_{(i)} = \sqrt{\lambda_{(i)}}$, $i = 1, \dots, p$, p denoting the common number of nonzero eigenvalues of A and B , and occur when $u = u_i = \alpha'_{(i)}X$ and $v = v_i = \beta'_{(i)}Y$, u_i and v_i being the i -th pair of canonical variables. Clearly, $\text{Var}(u_i)$ and $\text{Var}(v_i)$, $i = 1, \dots, p$, are both equal to one. Once again, best is taken to mean 'in the minimum mean square' sense. Hence, the following results:

Theorem 10.1.1. *Letting Σ , A , B , ρ , $\alpha_{(i)}$, $\beta_{(i)}$, u_i and v_i be as previously defined,*

$$\max_{\alpha' \Sigma_{11} \alpha = 1, \beta' \Sigma_{22} \beta = 1} [\alpha' \Sigma_{12} \beta] = \alpha'_{(1)} \Sigma_{12} \beta_{(1)} = \rho_{(1)} \quad (10.1.3)$$

where $\rho_{(1)}$ is the largest ρ or the largest canonical correlation, that is, the largest correlation between the first pair of canonical variables, $u = \alpha'X$ and $v = \beta'Y$, which is equal to the correlation between u_1 and v_1 , with $\rho_{(1)}^2 = \lambda_{(1)}$, the common largest eigenvalue of A and B . Similarly, we have

$$\min_{\alpha' \Sigma_{11} \alpha = 1, \beta' \Sigma_{22} \beta = 1} [\alpha' \Sigma_{12} \beta] = \alpha'_{(p)} \Sigma_{12} \beta_{(p)} = \rho_{(p)} \quad (10.1.4)$$

where $\rho_{(p)}$, which is the smallest nonzero value of ρ with $\rho_{(p)}^2 = \lambda_{(p)}$, the common smallest nonzero eigenvalue of A and B , represents the smallest canonical correlation between u and v or the correlation between u_p and v_p

This maximum correlation between the linear functions $\alpha'X$ and $\beta'Y$ or the correlation between the best predictors u_1 and v_1 or the maximum value of ρ is called the *first canonical correlation* between the sets X and Y in the sense the correlation between u_1 and v_1 attains its maximum value. When $p = 1$ or $q = 1$, the canonical correlation becomes the *multiple correlation*, and when $p = 1$ and $q = 1$, it is simply the correlation between two real scalar random variables. The matrix of the nonzero eigenvalues of A and B , denoted by Λ , is $\Lambda = \text{diag}(\lambda_{(1)}, \dots, \lambda_{(p)})$ when $p \leq q$ and Σ_{12} is of full rank p ; otherwise, p is replaced by q in Λ .

It should be noted that, for instance, the canonical variable $\beta'Y$ such that β satisfies the constraint $\beta' \Sigma_{22} \beta = 1$ is identical to $b' \Sigma_{22}^{-1/2} Y$ such that $b'b = 1$ since $\beta' \Sigma_{22} \beta = b' \Sigma_{22}^{-1/2} \Sigma_{22} \Sigma_{22}^{-1/2} b = b'b$. Accordingly, letting the $\lambda_{(i)}$'s as well as A and B be as previously defined, our definition of a canonical variable, that is, $u_i = \alpha'_{(i)}X$ and $v_i = \beta'_{(i)}Y$, coincides with the customary one, that is, $u_i^* = a'_i \Sigma_{11}^{-1/2} X$ where a_i is the eigenvector with respect to B which is associated with $\lambda_{(i)}$ and normalized by requiring that $a'_i a_i = 1$,

and $v_i^* = b_i' \Sigma_{22}^{-1/2} Y$ where b_i is an eigenvector with respect to B corresponding to $\lambda_{(i)}$ and such that $b_i' b_i = 1$. It can be readily proved that the canonical variables u_1^*, \dots, u_p^* (or equivalently the u_i 's) are uncorrelated, as $\text{Cov}(u_i^*, u_j^*) = a_i' \Sigma_{11}^{-1/2} \Sigma_{11} \Sigma_{11}^{-1/2} a_j = 0$ for $i \neq j$ since the normed eigenvectors a_i are orthogonal to one another. It can be similarly established that the v_i^* 's or, equivalently, the v_i 's are uncorrelated. Clearly, $\text{Cov}(u_i^*, u_j^*) = \text{Cov}(u_i, u_j) = 1$ and $\text{Cov}(v_i^*, v_j^*) = \text{Cov}(v_i, v_j) = 1$. We now demonstrate that, for $i \neq k$, the canonical variables, u_i and v_k are uncorrelated. First, consider the equation $A a = \lambda a$, that is,

$$\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a = \lambda a.$$

On pre-multiplying both sides by $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$, we obtain $B b = \lambda b$ where $b = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a$. Thus, if λ and a constitute an eigenvalue-eigenvector pair for A , then λ and b must also form an eigenvalue-eigenvector pair for B , and vice versa with $a = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} b$. By definition, $\text{Cov}(u_i^*, v_k^*) = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}} b_k$ where the vector $b_k = \theta \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k$, θ being a positive constant such that the Euclidean norm of b_k is one. Note that since $b_k' b_k = \theta^2 a_k' A a_k = \theta^2 a_k' \lambda_{(k)} a_k = 1$, θ must be equal to $1/\sqrt{\lambda_{(k)}}$. Thus, $b_k = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}}$ with $\Sigma_{11}^{-\frac{1}{2}} a_k = \alpha_{(k)}$ and $b_k = \Sigma_{22}^{-\frac{1}{2}} \beta_{(k)}$, which is equivalent to (iii) with $\alpha = \alpha_{(k)}$, $\beta = \beta_{(k)}$ and $\rho = \rho_{(k)}$, that is, $\beta_{(k)} = \Sigma_{22}^{-1} \Sigma_{21} \alpha_{(k)} / \rho_{(k)}$. Then, $\text{Cov}(u_i^*, v_k^*) = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}} =$ the (i, k) th element of $\text{diag}(\lambda_{(1)}, \dots, \lambda_{(p)}) / \sqrt{\lambda_{(k)}}$, which is equal to 0 whenever $i \neq k$. As expected, $\text{Cov}(u_k^*, v_k^*) = \sqrt{\lambda_{(k)}} = \rho_{(k)}$, and $\text{Cov}(u_i, v_k) = \alpha_{(i)}' \Sigma_{12} \beta_{(k)} \stackrel{(iii)}{=} \alpha_{(i)}' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha_{(k)} / \rho_{(k)} = a_i' \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} a_k / \sqrt{\lambda_{(k)}} = \text{Cov}(u_i^*, v_k^*)$ for $i, k = 1, \dots, p$, assuming that $p \leq q$ and Σ_{12} is of full rank; if $p \geq q$ and Σ_{12} is of full rank, A and B will then share q nonzero eigenvalues.

10.1.1. An invariance property

An interesting property of canonical correlations is now pointed out. Consider the following nonsingular transformations of X and Y : Let $X_1 = A_1 X$ and $Y_1 = B_1 Y$ where A_1 is a $p \times p$ nonsingular constant matrix and B_1 is a $q \times q$ constant nonsingular matrix so that $|A_1| \neq 0$ and $|B_1| \neq 0$. Now, consider the linear functions $\alpha' X_1 = \alpha' A_1 X$ and $\beta' Y_1 = \beta' B_1 Y$ whose variances and covariance are as follows:

$$\begin{aligned} \text{Var}(\alpha' Y_1) &= \text{Var}(\alpha' A_1 Y) = \alpha' A_1 \Sigma_{11} A_1' \alpha, \quad \text{Var}(\beta' Y_1) = \text{Var}(\beta' B_1 Y) = \beta' B_1 \Sigma_{22} B_1' \beta \\ \text{Cov}(\alpha' X_1, \beta' Y_1) &= \alpha' A_1 \Sigma_{12} B_1' \beta = \beta' B_1 \Sigma_{21} A_1' \alpha. \end{aligned}$$

On imposing the conditions $\text{Var}(\alpha'X_1) = 1$ and $\text{Var}(\beta'Y_1) = 1$, and maximizing $\text{Cov}(\alpha'X_1, \beta'Y_1)$ by means of the previously used procedure, we arrive at the equations

$$A_1 \Sigma_{12} B_1' \beta - \rho_1 A_1 \Sigma_{11} A_1' \alpha = 0 \quad (iv)$$

$$-\rho_2 B_1 \Sigma_{22} B_1' \beta + B_1 \Sigma_{21} A_1' \alpha = 0. \quad (v)$$

On pre-multiplying (iv) by α' and (v) by β' , one has $\rho_1 = \rho_2 \equiv \rho$, say. Equations (iv) and (v) can then be re-expressed as

$$\begin{bmatrix} -\rho A_1 \Sigma_{11} A_1' & A_1 \Sigma_{12} B_1' \\ B_1 \Sigma_{21} A_1' & -\rho B_1 \Sigma_{22} B_1' \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} A_1 & O \\ O & B_1 \end{bmatrix} \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} A_1' & O \\ O & B_1' \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Taking the determinant of the coefficient matrix and equating it to zero to determine the roots, we have

$$\left| \begin{bmatrix} A_1 & O \\ O & B_1 \end{bmatrix} \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \begin{bmatrix} A_1' & O \\ O & B_1' \end{bmatrix} \right| = 0 \Rightarrow \quad (10.1.5)$$

$$\left| \begin{bmatrix} -\rho \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\rho \Sigma_{22} \end{bmatrix} \right| = 0. \quad (10.1.6)$$

As can be seen from (10.1.6), (10.1.1) and (10.1.5) have the same roots ρ , which means that the canonical correlation ρ is invariant under nonsingular linear transformations. Observe that when Σ_{12} is of full rank p and $p \leq q$, $\rho_{(1)}, \dots, \rho_{(p)}$ corresponding to the nonzero roots of (10.1.1) or (10.1.6), encompasses all the canonical correlations, so that, in that case, we have a matrix of canonical correlations. Hence, the following result:

Theorem 10.1.2. *Let X , a $p \times 1$ vector of real scalar random variables x_1, \dots, x_p , and Y , a $q \times 1$ vector of real scalar random variables y_1, \dots, y_q , have a joint distribution. Then, the canonical correlations between X and Y are invariant under nonsingular linear transformations, that is, the canonical correlations between X and Y are the same as those between $A_1 X$ and $B_1 Y$ where $|A_1| \neq 0$ and $|B_1| \neq 0$.*

10.2. Pairs of Canonical Variables

As previously explained, $\lambda_{(1)}$ which denotes the largest eigenvalue of the matrix $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ or its symmetrized form $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$, as well the largest

eigenvalue of $B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ or $\Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$, also turns out to be equal to $\rho_{(1)}^2$, the square of the largest root of equation (10.1.1). Having evaluated $\lambda_{(1)}$, we compute the corresponding eigenvectors $\alpha_{(1)}$ and $\beta_{(1)}$ and normalize them via the constraints $\alpha'_{(1)} \Sigma_{11} \alpha_{(1)} = 1$ and $\beta'_{(1)} \Sigma_{22} \beta_{(1)} = 1$, which produces the first pair of canonical variables: $(u_1, v_1) = (\alpha'_{(1)} X, \beta'_{(1)} Y)$. We then take the second largest nonzero eigenvalue of A or B , denote it by $\lambda_{(2)}$, compute the corresponding eigenvectors $\alpha_{(2)}$ and $\beta_{(2)}$ and normalize them so that $\alpha'_{(2)} \Sigma_{11} \alpha_{(2)} = 1$ and $\beta'_{(2)} \Sigma_{22} \beta_{(2)} = 1$, which yields the second pair of canonical variables: $(u_2, v_2) = (\alpha'_{(2)} X, \beta'_{(2)} Y)$. Continuing this process with all of the p nonzero eigenvalues if $p \leq q$ and Σ_{12} is of full rank p , or with all of the q nonzero eigenvalues if $q \leq p$ and Σ_{21} is of full rank q , will produce a complete set of canonical variables pairs: $(u_i, v_i) = (\alpha'_{(i)} X, \beta'_{(i)} Y)$, $i = 1, \dots, p$ or q .

Since the symmetrized forms of A and B are symmetric and nonnegative definite, all of their eigenvalues will be nonnegative and all nonzero eigenvalues will be positive. As is explained in Chapter 1 and Mathai and Haubold (2017a), all the eigenvalues of real symmetric matrices are real and for such matrices, there exists a full set of orthogonal eigenvectors whether some of the roots are repeated or not. Hence, $\alpha'_{(j)} X$ will be uncorrelated with all the linear functions $\alpha'_{(r)} X$, $r = 1, 2, \dots, j - 1$, and $\beta'_{(j)} Y$ will be uncorrelated with $\beta'_{(r)} Y$, $r = 1, 2, \dots, j - 1$. When constructing the second pair of canonical variables, we may impose the condition that the second linear functions $\alpha' X$ and $\beta' Y$ must be uncorrelated with the first pair $\alpha'_{(1)} X$ and $\beta'_{(1)} Y$, respectively, by taking two more Lagrangian multipliers, adding the conditions $\text{Cov}(\alpha' X, \alpha'_{(1)} X) = \alpha' \Sigma_{11} \alpha_{(1)} = 0$ and $\beta' \Sigma_{22} \beta_{(1)} = 0$ to the optimizing function w and carrying out the optimization. We will then realize that these additional conditions are redundant and that the original optimizing equations are recovered, as was observed in the case of Principal Components. Similarly, we could incorporate the conditions $\alpha' \Sigma_{11} \alpha_{(r)} = 0$, $r = 1, \dots, j - 1$ when constructing $\alpha_{(j)}$ and similar conditions when constructing $\beta_{(j)}$. However, these uncorrelatedness conditions will become redundant in the optimization procedure. Note that $\lambda_{(1)} = \rho_{(1)}^2$ is the square of the first canonical correlation. Thus, the first canonical correlation is denoted by $\rho_{(1)}$. Similarly $\lambda_{(r)} = \rho_{(r)}^2$ is the square of the r -th canonical correlation, $r = 1, \dots, p$ when $p \leq q$ and Σ_{12} is of full rank p . That is, $\rho_{(1)}, \dots, \rho_{(p)}$, the p nonzero roots of (10.1.1) when $p \leq q$ and Σ_{12} is of full rank p , are canonical correlations, $\rho_{(r)}$ being called the r -th canonical correlation which is the r -th largest root of the determinantal equation (10.1.1). If $p \leq q$ and Σ_{12} is not of full rank p , then there will be fewer nonzero canonical correlations.

Example 10.2.1. Let $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$ be a 5×1 real vector random variable where X is 3×1 and Y is 2×1 . Let the covariance matrix of Z be Σ where

$$\Sigma = \begin{bmatrix} 3 & 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix},$$

$$\Sigma_{22} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \Sigma_{21} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

Construct the pairs of canonical variables.

Solution 10.2.1. We need the following quantities:

$$\begin{aligned} \Sigma_{22}^{-1} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \Sigma_{11}^{-1} = \frac{1}{15} \begin{bmatrix} 6 & -3 & 0 \\ -3 & 9 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \\ \Sigma_{22}^{-1} \Sigma_{21} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix}, \\ \Sigma_{11}^{-1} \Sigma_{12} &= \frac{1}{15} \begin{bmatrix} 6 & -3 & 0 \\ -3 & 9 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix}; \end{aligned}$$

$$A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21},$$

$$\begin{aligned} B &= \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \frac{1}{45} \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \\ &= \frac{1}{45} \begin{bmatrix} 20 & -10 \\ -7 & 26 \end{bmatrix}. \end{aligned}$$

Let us compute the eigenvalues of B since it is 2×2 whereas A is 3×3 . The characteristic equation of $45B$ is $(20 - \lambda)(26 - \lambda) - 70 = 0 \Rightarrow \lambda^2 - 46\lambda + 450 = 0$. The roots are $\lambda_1 = 23 + \sqrt{79}$, $\lambda_2 = 23 - \sqrt{79}$. Hence, the eigenvalues of B are $\rho_j = \frac{\lambda_j}{45}$, that

is, $\rho_1 = \frac{23+\sqrt{79}}{45}$, $\rho_2 = \frac{23-\sqrt{79}}{45}$. We have denoted the second set of real scalar random variables by Y , $Y' = [y_1, y_2]$. An eigenvector corresponding to ρ_1 is available from $(B - \rho_1 I)Y = O$. Since the right-hand side is null, we may omit the denominator. The first equation is then $(-3 - \sqrt{79})y_1 - 10y_2 = 0$. Taking $y_1 = 1$, $y_2 = -\frac{1}{10}(3 + \sqrt{79})$. It is easily verified that these values will also satisfy the second equation in $(B - \rho_1 I)Y = O$. An eigenvector, denoted by β_1 , is the following:

$$\beta_1 = \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix}.$$

We normalize β_1 through $\beta_1' \Sigma_{22} \beta_1 = 1$. To this end, consider

$$\begin{aligned} \beta_1' \Sigma_{22} \beta_1 &= [1, -\frac{1}{10}(3 + \sqrt{79})] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{1}{25}(79 - 2\sqrt{79}). \end{aligned}$$

Hence a normalized β_1 , denoted by $\beta_{(1)}$, and the corresponding canonical variable v_1 are the following:

$$\beta_{(1)} = \frac{5}{\sqrt{79 - 2\sqrt{79}}} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix}, \quad v_1 = \frac{5}{\sqrt{79 - 2\sqrt{79}}} [y_1 - \frac{1}{10}(3 + \sqrt{79})y_2].$$

The second eigenvalue of B is $\rho_2 = \frac{1}{45}(23 - \sqrt{79})$. An eigenvector corresponding to ρ_2 is available from the equation $(B - \rho_2 I)Y = O$. The second equation gives $-7y_1 + (3 + \sqrt{79})y_2 = 0$. Taking $y_2 = 1$, $y_1 = \frac{1}{7}(3 + \sqrt{79})$. Hence, an eigenvector corresponding to ρ_2 , denoted by β_2 , is the following:

$$\beta_2 = \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix}.$$

We normalize this vector through the constraint $\beta_2' \Sigma_{22} \beta_2 = 1$. Consider

$$\begin{aligned} \beta_2' \Sigma_{22} \beta_2 &= \left[\frac{1}{7}(3 + \sqrt{79}), 1 \right] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{1}{49}(316 + 26\sqrt{79}). \end{aligned}$$

Hence, the normalized eigenvector, denoted by $\beta_{(2)}$, and the corresponding canonical variable v_2 are

$$\beta_{(2)} = \frac{7}{\sqrt{316 + 26\sqrt{79}}} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix}, \quad v_2 = \frac{7}{\sqrt{316 + 26\sqrt{79}}} \left[\frac{1}{7}(3 + \sqrt{79})y_1 + y_2 \right].$$

We will obtain the eigenvectors resulting from $(A - \rho_1 I)X = O$ from the eigenvector β_1 instead of solving the equation relating to A , as the presence of the term $3 + \sqrt{79}$ can make the computations tedious. From equation (ii) of Sect. 10.1, we have

$$\begin{aligned}\alpha_1 &= \frac{1}{\rho_1} \Sigma_{11}^{-1} \Sigma_{12} \beta_1 \\ &= \frac{\sqrt{45}}{\sqrt{23 + \sqrt{79}}} \left(\frac{1}{15} \right) \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{\sqrt{45}}{15(\sqrt{23 + \sqrt{79}})} \begin{bmatrix} 6 + \frac{3}{10}(3 + \sqrt{79}) \\ -3 - \frac{9}{10}(3 + \sqrt{79}) \\ 5 - \frac{5}{10}(3 + \sqrt{79}) \end{bmatrix}.\end{aligned}$$

Let us normalize this vector by requiring that $\alpha_1' \Sigma_{11} \alpha_1 = 1$ or $\alpha_1' \Sigma_{11} \alpha_1 = \frac{1}{\rho_1^2} \beta_1' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \beta_1 = \gamma_1^2$, say:

$$\begin{aligned}\gamma_1^2 &= \frac{45}{15(23 + \sqrt{79})} \begin{bmatrix} 1, -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{45}{15(23 + \sqrt{79})} \begin{bmatrix} 1, -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \begin{bmatrix} 11 & 2 \\ 2 & 14 \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{10}(3 + \sqrt{79}) \end{bmatrix} \\ &= \frac{45}{15(23 + \sqrt{79})(25)} [553 + 11\sqrt{79}].\end{aligned}$$

Hence, the normalized α_1 , denoted by $\alpha_{(1)}$, is the following:

$$\alpha_{(1)} = \frac{\alpha_1}{|\gamma_1|} = \frac{5}{\sqrt{(15)}\sqrt{(553 + 11\sqrt{79})}} \begin{bmatrix} 6 + \frac{3}{10}(3 + \sqrt{79}) \\ -3 - \frac{9}{10}(3 + \sqrt{79}) \\ 5 - \frac{5}{10}(3 + \sqrt{79}) \end{bmatrix},$$

so that the corresponding canonical variable is

$$\begin{aligned}u_1 &= \frac{5}{\sqrt{(15)}\sqrt{(553 + 11\sqrt{79})}} \left\{ \left[6 + \frac{3}{10}(3 + \sqrt{79}) \right] x_1 - \left[3 + \frac{9}{10}(3 + \sqrt{79}) \right] x_2 \right. \\ &\quad \left. + \left[5 - \frac{1}{2}(3 + \sqrt{79}) \right] x_3 \right\}.\end{aligned}$$

Now, from the formula $\alpha_2 = \frac{1}{\rho_2} \Sigma_{11}^{-1} \Sigma_{12} \beta_2$, we have

$$\begin{aligned} \alpha_2 &= \frac{\sqrt{45}}{\sqrt{(23 - \sqrt{79})}} \left(\frac{1}{15} \right) \begin{bmatrix} 6 & -3 \\ -3 & 9 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{\sqrt{45}}{15(\sqrt{23 - \sqrt{79}})} \begin{bmatrix} \frac{6}{7}(3 + \sqrt{79}) - 3 \\ -\frac{3}{7}(3 + \sqrt{79}) + 9 \\ \frac{5}{7}(3 + \sqrt{79}) + 5 \end{bmatrix}. \end{aligned}$$

Let us normalize this vector via the constraint $\alpha_2' \Sigma_{11} \alpha_2 = 1$ or

$$\alpha_2' \Sigma_{11} \alpha_2 = \frac{1}{\rho_2^2} \beta_2' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \beta_2 = \gamma_2^2,$$

say. Thus,

$$\begin{aligned} \gamma_2^2 &= \frac{45}{15(23 - \sqrt{79})} \left[\frac{1}{7}(3 + \sqrt{79}), 1 \right] \begin{bmatrix} 11 & 2 \\ 2 & 14 \end{bmatrix} \begin{bmatrix} \frac{1}{7}(3 + \sqrt{79}) \\ 1 \end{bmatrix} \\ &= \frac{45}{15(23 - \sqrt{79})} \left[\frac{1}{49}(1738 + 94\sqrt{79}) \right], \end{aligned}$$

and the normalized vector α_2 , denoted by $\alpha_{(2)}$, is

$$\alpha_{(2)} = \frac{\alpha_2}{|\gamma_2|} = \frac{7}{\sqrt{15}\sqrt{(1738 + 94\sqrt{79})}} \begin{bmatrix} \frac{6}{7}(3 + \sqrt{79}) - 3 \\ -\frac{3}{7}(3 + \sqrt{79}) + 9 \\ \frac{5}{7}(3 + \sqrt{79}) + 5 \end{bmatrix},$$

so that the second canonical variable is

$$\begin{aligned} u_2 &= \frac{7}{\sqrt{15}\sqrt{(1738 + 94\sqrt{79})}} \left\{ \left[\frac{6}{7}(3 + \sqrt{79}) - 3 \right] x_1 - \left[\frac{3}{7}(3 + \sqrt{79}) + 9 \right] x_2 \right. \\ &\quad \left. + \left[\frac{5}{7}(3 + \sqrt{79}) + 5 \right] x_3 \right\}. \end{aligned}$$

Hence, the canonical pairs are (u_1, v_1) , (u_2, v_2) where u_j is the best predictor of v_j and vice versa for $j = 1, 2$. The pair of canonical variables (u_2, v_2) has the second largest canonical correlation. It is easy to verify that $\text{Cov}(u_1, u_2) = 0$ and $\text{Cov}(v_1, v_2) = 0$.

10.3. Estimation of the Canonical Correlations and Canonical Variables

Consider a simple random sample of size n from a population designated by the $(p + q) \times 1$ real vector $\begin{pmatrix} X \\ Y \end{pmatrix}$. Let the (corrected) sample sum of products matrix be denoted by

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{11} \text{ is } p \times p, \quad S_{22} \text{ is } q \times q,$$

where S_{11} is the sample sum of products matrix corresponding to the sample from the subvector X , whose (i, j) th element is of the form $\sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$ with the matrix (x_{ik}) denoting a sample of size n from X , S_{22} is the sample sum of products matrix corresponding to the subvector Y and $\frac{1}{n}S_{12}$ is the sample covariance between X and Y . Thus, denoting the estimates by hats, the estimates of Σ_{11} , Σ_{22} and Σ_{12} are $\hat{\Sigma}_{11} = \frac{1}{n}S_{11}$, $\hat{\Sigma}_{22} = \frac{1}{n}S_{22}$ and $\hat{\Sigma}_{12} = \frac{1}{n}S_{12}$, respectively. These will also be the maximum likelihood estimates if we assume normality, that is, if

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{p+q}(\nu, \Sigma), \quad \Sigma > O, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad (10.3.1)$$

where $\Sigma_{11} = \text{Cov}(X) > O$, $\Sigma_{22} = \text{Cov}(Y) > O$ and $\Sigma_{12} = \text{Cov}(X, Y)$. For the estimates of these submatrices, equation (10.1.1) will take the following form:

$$\begin{vmatrix} -t\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & -t\hat{\Sigma}_{22} \end{vmatrix} = 0 \Rightarrow \begin{vmatrix} -tS_{11} & S_{12} \\ S_{21} & -tS_{22} \end{vmatrix} = 0 \quad (10.3.2)$$

where t is the sample canonical correlation; the reader may also refer to Mathai and Haubold (2017b). Letting $\hat{\rho} = t$ be the estimated canonical correlation, whenever $p \leq q$, t^2 is an eigenvalue of the sample canonical correlation matrix given by

$$\hat{\Sigma}_{11}^{-\frac{1}{2}} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-\frac{1}{2}} = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} = R_{11}^{-\frac{1}{2}} R_{12} R_{22}^{-1} R_{21} R_{11}^{-\frac{1}{2}}. \quad (10.3.3)$$

Note that we have chosen the symmetric format for the sample canonical correlation matrix. Observe that the sample size n is omitted from the middle expression in (10.3.3) as it gets canceled. As well, the middle expression is expressed in terms of sample correlation matrices in the last expression appearing in (10.3.3). The conversion from a sample covariance matrix to a sample correlation matrix has previously been explained. Letting $S = (s_{ij})$ denote the sample sum of products matrix, we can write

$$S = S_1 R S_1, \quad S_1 = \text{diag}(\sqrt{s_{11}}, \dots, \sqrt{s_{p+q, p+q}}), \quad R = (r_{ij}) = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix},$$

where r_{ij} is the (i, j) th sample correlation coefficient, and for example, R_{11} is the $p \times p$ submatrix within the $(p + q) \times (p + q)$ matrix R . We will examine the distribution of t^2 when the population covariance submatrix $\Sigma_{12} = O$, as well as when $\Sigma_{12} \neq O$, in the case of a $(p + q)$ -variate real Gaussian population as given in (10.3.1), after considering an example to illustrate the computations of the canonical correlation ρ resulting from (10.1.1) and presenting an iterative procedure.

Example 10.3.1. Let X and Y be two real bivariate vector random variables and $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$. Consider the following simple random sample of size 5 from Z :

$$Z_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, Z_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \end{bmatrix}, Z_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix}, Z_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, Z_5 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

Construct the sample pairs of canonical variables.

Solution 10.3.1. Let us use the standard notation. The sample matrix is $\mathbf{Z} = [Z_1, \dots, Z_5]$, the sample average $\bar{Z} = \frac{1}{5}[Z_1 + \dots + Z_5]$, the matrix of sample averages is $\bar{\mathbf{Z}} = [\bar{Z}, \dots, \bar{Z}]$, the deviation matrix is $\mathbf{Z}_d = [Z_1 - \bar{Z}, \dots, Z_5 - \bar{Z}]$ and the sample sum of products matrix is $S = \mathbf{Z}_d \mathbf{Z}'_d$. These quantities are the following:

$$\mathbf{Z} = \begin{bmatrix} 1 & 2 & 2 & 0 & 0 \\ 2 & 0 & 1 & 1 & 1 \\ 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & -1 & 0 & 1 \end{bmatrix}, \bar{\mathbf{Z}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{Z}_d = \begin{bmatrix} 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & -1 & 0 & 1 \end{bmatrix}, S = \begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 2 & 2 & -2 \\ -1 & 2 & 4 & -3 \\ -1 & -2 & -3 & 4 \end{bmatrix},$$

where, as per our notation,

$$S_{11} = \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}, S_{12} = \begin{bmatrix} -1 & -1 \\ 2 & -2 \end{bmatrix},$$

$$S_{21} = \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix}, S_{22} = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}.$$

We need to compute the following items:

$$\begin{aligned} S_{11}^{-1} &= \frac{1}{7} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}, \quad S_{22}^{-1} = \frac{1}{7} \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}, \\ S_{11}^{-1} S_{12} &= \frac{1}{7} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 2 & -2 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix}, \\ S_{22}^{-1} S_{21} &= \frac{1}{7} \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix}. \end{aligned}$$

The matrices A and B are then the following (using the same notation as for the population values for convenience):

$$\begin{aligned} A &= S_{11}^{-1} S_{12} S_{22}^{-1} S_{21} = \frac{1}{7^2} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix} = \frac{2}{7^2} \begin{bmatrix} 14 & 4 \\ 7 & 16 \end{bmatrix}, \\ B &= S_{22}^{-1} S_{21} S_{11}^{-1} S_{12} = \frac{1}{7^2} \begin{bmatrix} -7 & 2 \\ -7 & -2 \end{bmatrix} \begin{bmatrix} 0 & -4 \\ 7 & -9 \end{bmatrix} = \frac{2}{7^2} \begin{bmatrix} 7 & 5 \\ -7 & 23 \end{bmatrix}. \end{aligned}$$

If the population covariance matrix of Z is Σ , an estimate of Σ is $\frac{S}{n}$ where S is the sample sum of products matrix and n is the sample size, which is also the maximum likelihood estimate of Σ if Z is Gaussian distributed. Instead of using $\frac{S}{n}$, we will work with S since the normalized eigenvectors of S and $\frac{S}{n}$ are identical, although the eigenvalues of $\frac{S}{n}$ are $\frac{1}{n}$ times the eigenvalues of S .

The eigenvalues of A are $\frac{2}{7^2}$ times the solution of $(14 - \lambda)(16 - \lambda) - 28 = 0 \Rightarrow \lambda_1 = 15 + \sqrt{29}$, $\lambda_2 = 15 - \sqrt{29}$, so that the eigenvalues of A , denoted by λ_{11} and λ_{12} , are $\lambda_{11} = (\frac{2}{7^2})(15 + \sqrt{29})$, $\lambda_{12} = \frac{2}{7^2}(15 - \sqrt{29})$. The eigenvalues of B are $\frac{2}{7^2}$ times the solutions of $(7 - \nu)(23 - \nu) + 35 = 0 \Rightarrow \nu_1 = 15 + \sqrt{29}$, $\nu_2 = 15 - \sqrt{29}$, so that the eigenvalues of B , denoted by ν_{21} and ν_{22} , are $\nu_{21} = \frac{2}{7^2}(15 + \sqrt{29})$, $\nu_{22} = \frac{2}{7^2}(15 - \sqrt{29})$, which, as expected, are the same as those of A . Corresponding to λ_{11} , an eigenvector from A is available from the equation

$$\begin{bmatrix} 14 - (15 + \sqrt{29}) & 4 \\ 7 & 16 - (15 + \sqrt{29}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

deleting $\frac{2}{7^2}$ from both sides. Thus, one solution is

$$\alpha_1 = \begin{bmatrix} 4/[1 + \sqrt{29}] \\ 1 \end{bmatrix}.$$

Let us normalize this vector through the constraint $\alpha_1' S_{11} \alpha_1 = 1$. Since

$$\alpha_1' S_{11} \alpha_1 = \left[\frac{4}{1 + \sqrt{29}}, 1 \right] \begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{4}{1 + \sqrt{29}} \\ 1 \end{bmatrix} = \frac{2}{7^2} (116 - 11\sqrt{29}),$$

the normalized eigenvector, denoted by $\alpha_{(1)}$, and the corresponding sample canonical variable, denoted by u_1 , are

$$\alpha_{(1)} = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \begin{bmatrix} \frac{4}{1 + \sqrt{29}} \\ 1 \end{bmatrix} \text{ and } u_1 = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \left[\frac{4}{1 + \sqrt{29}} x_1 + x_2 \right].$$

The eigenvalues of B are also the same as $v_1 = 15 + \sqrt{29}$, $v_2 = 15 - \sqrt{29}$. Let us compute an eigenvector corresponding to the eigenvalue v_1 obtained from B . This eigenvector can be determined from the equation

$$\begin{bmatrix} -8 - \sqrt{29} & 5 \\ -7 & 8 - \sqrt{29} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives one solution as

$$\beta_1 = \begin{bmatrix} 5/(8 + \sqrt{29}) \\ 1 \end{bmatrix}.$$

Let us normalize under the constraint $\beta_1' S_{22} \beta_1 = 1$. Since

$$\beta_1' S_{22} \beta_1 = \left[\frac{5}{8 + \sqrt{29}}, 1 \right] \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} \frac{5}{8 + \sqrt{29}} \\ 1 \end{bmatrix} = \frac{2}{7^2} (116 - 11\sqrt{29}),$$

the normalized eigenvector, denoted by $\beta_{(1)}$, and the corresponding canonical variable denoted by v_1 , are the following:

$$\beta_{(1)} = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \begin{bmatrix} \frac{5}{8 + \sqrt{29}} \\ 1 \end{bmatrix} \text{ and } v_1 = \frac{7}{\sqrt{2}\sqrt{116 - 11\sqrt{29}}} \left[\frac{5}{8 + \sqrt{29}} y_1 + y_2 \right].$$

Therefore, one pair of canonical variables is (u_1, v_1) where u_1 is the best predictor of v_1 and vice versa. Now, consider $\lambda_2 = 15 - \sqrt{29}$ and $v_2 = 15 - \sqrt{29}$. Proceed as in the above case to obtain the second pair of canonical variables (u_2, v_2) .

10.3.1. An iterative procedure

Without any loss of generality, let $p \leq q$. We will illustrate the procedure for the population values for convenience. Consider the matrix A as previously defined, that is, $A = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$, and ρ , the canonical correlation which is a solution of (10.1.1) with $\rho^2 = \lambda$ where λ is an eigenvalue of A . When p is small, we may directly solve the determinantal equation (10.1.1) and evaluate the roots which are the canonical correlations. We are now illustrating the computations for the population values. When p is large, direct evaluation could prove tedious without resorting to computational software packages. In this case, the following iterative procedure may be employed. Let λ be an eigenvalue of A and α the corresponding eigenvector. We want to evaluate $\lambda = \rho^2$ and α , but we cannot solve (10.1.1) directly when p is large. In that case, take an initial trial vector γ_0 and normalize it via the constraint $\gamma_0' \Sigma_{11} \gamma_0 = 1$ so that $\alpha_0' \Sigma_{11} \alpha_0 = 1$, α_0 being the normalized γ_0 . Then, $\alpha_0 = \frac{1}{\sqrt{\gamma_0' \Sigma_{11} \gamma_0}} \gamma_0$. Now, consider the equation

$$A \alpha_0 = \gamma_1.$$

If α_0 happens to be the eigenvector $\alpha_{(1)}$ corresponding to the largest eigenvalue $\lambda_{(1)}$ of A then $A \alpha_0 = \lambda_{(1)} \alpha_{(1)}$; $A \alpha_0 = \gamma_1 \Rightarrow \gamma_1' \Sigma_{11} \gamma_1 = \lambda_{(1)}^2 \alpha_{(1)}' \Sigma_{11} \alpha_{(1)} = \lambda_{(1)}^2$ since $\alpha_{(1)}' \Sigma_{11} \alpha_{(1)} = 1$. Then $\rho_{(1)}^2 = \lambda_{(1)} = \sqrt{\gamma_1' \Sigma_{11} \gamma_1}$. This gives the motivation for the iterative procedure. Consider the equation

$$A \alpha_i = \gamma_{i+1}, \quad \alpha_i = \frac{1}{\sqrt{\gamma_i' \Sigma_{11} \gamma_i}} \gamma_i, \quad i = 0, 1, \dots \quad (i)$$

Continue the iteration process. At each stage compute $\delta_i = \alpha_i' \Sigma_{11} \alpha_i$ while ensuring that δ_i is increasing. Halt the iteration when $\gamma_j = \gamma_{j-1}$ approximately, that is, when $\alpha_j = \alpha_{j-1}$ approximately, which indicates that γ_j converges to some vector γ . At this stage, the normalized γ is $\alpha_{(1)}$, the eigenvector corresponding to the largest eigenvalue $\lambda_{(1)}$ of A . Then, the largest eigenvalue $\lambda_{(1)}$ of A is given by $\lambda_{(1)} = \sqrt{\gamma' \Sigma_{11} \gamma}$. Thus, as a result of the iteration process specified by equation (i),

$$\lim_{j \rightarrow \infty} \alpha_j = \alpha_{(1)} \text{ and } + \sqrt{\lim_{j \rightarrow \infty} \gamma_j' \Sigma_{11} \gamma_j} = \lambda_{(1)}. \quad (ii)$$

These initial iterations produce the largest eigenvalue $\lambda_{(1)} = \rho_{(1)}^2$ and the corresponding eigenvector $\alpha_{(1)}$. From (10.1.1), we have

$$\Sigma_{22}^{-1} \Sigma_{21} \alpha = \rho \beta \Rightarrow \frac{1}{\rho} \Sigma_{22}^{-1} \Sigma_{21} \alpha = \beta. \quad (iii)$$

Substitute the computed $\rho_{(1)}$ and $\alpha_{(1)}$ in (iii) to obtain $\beta_{(1)}$, the eigenvector corresponding to the largest eigenvalue $\lambda_{(1)}$ of $B = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$. This completes the first stage of the iteration process. Now, consider $A_2 = A - \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)}$. Observe that $\alpha_{(1)} \alpha'_{(1)}$ is a $p \times p$ matrix. In general, we can express a symmetric matrix in terms of its eigenvalues and normalized eigenvectors as follows:

$$A = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} = \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)} + \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)} + \cdots + \lambda_{(p)} \alpha_{(p)} \alpha'_{(p)}, \quad (iv)$$

as explained in Chapter 1 or Mathai and Haubold (2017a). Carry out the second stage of the iteration process on A_2 as indicated in (i). This will produce the second largest eigenvalue $\lambda_{(2)}$ of A and the corresponding eigenvector $\alpha_{(2)}$. Then, compute the corresponding $\beta_{(2)}$ via the procedure given in (iii). This will complete the second stage. For the next stage, consider

$$A_3 = A_2 - \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)} = A - \lambda_{(1)} \alpha_{(1)} \alpha'_{(1)} - \lambda_{(2)} \alpha_{(2)} \alpha'_{(2)}$$

and perform the iterative steps (i) to (iv). This will produce $\lambda_{(3)}$, $\alpha_{(3)}$ and $\beta_{(3)}$. Keep on iterating until all the p eigenvalues $\lambda_{(1)}, \dots, \lambda_{(p)}$ of A as well as $\alpha_{(j)}$ and $\beta_{(j)}$, the corresponding eigenvectors of A and B are obtained for $j = 1, \dots, p$.

In the case of sample eigenvalues and eigenvectors, start with the sample matrices

$$\begin{aligned} \hat{A} &= \hat{\Sigma}_{11}^{-\frac{1}{2}} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-\frac{1}{2}} = R_{11}^{-\frac{1}{2}} R_{12} R_{22}^{-1} R_{21} R_{11}^{-\frac{1}{2}} \\ \hat{B} &= \hat{\Sigma}_{22}^{-\frac{1}{2}} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-\frac{1}{2}} = R_{22}^{-\frac{1}{2}} R_{21} R_{11}^{-1} R_{12} R_{22}^{-\frac{1}{2}}. \end{aligned}$$

Carry out the iteration steps (i) to (iv) on \hat{A} to obtain the sample eigenvalues, denoted by $\hat{\rho}_{(j)} = t_{(j)}$, $j = 1, \dots, p$ for $p \leq q$, where $t_{(j)}$ is the j -th sample canonical correlation, and the corresponding eigenvectors of \hat{A} denoted by $a_{(j)}$ as well as those of \hat{B} denoted by $b_{(j)}$.

Example 10.3.2. Consider the real vectors $X' = (x_1, x_2)$, $Y' = (y_1, y_2, y_3)$, and let $Z' = (X', Y')$ where x_1, x_2, y_1, y_2, y_3 are real scalar random variables. Let the covariance matrix of Z be $\Sigma > O$ where

$$\begin{aligned} \text{Cov}(Z) &= \text{Cov} \begin{bmatrix} X \\ Y \end{bmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma_{12} = \Sigma'_{21}, \\ \text{Cov}(X) &= \Sigma_{11}, \quad \text{Cov}(Y) = \Sigma_{22}, \quad \text{Cov}(X, Y) = \Sigma_{12}, \end{aligned}$$

with

$$\Sigma_{11} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \Sigma_{12} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \quad \Sigma_{22} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Consider the problem of predicting X from Y and vice versa. Obtain the best predictors by constructing pairs of canonical variables.

Solution 10.3.2. Let us first compute the inverses Σ_{11}^{-1} , Σ_{22}^{-1} and $\Sigma_{11}^{-1}\Sigma_{12}$, $\Sigma_{22}^{-1}\Sigma_{21}$. We are taking the non-symmetric form of A as the symmetric form requires more calculations. Either way, the eigenvalues are identical. On directly applying the formula $C^{-1} = \frac{1}{|C|} \times [\text{the matrix of cofactors}]'$, we have

$$\begin{aligned}\Sigma_{11}^{-1} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \Sigma_{22}^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \Sigma_{11}^{-1}\Sigma_{12} &= \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix}, \\ \Sigma_{22}^{-1}\Sigma_{21} &= \begin{bmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}.\end{aligned}$$

Thus, the non-symmetric forms of A and B are

$$\begin{aligned}A &= \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \frac{1}{3} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -2 \\ 2 & 10 \end{bmatrix}, \\ B &= \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \frac{1}{3} \begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 1 & -3 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 4 & -6 & 4 \\ -2 & 6 & -2 \\ 2 & 0 & 2 \end{bmatrix}.\end{aligned}$$

Let us compute the eigenvalues of $3A$. Consider $\begin{vmatrix} 2-\lambda & -2 \\ 2 & 10-\lambda \end{vmatrix} = 0 \Rightarrow (2-\lambda)(10-\lambda) + 4 = 0$, which gives

$$\lambda = \frac{12 \pm \sqrt{(12)^2 - 4(24)}}{2} = 6 + 2\sqrt{3}, 6 - 2\sqrt{3},$$

the eigenvalues of A being $\lambda_{(1)} = 2 + \frac{2}{3}\sqrt{3}$, $\lambda_{(2)} = 2 - \frac{2}{3}\sqrt{3}$. These are the squares of the canonical correlation coefficient ρ resulting from (10.1.1). Let us determine the eigenvectors corresponding to $\lambda_{(1)}$ and $\lambda_{(2)}$. Our notations for the linear functions of X and Y are $u = \alpha'X$ and $v = \beta'Y$; in this case, $\alpha' = (\alpha_1, \alpha_2)$ and $\beta' = (\beta_1, \beta_2, \beta_3)$. Then, the eigenvector α , corresponding to $\lambda_{(1)}$ is obtained from the equation

$$\begin{bmatrix} \frac{2}{3} - (2 + \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & \frac{10}{3} - (2 + \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \quad (i)$$

$$-(2 + \sqrt{3})\alpha_1 - \alpha_2 = 0 \Rightarrow \alpha_1 = 1, \alpha_2 = -(2 + \sqrt{3}).$$

Observe that since (i) is a singular system of linear equations, we need only consider one equation and we can preassign a value for α_1 or α_2 . Taking $\alpha_1 = 1$, let us normalize the resulting vector via the constraint $\alpha' \Sigma_{11} \alpha = 1$. Since

$$\begin{aligned} \alpha' \Sigma_{11} \alpha &= [1 \quad -(2 + \sqrt{3})] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -(2 + \sqrt{3}) \end{bmatrix} \\ &= 12 + 6\sqrt{3} \equiv \gamma_1, \end{aligned}$$

the normalized α , denoted by $\alpha_{(1)}$, is

$$\alpha_{(1)} = \frac{1}{\sqrt{\gamma_1}} \begin{bmatrix} 1 \\ -(2 + \sqrt{3}) \end{bmatrix} \Rightarrow u_1 = \frac{1}{\sqrt{\gamma_1}} [x_1 - (2 + \sqrt{3})x_2]. \quad (ii)$$

Now, the eigenvector corresponding to the second eigenvalue $\lambda_{(2)}$ is such that

$$\begin{bmatrix} \frac{2}{3} - (2 - \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & \frac{10}{3} - (2 - \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$(-2 + \sqrt{3})\alpha_1 - \alpha_2 = 0 \Rightarrow \alpha_1 = 1, \alpha_2 = -2 + \sqrt{3}.$$

Since

$$\begin{aligned} \alpha' \Sigma_{11} \alpha &= [1 \quad -2 + \sqrt{3}] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 + \sqrt{3} \end{bmatrix} \\ &= 12 - 6\sqrt{3} \equiv \gamma_2, \end{aligned}$$

the normalized α such that $\alpha' \Sigma_{11} \alpha = 1$ is

$$\alpha_{(2)} = \frac{1}{\sqrt{\gamma_2}} \begin{bmatrix} 1 \\ -2 + \sqrt{3} \end{bmatrix} \Rightarrow u_2 = \frac{1}{\sqrt{\gamma_2}} [x_1 + (-2 + \sqrt{3})x_2]. \quad (iii)$$

Observe that computing the eigenvalues of B from the equation $|B - \lambda_{(j)}I| = 0$ will be difficult. However, we know that they are $\lambda_{(1)}$ and $\lambda_{(2)}$ as given above, the third one being equal to zero. So, let us first verify that $3\lambda_{(1)}$ is an eigenvalue of $3B$. Consider

$$\begin{aligned}
|3B - (6 + 2\sqrt{3})I| &= \begin{vmatrix} 4 - (6 + 2\sqrt{3}) & -6 & 4 \\ -2 & 6 - (6 + 2\sqrt{3}) & -2 \\ 2 & 0 & 2 - (6 + 2\sqrt{3}) \end{vmatrix} \\
&= 2 \times 2 \times 2 \begin{vmatrix} 1 & \sqrt{3} & 1 \\ -(1 + \sqrt{3}) & -3 & 2 \\ 1 & 0 & -(2 + \sqrt{3}) \end{vmatrix} \\
&= 8 \begin{vmatrix} 1 & \sqrt{3} & 1 \\ 0 & \sqrt{3} & 3 + \sqrt{3} \\ 0 & -\sqrt{3} & -(3 + \sqrt{3}) \end{vmatrix} = 0.
\end{aligned}$$

The operations performed are the following: Taking out 2 from each row; interchanging the second and the first rows; adding $(1 + \sqrt{3})$ times the first row to the second row and adding minus one times the first row to the third row. Similarly, it can be verified that $\lambda_{(2)}$ is also an eigenvalue of B . Moreover, since the third row of B is equal to the sum of its first two rows, B is singular, which means that the remaining eigenvalue must be zero. In Example 10.2.1, we made use of the formula resulting from (ii) of Sect. 10.1 for determining the second set of canonical variables. In this case, they will be directly computed from B to illustrate a different approach. Let us now determine the eigenvectors with respect to B , corresponding to $\lambda_{(1)}$ and $\lambda_{(2)}$:

$$B = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \frac{1}{3} \begin{bmatrix} 4 & -6 & 4 \\ -2 & 6 & -2 \\ 2 & 0 & 2 \end{bmatrix};$$

$$(B - \lambda_{(1)}I)\beta = O$$

$$\begin{aligned}
\Rightarrow & \begin{bmatrix} \frac{4}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) & -\frac{6}{3} & \frac{4}{3} \\ -\frac{2}{3} & \frac{6}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & 0 & \frac{2}{3} - (\frac{6}{3} + \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
\Rightarrow & \begin{bmatrix} -(1 + \sqrt{3}) & -3 & 2 \\ -1 & -\sqrt{3} & -1 \\ 1 & 0 & -(2 + \sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\end{aligned}$$

This yields the equations

$$-(1 + \sqrt{3})\beta_1 - 3\beta_2 + 2\beta_3 = 0 \tag{iv}$$

$$-\beta_1 - \sqrt{3}\beta_2 - \beta_3 = 0 \tag{v}$$

$$\beta_1 - (2 + \sqrt{3})\beta_3 = 0, \tag{vi}$$

whose solution in terms of an arbitrary β_3 is $\beta_2 = -(1 + \sqrt{3})\beta_3$ and $\beta_1 = (2 + \sqrt{3})\beta_3$. Taking $\beta_3 = 1$, we have the solution, $\beta_3 = 1$, $\beta_2 = -(1 + \sqrt{3})$, $\beta_1 = (2 + \sqrt{3})$. Let us normalize the resulting vector via the constraint $\beta' \Sigma_{22} \beta = 1$:

$$\begin{aligned} \beta' \Sigma_{22} \beta &= \begin{bmatrix} 2 + \sqrt{3} & -(1 + \sqrt{3}) & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \\ &= 6 + 2\sqrt{3} = \delta_1. \end{aligned}$$

Thus, the normalized β , denoted by $\beta_{(1)}$, is

$$\begin{aligned} \beta_{(1)} &= \frac{1}{\sqrt{\delta_1}} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \Rightarrow \\ v_1 &= \frac{1}{\sqrt{\delta_1}} [(2 + \sqrt{3})y_1 - (1 + \sqrt{3})y_2 + y_3]. \end{aligned} \tag{vii}$$

Observe that we could also have utilized (iii) of Sect. 10.3.1 to evaluate $\beta_{(1)}$ and $\beta_{(2)}$ from $\alpha_{(1)}$ and $\alpha_{(2)}$. Consider the second eigenvalue $\lambda_{(2)} = \frac{6}{3} - \frac{2}{3}\sqrt{3}$ and the equation $(B - \lambda_{(2)}I)\beta = O$, that is,

$$\begin{aligned} \begin{bmatrix} \frac{4}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) & -\frac{6}{3} & \frac{4}{3} \\ -\frac{2}{3} & \frac{6}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) & -\frac{2}{3} \\ \frac{2}{3} & 0 & \frac{2}{3} - (\frac{6}{3} - \frac{2}{3}\sqrt{3}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} -1 + \sqrt{3} & -3 & 2 \\ -1 & \sqrt{3} & -1 \\ 1 & 0 & -2 + \sqrt{3} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \end{aligned}$$

which leads to the equations

$$(-1 + \sqrt{3})\beta_1 - 3\beta_2 + 2\beta_3 = 0 \quad (\text{viii})$$

$$-\beta_1 + \sqrt{3}\beta_2 - \beta_3 = 0 \quad (\text{ix})$$

$$\beta_1 + (-2 + \sqrt{3})\beta_3 = 0. \quad (\text{x})$$

Thus, when $\beta_3 = 1$, $\beta_1 = 2 - \sqrt{3}$ and $\beta_2 = \sqrt{3} - 1$. Subject to the constraint $\beta' \Sigma_{22} \beta = 1$, we have

$$\begin{aligned} \beta' \Sigma_{22} \beta &= [(2 - \sqrt{3}) \quad (\sqrt{3} - 1) \quad 1] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 - \sqrt{3} \\ \sqrt{3} - 1 \\ 1 \end{bmatrix} \\ &= 6 - 2\sqrt{3} = \delta_2. \end{aligned}$$

Hence, the normalized β is

$$\begin{aligned} \beta_{(2)} &= \frac{1}{\sqrt{\delta_2}} \begin{bmatrix} 2 - \sqrt{3} \\ \sqrt{3} - 1 \\ 1 \end{bmatrix} \Rightarrow \\ v_2 &= \frac{1}{\sqrt{\delta_2}} [(2 - \sqrt{3})y_1 + (\sqrt{3} - 1)y_2 + y_3]. \quad (\text{xi}) \end{aligned}$$

The reader may also verify that this solution for $\beta_{(2)}$ is identical to that coming from (iii) of Sect. 10.3.1. Thus, the canonical pairs are the following: From (ii) and (vii), we have the first canonical pair (u_1, v_1) , the second pair (u_2, v_2) resulting from (iii) and (xi). This means u_1 is the best predictor of v_1 and vice versa, and that u_2 is the second best predictor of v_2 and vice versa.

Let us ensure that no computational errors have been committed. Consider

$$\begin{aligned} \alpha'_{(1)} \Sigma_{12} \beta_{(1)} &= \frac{1}{\sqrt{\gamma_1 \delta_1}} [1, -(2 + \sqrt{3})] \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 + \sqrt{3} \\ -(1 + \sqrt{3}) \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{\gamma_1 \delta_1}} 4(3 + 2\sqrt{3}), \end{aligned}$$

with

$$\gamma_1 \delta_1 = 6(2 + \sqrt{3})2(3 + \sqrt{3}) = 12(9 + 5\sqrt{3}),$$

so that

$$\begin{aligned} \frac{[\alpha'_{(1)} \Sigma_{12} \beta_{(1)}]^2}{\gamma_1 \delta_1} &= \frac{16(3 + 2\sqrt{3})^2}{12(9 + 5\sqrt{3})} \\ &= \frac{16(21 + 12\sqrt{3})}{12(9 + 5\sqrt{3})} = \frac{4(7 + 4\sqrt{3})}{9 + 3\sqrt{3}} = \frac{4(7 + 4\sqrt{3})(9 - 5\sqrt{3})}{6} \\ &= \frac{2}{3}(3 + \sqrt{3}) = 2 + \frac{2}{\sqrt{3}} = 2 + \frac{2}{3}\sqrt{3} = \lambda_{(1)} : \text{the largest eigenvalue of } A, \end{aligned}$$

which corroborates the results obtained for $\alpha_{(1)}$, $\beta_{(1)}$ and $\lambda_{(1)}$. Similarly, it can be verified that $\alpha_{(2)}$, $\beta_{(2)}$ and $\lambda_{(2)}$ have been correctly computed.

10.4. The Sampling Distribution of the Canonical Correlation Matrix

Consider a simple random sample of size n from $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$. Let the $(p + q) \times (p + q)$ sample sum of products matrix be denoted by S and let Z have a real $(p + q)$ -variate standard Gaussian density. Then S has a real $(p + q)$ -variate Wishart distribution with the identity matrix as its parameter matrix and $m = n - 1$ degrees of freedom, n being the sample size. Letting the density of S be denoted by $f(S)$,

$$f(S) = \frac{|S|^{\frac{m}{2} - \frac{p+q-1}{2}}}{2^{\frac{m(p+q)}{2}} \Gamma_{p+q}(\frac{m}{2})} e^{-\frac{1}{2}\text{tr}(S)}, \quad S > O, \quad m \geq p + q. \quad (10.4.1)$$

Let us partition S as follows:

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{11} \text{ is } p \times p, \quad S_{22} \text{ is } q \times q,$$

and let $dS = dS_{11} \wedge dS_{22} \wedge dS_{12}$. Note that $\text{tr}(S) = \text{tr}(S_{11}) + \text{tr}(S_{22})$ and

$$\begin{aligned} |S| &= |S_{22}| |S_{11} - S_{12} S_{22}^{-1} S_{21}| \\ &= |S_{22}| |S_{11}| |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}|. \end{aligned}$$

Letting $U = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}}$, $dU = |S_{11}|^{-\frac{q}{2}} |S_{22}|^{-\frac{p}{2}} dS_{12}$ for fixed S_{11} and S_{22} , so that the joint density of S_{11} , S_{22} and S_{12} is given by

$$\begin{aligned} f_1(S) dS_{11} \wedge dS_{22} \wedge dS_{12} &= \frac{|S_{11}|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(S_{11})}}{2^{\frac{mp}{2}} \Gamma_p(\frac{m}{2})} dS_{11} \\ &\times \frac{|S_{22}|^{\frac{m}{2} - \frac{q+1}{2}} e^{-\frac{1}{2}\text{tr}(S_{22})}}{2^{\frac{mq}{2}} \Gamma_q(\frac{m}{2})} dS_{22} \\ &\times \frac{\Gamma_p(\frac{m}{2}) \Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} |I - UU'|^{\frac{m}{2} - \frac{p+q-1}{2}} dU. \end{aligned} \quad (10.4.2)$$

It is seen from (10.4.2) that S_{11} , S_{22} and U are mutually independently distributed, and so are S_{11} , S_{22} and $W = UU'$. Further, $S_{11} \sim W_p(m, I)$ and $S_{22} \sim W_q(m, I)$. Note that $W = UU' = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}$ is the sample canonical correlation matrix. It follows from Theorem 4.2.3 of Chapter 4, that for $p \leq q$ and U of full rank p ,

$$dU = \frac{\pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} dW. \quad (10.4.3)$$

After integrating out S_{11} and S_{22} from (10.4.2) and substituting for dS_{12} , we obtain the following representation of the density of W :

$$f_2(W) = \frac{\pi^{\frac{pq}{2}}}{\Gamma_p(\frac{q}{2})} \frac{\Gamma_p(\frac{m}{2}) \Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} |I - W|^{\frac{m-q}{2} - \frac{p+1}{2}},$$

where

$$\frac{\Gamma_q(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})} = \frac{\pi^{\frac{q(q-1)}{4}}}{\pi^{\frac{(p+q)(p+q-1)}{4}}} \frac{\Gamma(\frac{m}{2}) \cdots \Gamma(\frac{m}{2} - \frac{q-1}{2})}{\Gamma(\frac{m}{2}) \cdots \Gamma(\frac{m}{2} - \frac{p+q-1}{2})} = \frac{1}{\pi^{\frac{pq}{2}} \Gamma_p(\frac{m-q}{2})}.$$

Hence, the density of W is

$$f_2(W) = \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2}) \Gamma_p(\frac{m-q}{2})} |W|^{\frac{q}{2} - \frac{p+1}{2}} |I - W|^{\frac{m-q}{2} - \frac{p+1}{2}}. \quad (10.4.4)$$

Thus, the following result:

Theorem 10.4.1. *Let Z , S , S_{11} , S_{22} , S_{12} , U and W be as defined above. Then, for $p \leq q$ and U of full rank p , the $p \times p$ canonical correlation matrix $W = UU'$ has the real matrix-variate type-I beta density with the parameters $(\frac{q}{2}, \frac{m-q}{2})$ that is specified in (10.4.4) with $m \geq p + q$, $m = n - 1$.*

When $q \leq p$ and S_{21} is of full rank q , the canonical correlation matrix $W_1 = U'U = S_{22}^{-\frac{1}{2}} S_{21} S_{11}^{-1} S_{12} S_{22}^{-\frac{1}{2}}$ will have the density given in (10.4.4) with p and q interchanged. Suppose that $p \leq q$ and we would like to consider the density of $W_1 = U'U$. In this case $U'U$ is real positive semi-definite as the rank of U is $p \leq q$. However, on expanding the following determinant in two different ways:

$$\begin{vmatrix} I_p & U \\ U' & I_q \end{vmatrix} = |I_p - UU'| = |I_q - U'U|,$$

it follows from (10.4.2) that the $q \times q$ matrix $U'U$ has a distribution that is equivalent to that of the $p \times p$ matrix UU' , as given in (10.4.4). The distribution of the sample canonical correlation matrix has been derived in Mathai (1981) for a Gaussian population under the assumption that $\Sigma_{12} \neq O$.

10.4.1. The joint density of the eigenvalues and eigenvectors

Without any loss of generality, let $p \leq q$ and U be of full rank p . Let W denote the sample canonical correlation matrix whose density is as given in (10.4.4) for the case when the population canonical matrix is a null matrix. Let the eigenvalues of W be distinct and such that $1 > v_1 > v_2 > \dots > v_p > 0$. Observe that $v_j = r_{(j)}^2$ where $r_{(j)}$, $j = 1, \dots, p$ are the sample canonical correlations. For a unique $p \times p$ orthonormal matrix Q , $QQ' = I$, $Q'Q = I$, we have $Q'WQ = \text{diag}(v_1, \dots, v_p) \equiv D$. Consider the transformation from W to D and the normalized eigenvectors of W , which constitute the columns of Q . Then, as is explained in Theorem 8.2.1 or Theorem 4.4 of Mathai (1997),

$$dW = \left[\prod_{i < j} (v_i - v_j) \right] dD \wedge h(Q) \tag{10.4.5}$$

where $h(Q) = \wedge[(dQ)Q']$ is the differential element associated with Q , and we have the following result:

Theorem 10.4.2. *The joint density of the distinct eigenvalues $1 > v_1 > v_2 > \dots > v_p > 0$, $p \leq q$, of $W = UU'$ whose density is specified in (10.4.4), U being assumed to be of full rank p , and the normalized eigenvectors corresponding to v_1, \dots, v_p , denoted by $f_3(D, Q)$, is the following:*

$$\begin{aligned} f_3(D, Q)dD \wedge h(Q) &= \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2})\Gamma_p(\frac{m-q}{2})} \left[\prod_{j=1}^p v_j^{\frac{q-p+1}{2}} \right] \left[\prod_{j=1}^p (1 - v_j)^{\frac{m-q}{2} - \frac{p+1}{2}} \right] \\ &\quad \times \left[\prod_{i < j} (v_i - v_j) \right] dD \wedge h(Q) \end{aligned} \tag{10.4.6}$$

where $h(Q)$ is as defined in (10.4.5). To obtain the joint density of the squares of the sample canonical correlations $r_{(j)}^2$ and the corresponding canonical vectors, it suffices to replace v_j by $r_{(j)}^2$, $1 > r_{(1)}^2 > \cdots > r_{(p)}^2 > 0$, $-1 < r_{(j)} < 1$, $j = 1, \dots, p$.

The joint density of the eigenvalues can be determined by integrating out $h(Q)$ from (10.4.6) in this real case. It follows from Theorem 4.2.2 that

$$\int_{O_p} h(Q) = \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})}, \quad (10.4.7)$$

this result being also stated in Mathai (1997). For the complex case, the expression on the right-hand side of (10.4.7) is $\pi^{p(p-1)}/\tilde{\Gamma}_p(p)$. Hence, the joint density of the eigenvalues or, equivalently, the density of D and the density of Q are the following:

Theorem 10.4.3. *When $p \leq q$ and U is of full rank p , the joint density of the distinct eigenvalues $1 > v_1 > \cdots > v_p > 0$ of the canonical correlation matrix W in (10.4.4), which is available from (10.4.6) and denoted by $f_4(D)$, is*

$$\begin{aligned} f_4(D) &= \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{q}{2})\Gamma_p(\frac{m-q}{2})} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p(\frac{p}{2})} \left[\prod_{j=1}^p v_j^{\frac{q}{2} - \frac{p+1}{2}} \right] \\ &\times \left[\prod_{j=1}^p (1 - v_j)^{\frac{m-q}{2} - \frac{p+1}{2}} \right] \left[\prod_{i < j} (v_i - v_j) \right], \end{aligned} \quad (10.4.8)$$

and the joint density of the normalized eigenvectors associated with W , denoted by $f_5(Q)$, is given by

$$f_5(Q) = \frac{\Gamma_p(\frac{p}{2})}{\pi^{\frac{p^2}{2}}} h(Q) \quad (10.4.9)$$

where $h(Q)$ is as defined in (10.4.5).

To obtain the joint density of the squares of the sample canonical correlations $r_{(j)}^2$, one should replace v_j by $r_{(j)}^2$, $j = 1, 2, \dots, p$, in (10.4.8).

Example 10.4.1. Verify that (10.4.8) is a density for $p = 2$, $m - q = p + 1$, with q being a free parameter.

Solution 10.4.1. For $m - q = p + 1$, $p = 2$, the right-hand side of (10.4.8) becomes

$$\begin{aligned} & \frac{\Gamma_p\left(\frac{q+p+1}{2}\right)}{\Gamma_p\left(\frac{q}{2}\right)\Gamma_p\left(\frac{p+1}{2}\right)} \frac{\pi^{\frac{p^2}{2}}}{\Gamma_p\left(\frac{p}{2}\right)} (v_1 v_2)^{\frac{q-(p+1)}{2}} (v_1 - v_2) \\ &= \frac{\Gamma_2\left(\frac{q+3}{2}\right)}{\Gamma_2\left(\frac{q}{2}\right)\Gamma_2\left(\frac{p+1}{2}\right)} \frac{\pi^2}{\Gamma_2(1)} (v_1 v_2)^{\frac{m-3}{2}} (v_1 - v_2). \end{aligned} \quad (i)$$

The constant part simplifies as follows:

$$\frac{\Gamma_2\left(\frac{q+3}{2}\right)}{\Gamma_2\left(\frac{q}{2}\right)\Gamma_2\left(\frac{p+1}{2}\right)} \frac{\pi^2}{\Gamma_2(1)} = \frac{\Gamma\left(\frac{q+3}{2}\right)\Gamma\left(\frac{q+2}{2}\right)}{\sqrt{\pi}\left[\Gamma\left(\frac{q}{2}\right)\Gamma\left(\frac{q-1}{2}\right)\right]\left[\Gamma\left(\frac{3}{2}\right)\Gamma(1)\right]} \frac{\pi^2}{\sqrt{\pi}\Gamma(1)\Gamma\left(\frac{1}{2}\right)}; \quad (ii)$$

now, noting that

$$\Gamma\left(\frac{q+3}{2}\right) = \left(\frac{q+1}{2}\right)\left(\frac{q-1}{2}\right)\Gamma\left(\frac{q-1}{2}\right) \text{ and } \Gamma\left(\frac{q+2}{2}\right) = \frac{q}{2}\Gamma\left(\frac{q}{2}\right),$$

and substituting these values in (ii), the constant part becomes

$$\frac{\left(\frac{q+1}{2}\right)\left(\frac{q-1}{2}\right)\left(\frac{q}{2}\right)}{\sqrt{\pi}\left(\frac{1}{2}\right)\sqrt{\pi}} \frac{\pi^2}{\sqrt{\pi}\sqrt{\pi}} = \frac{(q-1)q(q+1)}{4}. \quad (iii)$$

Let us show that the total integral equals 1. The integral part is the following:

$$\begin{aligned} & \int_{v_1=0}^1 \int_{v_2=0}^{v_1} (v_1 v_2)^{\frac{q-3}{2}} (v_1 - v_2) dv_1 \wedge dv_2 \\ &= \int_0^1 v_1^{\frac{q-1}{2}} \left[\int_{v_2=0}^{v_1} v_2^{\frac{q-3}{2}} dv_2 \right] dv_1 - \int_0^1 v_1^{\frac{q-3}{2}} \left[\int_{v_2=0}^{v_1} v_2^{\frac{q-1}{2}} dv_2 \right] dv_1 \\ &= \int_0^1 \frac{v_1^{q-1}}{\left(\frac{q-1}{2}\right)} dv_1 - \int_0^1 \frac{v_1^{q-1}}{\left(\frac{q+1}{2}\right)} dv_1 = \frac{1}{q\left(\frac{q-1}{2}\right)} - \frac{1}{q\left(\frac{q+1}{2}\right)} \\ &= \frac{4}{(q-1)q(q+1)}. \end{aligned} \quad (iv)$$

The product of (iii) and (iv) being equal to 1, this verifies that (10.4.8) is a density for $m - q = p + 1$, $p = 2$. This completes the computations.

10.4.2. Testing whether the population canonical correlations equal zero

In its symmetric form, whenever $\Sigma_{12} = O$, the population canonical correlation matrix is a null matrix, that is, $\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}} = O$. Thus, when $\Sigma_{12} = O$, the canonical correlations are equal to zero and vice versa. As was explained in Sect. 6.8.2, we have a one-to-one function of u_4 , the likelihood ratio criterion for testing this hypothesis in the case of a Gaussian distributed population. It was established that

$$u_4 = \frac{|S|}{|S_{11}| |S_{22}|} = |I - S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}}| = |I - UU'| = |I - W| = \prod_{j=1}^p (1 - r_{(j)}^2) \quad (10.4.10)$$

where $r_{(j)}$, $j = 1, \dots, p$, are the sample canonical correlations. It can also be seen from (10.4.2) that, when U is of full rank p , U has a rectangular matrix-variate type-1 beta distribution and $W = UU'$ has a real matrix-variate type-1 beta distribution. Since it has been determined in Sect. 6.8.2, that under H_o , the h -th moment of u_4 for an arbitrary h is given by

$$E[u_4^h | H_o] = c \frac{\prod_{j=p+1}^{p+q} \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}{\prod_{j=1}^q \Gamma(\frac{m}{2} - \frac{j-1}{2} + h)}, \quad m = n - 1, \quad (10.4.11)$$

where n is the sample size and c is such that $E[u_4^0 | H_o] = 1$, the density of u_4 is expressible in terms of a G-function. It was also shown in the same section that $-n \ln u_4$ is asymptotically distributed as a real chisquare random variable having $\frac{(p+q)(p+q-1)}{2} - \frac{p(p-1)}{2} - \frac{q(q-1)}{2} = pq$ degrees of freedom, which corresponds to the number of parameters restricted by the hypothesis $\Sigma_{12} = O$ since there are pq free parameters in Σ_{12} . Thus, the following result:

Theorem 10.4.4. *Consider the hypothesis $H_o: \rho_{(1)} = \dots = \rho_{(p)} = 0$, that is, the population canonical correlations $\rho_{(j)}$, $j = 1, \dots, p$, are all equal to zero, which is equivalent to the hypothesis $H_o: \Sigma_{12} = O$. Let u_4 denote the $(2/n)$ -th root of the likelihood ratio criterion for testing this hypothesis. Then, as the sample size $n \rightarrow \infty$, under H_o ,*

$$-n \ln u_4 = -2 \ln(\text{the likelihood ratio criterion}) \rightarrow \chi_{pq}^2, \quad (10.4.12)$$

χ_v^2 denoting a real chisquare random variable having v degrees of freedom.

An illustrative numerical example has already been presented in Chap. 6.

Note 10.1. We have initially assumed that $\Sigma > O$, $\Sigma_{11} > O$ and $\Sigma_{22} > O$. However, $\Sigma_{12} = \Sigma'_{21}$ may or may not be of full rank or some of its elements could be equal to zero. Note that $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ is either positive definite or positive semi-definite. Whenever $p \leq q$ and Σ_{12} is of rank p , $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} > O$ and, in this instance, all the p canonical correlations are positive. If Σ_{12} is not of full rank, then some of the eigenvalues of W as previously defined, as well as the corresponding canonical correlations will be equal to zero and, in the event that $q \leq p$, similar statements would hold with respect to Σ_{21} , W' and the resulting canonical correlations. This aspect will not be further investigated from an inferential standpoint.

Note 10.2. Consider the regression of X on Y , that is, $E[X|Y]$, when $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ has the following real $(p + q)$ -variate normal distribution:

$$Z \sim N_{p+q}(\mu, \Sigma), \Sigma > O, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

$$\Sigma_{11} = \text{Cov}(X) > O \text{ is } p \times p, \Sigma_{22} = \text{Cov}(Y) > O \text{ is } q \times q.$$

Then, from equation (3.3.5), we have

$$E[X|Y] = \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(Y - \mu_{(2)})$$

where $\mu' = (\mu'_{(1)}, \mu'_{(2)})$ and

$$\text{Cov}(X|Y) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Regression analysis is performed on the conditional space where Y is either composed of non-random real scalar variables or given values of real scalar random variables, whereas canonical correlation analysis is carried out in the entire space of Z . Clearly, these techniques involve distinct approaches. When Y is given values of random variables, then Σ_{12} and Σ_{22} can make sense. In this instance, the hypothesis $H_o : \Sigma_{12} = O$, in which case the regression coefficient matrix is a null matrix or, equivalently, the hypothesis that Y does not contribute to predicting X , implies that the canonical correlation matrix $\Sigma_{11}^{-\frac{1}{2}}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-\frac{1}{2}}$ is as well a null matrix. Accordingly, in this case, the ‘no regression’ hypothesis $\Sigma_{12} = O$ (no contribution of Y in predicting X) is equivalent to the hypothesis that the canonical correlations are equal to zero and vice versa.

10.5. The General Sampling Distribution of the Canonical Correlation Matrix

Let the $(p + q) \times 1$ real vector random variable $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N_{p+q}(\mu, \Sigma), \Sigma > O$. Consider a simple random sample of size n from this Gaussian population and let the

sample sum of products matrix S be partitioned as in the preceding section. Let the sample canonical correlation matrix be denoted by R and the corresponding population canonical correlation, by P , that is,

$$R = S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-1} S_{21} S_{11}^{-\frac{1}{2}} \text{ and } P = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}.$$

We now examine the distribution of R , assuming that $P \neq O$. Letting the determinant of $I - R$ be denoted by u , we have

$$u = |I - R| = |S_{11} - S_{12} S_{22}^{-1} S_{21}| |S_{11}|^{-1} = \frac{|S|}{|S_{11}| |S_{22}|}.$$

Thus, the h -th moment of u is

$$E[u^h] = E\left[\frac{|S|}{|S_{11}| |S_{22}|}\right]^h = E[|S|^h |S_{11}|^{-h} |S_{22}|^{-h}].$$

Since S , S_{11} and S_{22} are functions of S , we can integrate out over the density of S , namely the Wishart density with $m = n - 1$ degrees of freedom and parameter matrix $\Sigma > O$. Then for $m \geq p + q$,

$$E[u^h] = \frac{1}{|2\Sigma|^{\frac{m}{2}} \Gamma_{p+q}(\frac{m}{2})} \int_{S>O} \left[\frac{|S|}{|S_{11}| |S_{22}|}\right]^h |S|^{\frac{m}{2} - \frac{p+q+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)} dS. \quad (10.5.1)$$

Let us substitute S to $\frac{1}{2}S$ so that 2 will vanish from the factors containing 2, and let us replace $|S_{11}|^{-h}$ and $|S_{22}|^{-h}$ by equivalent integrals:

$$\begin{aligned} |S_{11}|^{-h} &= \frac{1}{\Gamma_p(h)} \int_{Y_1>O} |Y_1|^{h - \frac{p+1}{2}} e^{-\text{tr}(Y_2 S_{11})} dY_1, \quad \Re(h) > \frac{p-1}{2}; \\ |S_{22}|^{-h} &= \frac{1}{\Gamma_q(h)} \int_{Y_2>O} |Y_2|^{h - \frac{q+1}{2}} e^{-\text{tr}(Y_2 S_{22})} dY_2, \quad \Re(h) > \frac{q-1}{2}. \end{aligned}$$

Then,

$$\begin{aligned} E[u^h] &= \frac{1}{|\Sigma|^{\frac{m}{2}} \Gamma_{p+q}(\frac{m}{2})} \frac{1}{\Gamma_p(h) \Gamma_q(h)} \int_{Y_1>O} \int_{Y_2>O} |Y_1|^{h - \frac{p+1}{2}} |Y_2|^{h - \frac{q+1}{2}} \int_{S>O} |S|^{\frac{m}{2} + h - \frac{p+q+1}{2}} \\ &\quad \times e^{-\text{tr}(\Sigma^{-1}S + YS)} dS \wedge dY_1 \wedge dY_2 \quad (10.5.2) \end{aligned}$$

where

$$\text{tr}(YS) = \text{tr}\left\{ \begin{pmatrix} Y_1 & O \\ O & Y_2 \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \right\} = \text{tr}(Y_1 S_{11}) + \text{tr}(Y_2 S_{22}), \quad Y = \begin{pmatrix} Y_1 & O \\ O & Y_2 \end{pmatrix}.$$

Integrating over S in (10.5.2) gives

$$E[u^h] = \frac{\Gamma_{p+q}(\frac{m}{2} + h)}{\Gamma_{p+q}(\frac{m}{2})} \frac{1}{|\Sigma|^{\frac{m}{2}} \Gamma_p(h) \Gamma_q(h)} \int_{Y_1 > 0} \int_{Y_2 > 0} |Y_1|^{h - \frac{p+1}{2}} |Y_2|^{h - \frac{q+1}{2}} \times |\Sigma^{-1} + Y|^{-(\frac{m}{2} + h)} dY_1 \wedge dY_2.$$

Let

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \Rightarrow [\Sigma^{-1} + Y] = \begin{bmatrix} \Sigma^{11} + Y_1 & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} + Y_2 \end{bmatrix}.$$

Then, the determinant can be expanded as follows:

$$\begin{aligned} |\Sigma^{-1} + Y| &= |\Sigma^{22} + Y_2| |\Sigma^{11} + Y_1 - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}| \\ &= |\Sigma^{22} + Y_2| |Y_1 + B|, \quad B = \Sigma^{11} - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}, \end{aligned}$$

so that

$$|\Sigma^{-1} + Y|^{-(\frac{m}{2} + h)} = |\Sigma^{22} + Y_2|^{-(\frac{m}{2} + h)} |I + B^{-1} Y_1|^{-(\frac{m}{2} + h)} |B|^{-(\frac{m}{2} + h)}.$$

Collecting the factors containing Y_1 and integrating out, we have

$$\frac{1}{\Gamma_p(h)} \int_{Y_1 > 0} |Y_1|^{h - \frac{p+1}{2}} |I + B^{-1} Y_1|^{-(\frac{m}{2} + h)} dY_1 = \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} |B|^h, \quad \Re(h) > \frac{p-1}{2},$$

and $|B|^{-(\frac{m}{2} + h)} |B|^h = |B|^{-\frac{m}{2}}$. Noting that

$$\begin{vmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} + Y_2 \end{vmatrix} = \begin{cases} |\Sigma^{11}| |\Sigma^{22} + Y_2 - \Sigma^{21}(\Sigma^{11})^{-1} \Sigma^{12}| \\ |\Sigma^{22} + Y_2| |\Sigma^{11} - \Sigma^{12}(\Sigma^{22} + Y_2)^{-1} \Sigma^{21}| \end{cases} = |\Sigma^{22} + Y_2| |B|,$$

$|B|$ can be expressed in the following form:

$$|B| = \frac{|\Sigma^{11}| |Y_2 + C|}{|\Sigma^{22} + Y_2|}, \quad C = \Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1} \Sigma^{12} = \Sigma_{22}^{-1}, \quad (i)$$

so that,

$$\begin{aligned} |B|^{-\frac{m}{2}} &= |\Sigma^{11}|^{-\frac{m}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{\frac{m}{2}} \Rightarrow \\ |B|^{-\frac{m}{2}} |Y + \Sigma^{22}|^{-(\frac{m}{2} + h)} &= |\Sigma^{11}|^{-\frac{m}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{-h}. \end{aligned} \quad (ii)$$

Collecting all the factors containing Y_2 and integrating out, we have the following:

$$\begin{aligned} & \frac{1}{\Gamma_q(h)} \int_{Y_2 > O} |Y_2|^{h-\frac{q+1}{2}} |Y_2 + \Sigma_{22}^{-1}|^{-\frac{m}{2}} |Y_2 + \Sigma^{22}|^{-h} dY_2 \\ &= \frac{|\Sigma_{22}|^{\frac{m}{2}+h}}{\Gamma_q(h)} \int_{Y_2 > O} |Y_2|^{h-\frac{q+1}{2}} |I + \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}}|^{-\frac{m}{2}} \\ & \quad \times |\Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}} + \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}|^{-h} dY_2 \\ &= \frac{|\Sigma_{22}|^{\frac{m}{2}}}{\Gamma_q(h)} \int_{W > O} |W|^{h-\frac{q+1}{2}} |I + W|^{-\frac{m}{2}} |W + \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}|^{-h} dW, \quad W = \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}}, \end{aligned} \tag{iii}$$

as $W = \Sigma_{22}^{\frac{1}{2}} Y_2 \Sigma_{22}^{\frac{1}{2}} \Rightarrow dY_2 = |\Sigma_{22}|^{-\frac{q+1}{2}} dW$. Now, letting $W = U^{-1} - I$, so that $dW = |U|^{-(q+1)} dU$ with $O < U < I$, the expression in (iii), denoted by δ , becomes

$$\delta = \frac{|\Sigma_{22}|^{\frac{m}{2}}}{\Gamma_q(h)} \int_{O < U < I} |U|^{\frac{m}{2}-\frac{q+1}{2}} |I - U|^{h-\frac{q+1}{2}} |I - AU|^{-h} dU$$

where $A = I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}$. Note that since $|\Sigma|^{\frac{m}{2}} = |\Sigma_{22}|^{\frac{m}{2}} |\Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1} \Sigma_{21}|^{\frac{m}{2}} = |\Sigma_{22}|^{\frac{m}{2}} |\Sigma^{11}|^{-\frac{m}{2}}$, $|\Sigma|^{\frac{m}{2}}$ in the denominator of the constant part gets canceled out, the remaining constant expression being

$$\frac{\Gamma_{p+q}(\frac{m}{2} + h)}{\Gamma_{p+q}(\frac{m}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)}. \tag{iv}$$

The integral part of δ can be evaluated by making use of Euler's representation of a Gauss' hypergeometric function of matrix argument, which as given in formula (5.2.15) of Mathai (1997), is

$$\frac{\Gamma_q(a)\Gamma_q(c-a)}{\Gamma_q(c)} {}_2F_1(a, b; c; X) = \int_{O < Z < I} |Z|^{a-\frac{q+1}{2}} |I - Z|^{c-a-\frac{q+1}{2}} |I - XZ|^{-b} dZ$$

where $O < Z < I$ and $O < X < I$ are $q \times q$ real matrices. Thus, δ can be expressed as the follows:

$$\delta = \frac{\Gamma_q(\frac{m}{2})\Gamma_q(h)}{\Gamma_q(\frac{m}{2} + h)\Gamma_q(h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right),$$

so that

$$E[u^h] = \frac{\Gamma_{p+q}(\frac{m}{2} + h)\Gamma_p(\frac{m}{2})}{\Gamma_{p+q}(\frac{m}{2})\Gamma_p(\frac{m}{2} + h)} \frac{\Gamma_q(\frac{m}{2})}{\Gamma_q(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right). \tag{10.5.3}$$

For $p \leq q$, it follows from the definition of the matrix-variate gamma function that $\Gamma_{p+q}(\alpha) = \pi^{-p q/2} \Gamma_q(\alpha) \Gamma_p(\alpha - q/2)$. Thus, the constant part in (10.5.3) simplifies to

$$\frac{\Gamma_p(\frac{m}{2} - \frac{q}{2} + h)}{\Gamma_p(\frac{m}{2} - \frac{q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)}.$$

Then,

$$E[u^h] = \frac{\Gamma_p(\frac{m-q}{2} + h)}{\Gamma_p(\frac{m-q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{22}^{\frac{1}{2}} \Sigma^{22} \Sigma_{22}^{\frac{1}{2}}\right) \quad (10.5.4)$$

for $m \geq p + q$, $\Re(h) > -\frac{m}{2} + \frac{p-1}{2} + \frac{q}{2}$, $p \leq q$. Had Y_2 been integrated out first instead of Y_1 , we would have ended up with a hypergeometric function having $I - \Sigma_{11}^{\frac{1}{2}} \Sigma^{11} \Sigma_{11}^{\frac{1}{2}}$ as its argument, that is,

$$E[u^h] = \frac{\Gamma_p(\frac{m-q}{2} + h)}{\Gamma_p(\frac{m-q}{2})} \frac{\Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, h; \frac{m}{2} + h; I - \Sigma_{11}^{\frac{1}{2}} \Sigma^{11} \Sigma_{11}^{\frac{1}{2}}\right). \quad (10.5.5)$$

10.5.1. The sampling distribution of the multiple correlation coefficient

When $p = 1$ and $q > 1$, $r_{1(1\dots q)}^2$ is equal to the square of the sample multiple correlation coefficient $r_{1(1\dots q)}$. In this case, the argument in (10.5.5) is a real scalar quantity that is equal to $1 - \sigma_{11} \sigma^{11}$, the real matrix-variate $\Gamma_p(\cdot)$ functions are simply $\Gamma(\cdot)$ functions and $u = 1 - r_{1(1\dots q)}^2$. Letting $y = r_{1(1\dots q)}^2$, $E[1 - y]^h$ is available from (10.5.5) for $p = 1$ and the argument of the ${}_2F_1$ hypergeometric function is then

$$1 - \sigma_{11} \sigma^{11} = 1 - \sigma_{11} (\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1} = -\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}. \quad (10.5.6)$$

By taking the inverse Mellin transform of (10.5.5) for $h = s - 1$ and $p = 1$, we can express the density $f(y)$ of the square of the sample multiple correlation as follows:

$$f(y) = \frac{(1 - \rho^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}{\Gamma(\frac{m-q}{2}) \Gamma(\frac{q}{2})} y^{\frac{q}{2}-1} (1 - y)^{\frac{m-q}{2}-1} {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{q}{2}; \rho^2 y\right) \quad (10.5.7)$$

where ρ^2 is the population multiple correlation squared, that is, $\rho^2 = [\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}] / \sigma_{11}$. We can verify the result by computing the h -th moment of $1 - y$

in (10.5.7). The h -th moment can be determined as follows by expanding the ${}_2F_1$ function and then integrating:

$$E[1 - y]^h = \frac{(1 - \rho^2)^{\frac{m}{2}} \Gamma(\frac{m}{2})}{\Gamma(\frac{m-q}{2}) \Gamma(\frac{q}{2})} \sum_{k=0}^{\infty} \frac{(\frac{m}{2})_k (\frac{m}{2})_k (\rho^2)^k}{(\frac{q}{2})_k k!} \\ \times \int_0^1 y^{\frac{q}{2}+k-1} (1 - y)^{\frac{m-q}{2}+h-1} dy,$$

the integral part being

$$\frac{\Gamma(\frac{q}{2} + k) \Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m}{2} + h + k)} = \frac{\Gamma(\frac{q}{2}) \Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m}{2} + h)} \frac{(\frac{q}{2})_k}{(\frac{m}{2} + h)_k},$$

so that

$$E[1 - y]^h = (1 - \rho^2)^{\frac{m}{2}} \frac{\Gamma(\frac{m-q}{2} + h)}{\Gamma(\frac{m-q}{2})} \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2} + h)} {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{m}{2} + h; \rho^2\right). \quad (10.5.8)$$

On applying the relationship,

$${}_2F_1(a, b; c; z) = (1 - z)^{-b} {}_2F_1\left(c - a, b; c; \frac{z}{z - 1}\right), \quad (10.5.9)$$

we have

$${}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{m}{2} + h; \rho^2\right) = (1 - \rho^2)^{-\frac{m}{2}} {}_2F_1\left(h, \frac{m}{2}; \frac{m}{2} + h; \frac{\rho^2}{\rho^2 - 1}\right),$$

with

$$\frac{\rho^2}{\rho^2 - 1} = -\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}},$$

which agrees with (10.5.6). Observe that $(1 - \rho^2)^{\frac{m}{2}}$ gets canceled out so that (10.5.8) agrees with (10.5.5) for $p = 1$.

We can also obtain a representation of the density of the sample canonical correlation matrix whose M-transform is as given in (10.5.5) for $p \leq q$. This can be achieved by duplicating the steps utilized for the particular case considered in this section, which yields the following density:

$$f(R) = \frac{|I - P|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})}{\Gamma_p(\frac{m-q}{2}) \Gamma_p(\frac{q}{2})} |R|^{\frac{q-p+1}{2}} |I - R|^{\frac{m-q-p+1}{2}} \times {}_2F_1\left(\frac{m}{2}, \frac{m}{2}; \frac{q}{2}; P^{\frac{1}{2}} R P^{\frac{1}{2}}\right) \tag{10.5.10}$$

where $P = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}$ is the population canonical correlation matrix. Note that a function giving rise to a certain M-transform need not be unique. However, by making use of the Laplace transform and its inverse in the real matrix-variate case, Mathai (1981) has shown that the function specified in (10.5.10) is actually the unique density of R .

Exercises 10

10.1. In Example 10.3.2, verify that

$$\frac{[\alpha'_{(2)} \Sigma_{12} \beta_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2$$

where λ_2 is the second largest eigenvalue of the canonical correlation matrix A .

10.2. In Example 10.3.2, use equation (10.1.1) or equation (ii) preceding it with $\rho_1 = \rho_2 = \rho$ and evaluate $\beta_{(1)}$ and $\beta_{(2)}$ from $\alpha_{(1)}$ and $\alpha_{(2)}$. Obtain β first, normalize it subject to the constraint $\beta' \Sigma_{22} \beta = 1$ and then obtain $\beta_{(1)}$ and $\beta_{(2)}$. Then verify the results

$$\frac{[\alpha'_{(1)} \Sigma_{12} \beta_{(1)}]^2}{\gamma_1 \delta_1} = \lambda_1 \text{ and } \frac{[\alpha'_{(2)} \Sigma_{12} \beta_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2$$

where λ_1 and λ_2 are the largest and second largest eigenvalues of the canonical correlation matrix A .

10.3. Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{Cov}(X) = \Sigma_{11} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

$$\text{Cov}(Y) = \Sigma_{22} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}, \text{Cov}(X, Y) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 1 & 0 \end{bmatrix},$$

where x_1, x_2, x_3, y_1, y_2 are real scalar random variables. Evaluate the following where the notations of this chapter are utilized: (1): The canonical correlations $\rho_{(1)}$ and $\rho_{(2)}$; (2): The first pair of canonical variables (u_1, v_1) by direct evaluation as done in Example 10.3.2; (3): Verify that

$$\frac{[\beta'_{(1)} \Sigma_{21} \alpha_{(1)}]^2}{\gamma_1 \delta_1} = \lambda_1 : \text{the largest eigenvalue of } B$$

where $B = \Sigma_{22}^{-\frac{1}{2}} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}$; (4): Evaluate the second pair of canonical variables (u_2, v_2) by using equation (10.1.1) for constructing $\alpha_{(1)}$ and $\alpha_{(2)}$ after obtaining $\beta_{(1)}$ and $\beta_{(2)}$; (5): Verify that

$$\frac{[\beta'_{(2)} \Sigma_{21} \alpha_{(2)}]^2}{\gamma_2 \delta_2} = \lambda_2 : \text{the second largest eigenvalue of } B.$$

10.4. Repeat Problem 10.3 with X, Y and their associated covariance matrices defined as follows:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \text{Cov}(X) = \Sigma_{11} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix},$$

$$\text{Cov}(Y) = \Sigma_{22} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \text{Cov}(X, Y) = \Sigma_{12} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$$

where $x_1, x_2, x_3, y_1, y_2, y_3$ are real scalar random variables. As well, compute the three pairs of canonical variables.

10.5. Show that the M-transform in (10.5.5) is available from the density specified in (10.5.10).

References

- A.M. Mathai (1981): Distribution of the canonical correlation matrix, *Annals of the Institute of Statistical Mathematics*, **33**(A), 35–43.
- A.M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.
- A.M. Mathai and H.J. Haubold (2017a): *Linear Algebra for Physicists and Engineers*, De Gruyter, Germany.
- A.M. Mathai and H.J. Haubold (2017b): *Probability and Statistics for Physicists and Engineers*, De Gruyter, Germany.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 11

Factor Analysis



11.1. Introduction

We will utilize the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. In the real and complex cases, the determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

Factor analysis is a statistical method aiming to identify a relatively small number of underlying (unobserved) factors that could explain certain interdependencies among a larger set of observed variables. Factor analysis also proves useful for analyzing causal mechanisms. As a statistical technique, Factor Analysis was originally developed in connection with psychometrics. It has since been utilized in operations research, finance and biology, among other disciplines. For instance, a score available on an intelligence test will often assess several intellectual faculties and cognitive abilities. It is assumed that a

certain linear function of the contributions from these various mental factors is producing the final score. Hence, there is a parallel to be made with analysis of variance as well as design of experiments and linear regression models.

11.2. Linear Models from Different Disciplines

In order to introduce the current topic, we will first examine a linear regression model and an experimental design model.

11.2.1. A linear regression model

Let x be a real scalar random variable and let t_1, \dots, t_r be either r real fixed numbers or given values of r real random variables. Let the conditional expectation of x , given t_1, \dots, t_r , be of the form

$$E[x|t_1, \dots, t_r] = a_0 + a_1 t_1 + \dots + a_r t_r$$

or the corresponding model be

$$x = a_0 + a_1 t_1 + \dots + a_r t_r + e$$

where a_0, a_1, \dots, a_r are unknown constants, t_1, \dots, t_r are given values and e is the error component or the sum total of contributions coming from unknown or uncontrolled factors plus the experimental error. For example, x might be an inflation index with respect to a particular base year, say 2010. In this instance, t_1 may be the change or deviation in the average price per kilogram of certain staple vegetables from the base year 2010, t_2 may be the change or deviation in the average price of a kilogram of rice compared to the base year, t_3 may be the change or deviation in the average price of flour per kilogram with respect to the base year, and so on, and t_r may be the change or deviation in the average price of milk per liter compared to the base year 2010. The notation t_j , $j = 1, \dots, r$, is utilized to designate the given values as well as the corresponding random variables. Since we are taking deviations from the base values, we may assume without any loss of generality that the expected value of t_j is zero, that is, $E[t_j] = 0$, $j = 1, \dots, r$. We may also take the expected value of the error term e to be zero, that is, $E[e] = 0$. Now, let x_1 be the inflation index, x_2 be the caloric intake index per person, x_3 be the general health index and so on. In all these cases, the same t_1, \dots, t_r can act as the independent variables in a regression set up. Thus, in such a situation, a multivariate linear regression model will have the following format:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pr} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_r \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}, \quad (11.2.1)$$

and we may write this model as

$$X = M + \Lambda F + \epsilon$$

where $\Lambda = (\lambda_{ij})$ is a $p \times r$, $r \leq p$, matrix of full rank r , ϵ is $p \times 1$ and F is $r \times 1$. In (11.2.1), $\lambda_{ij} = a_{ij}$ and $f_j = t_j$. Then, $E[X] = M + \Lambda E[F] + E[\epsilon] = M$ since we have assumed that $E[F] = O$ (a null matrix) and $E[\epsilon] = O$. When F and ϵ are uncorrelated, the covariance matrix associated with X , denoted by $\text{Cov}(X) = \Sigma$, is the following:

$$\begin{aligned} \Sigma = \text{Cov}(X) &= E\{(X - M)(X - M)'\} = E\{(\Lambda F + \epsilon)(\Lambda F + \epsilon)'\} \\ &= \Lambda \text{Cov}(F) \Lambda' + \text{Cov}(\epsilon) + O \\ &= \Lambda \Phi \Lambda' + \Psi \end{aligned} \tag{11.2.2}$$

where the covariance matrices of F and ϵ are respectively denoted by $\Phi > O$ (positive definite) and $\Psi > O$. In the above formulation, F is taken to be a real vector random variable. In a simple linear model, the covariance matrix of ϵ , namely Ψ , is usually taken as $\sigma^2 I$ where $\sigma^2 > 0$ is a real scalar quantity and I is the identity matrix. In a more general setting, Ψ can be taken to be a diagonal matrix whose diagonal elements are positive; in such a model, the e_j 's are uncorrelated and their variances need not be equal. It will be assumed that the covariance matrix Ψ in (11.2.2) is a diagonal matrix having positive diagonal elements.

11.2.2. A design of experiment model

Consider a completely randomized experiment where one set of treatments are undertaken. In this instance, the experimental plots are assumed to be fully homogeneous with respect to all the known factors of variation that may affect the response. For example, the observed value may be the yield of a particular variety of corn grown in an experimental plot. Let the set of treatments be r different fertilizers F_1, \dots, F_r , the effects of these fertilizers being denoted by $\alpha_1, \dots, \alpha_r$. If no fertilizer is applied, the yield from a test plot need not be zero. Let μ_1 be a general effect when F_1 is applied so that we may regard α_1 as a deviation from this effect μ_1 due to F_1 . Let e_1 be the sum total of the contributions coming from all unknown or uncontrolled factors plus the experimental error, if any, when F_1 is applied. Then a simple linear one-way classification model for F_1 is

$$x_1 = \mu_1 + \alpha_1 + e_1,$$

with x_1 representing the yield from the test plot where F_1 was applied. Then, corresponding to F_1, \dots, F_r , $r = p$ we have the following system:

$$\begin{aligned} x_1 &= \mu_1 + \alpha_1 + e_1 \\ &\vdots \\ x_p &= \mu_p + \alpha_p + e_p \end{aligned}$$

or, in matrix notation,

$$X = M + \Lambda F + \epsilon \quad (11.2.3)$$

where

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}, \quad F = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}, \quad \text{and } \Lambda = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

In this case, the elements of Λ are dictated by the design itself. If the vector F is fixed, we call the model specified by (11.2.3), the fixed effect model, whereas if F is assumed to be random, then it is referred to as the random effect model. With a single observation per cell, as stated in (11.2.3), we will not be able to estimate the parameters or test hypotheses. Thus, the experiment will have to be replicated. So, let the j -th replicated observation vector be

$$X_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{pj} \end{bmatrix}, \quad j = 1, \dots, n,$$

Σ , Φ , and Ψ remaining the same for each replicate within the random effect model. Similarly, for the regression model given in (11.2.1), the j -th replication or repetition vector will be $X'_j = (x_{1j}, \dots, x_{pj})$ with Σ , Φ and Ψ therein remaining the same for each sample.

We will consider a general linear model encompassing those specified in (11.2.1) and (11.2.3) and carry out a complete analysis that will involve verifying the existence and uniqueness of such a model, estimating its parameters and testing various types of hypotheses. The resulting technique is referred to as Factor Analysis.

11.3. A General Linear Model for Factor Analysis

Consider the following general linear model:

$$X = M + \Lambda F + \epsilon \quad (11.3.1)$$

where

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad M = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix},$$

$$A = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pr} \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} f_1 \\ \vdots \\ f_r \end{bmatrix}, \quad r \leq p,$$

with the μ_j 's, λ_{ij} 's, f_j 's being real scalar parameters, the x_j 's, $j = 1, \dots, p$, being real scalar quantities, and A being of dimension $p \times r$, $r \leq p$, and of full rank r . When considering expected values, variances, covariances, etc., X , F , and ϵ will be assumed to be random quantities; however, when dealing with estimates, X will represent a vector of observations. This convention will be employed throughout this chapter so as to avoid a multiplicity of symbols for the variables and the corresponding observations.

From a geometrical perspective, the r columns of A , which are linearly independent, span an r -dimensional subspace in the p -dimensional Euclidean space. In this case, the $r \times 1$ vector F is a point in this r -subspace and this subspace is usually called the *factor space*. Then, right-multiplying the $p \times r$ matrix A by a matrix will correspond to employing a new set of coordinate axes for the factor space.

Factor Analysis is a subject dealing with the identification or unique determination of a model of the type specified in (11.3.1), as well as the estimation of its parameters and the testing of various related hypotheses. The subject matter was originally developed in connection with intelligence testing. Suppose that a test is administered to an individual to evaluate his/her mathematical skills, spatial perception, language abilities, etc., and that the score obtained is recorded. There will be a component in the model representing the expected score. If the test is administered to 10th graders belonging to a particular school, the grand average of such test scores among all 10th graders across the nation could be taken as the expected score. Then, inputs associated to various intellectual faculties or combinations thereof will come about. All such factors may be contributing towards the observed test score. If f_1, \dots, f_r , are the contributions coming from r factors corresponding to specific intellectual abilities, then, when a linear model is assumed, a certain linear combination of these inputs will constitute the final quantity accounting for the observed test score. A test score, x_1 , may then result from a linear model of the following form:

$$x_1 = \mu_1 + \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1r}f_r + e_1$$

with $\lambda_{11}, \dots, \lambda_{1r}$ being the coefficients of f_1, \dots, f_r , where f_1, \dots, f_r , are contributions from r factors toward x_1 ; these factors may be called the *main intellectual factors* in this case, and the coefficients $\lambda_{11}, \dots, \lambda_{1r}$ may be referred to as the *factor loadings* for these main factors. In this context, μ_1 is the general expected value and e_1 is the error component or the sum total of contributions originating from all unknown factors plus the experimental error, if any. Note that the contributions f_1, \dots, f_r due to the main intellectual factors can vary from individual to individual, and hence it is appropriate to treat f_1, \dots, f_r as random variables rather than fixed unknown quantities. These f_1, \dots, f_r are not observable as in the case of the design model in (11.2.3), whereas in the regression type model specified by (11.2.1), they may take on the recorded values of the observable variables which are called the independent variables. Thus, the model displayed in (11.3.1) may be analyzed either by treating f_1, \dots, f_r as fixed quantities or as random variables. If they are treated as random variables, we can assume that f_1, \dots, f_r follow some joint distribution. Usually, joint normality is presupposed for f_1, \dots, f_r . Since f_1, \dots, f_r are deviations from the general effect μ_1 due to certain main intellectual faculties under consideration, it may be assumed that the expected value is a null vector, that is, $E[F] = O$. We will denote the covariance matrix associated with F as Φ : $\text{Cov}(F) = \Phi > O$ (real positive definite). Note that the model's error term e_j is always a random variable. Letting x_1, \dots, x_p be the test scores on p individuals, we have the error vector $\epsilon' = (e_1, \dots, e_p)$. Without any loss of generality, we may take the expected value of ϵ as being a null vector, that is, $E[\epsilon] = O$. For a very simple situation, we may assume the covariance matrix associated with ϵ to be $\text{Cov}(\epsilon) = \sigma^2 I$ where $\sigma^2 > 0$ is a real positive scalar quantity and I is the identity matrix. For a somewhat more general situation, we may take $\text{Cov}(\epsilon) = \Psi$ where Ψ is a real positive definite diagonal matrix, or a diagonal matrix with positive diagonal elements. In the most general case, we may take Ψ to be a real positive definite matrix. It will be assumed that Ψ is diagonal with positive diagonal elements in the model (11.3.1), and that F and ϵ are uncorrelated. Thus, letting Σ be the covariance matrix of X , we have

$$\begin{aligned}\Sigma &= E[(X - M)(X - M)'] = E[(\Lambda F + \epsilon)(\Lambda F + \epsilon)'] \\ &= \Lambda E(F F') \Lambda' + E(\epsilon \epsilon') + O \\ &= \Lambda \Phi \Lambda' + \Psi\end{aligned}\tag{11.3.2}$$

where Σ , Φ and Ψ , with $\Sigma = \Lambda \Phi \Lambda' + \Psi$, are all assumed to be real positive definite matrices.

11.3.1. The identification problem

Is the model specified by (11.3.1) unique or could it represent different situations? In other words, does it make sense as a model as stated? Given any $r \times r$ nonsingular matrix A , let $AF = F^*$ and $\Lambda A^{-1} = \Lambda^*$. Then, $\Lambda^* F^* = \Lambda A^{-1} AF = \Lambda F$. In other words,

$$X = M + \Lambda F + \epsilon = M + \Lambda^* F^* + \epsilon. \quad (11.3.3)$$

Consequently, the model in (11.3.1) is not identified, that is, it is not uniquely determined.

The identification problem can also be stated as follows: Does there exist a real positive definite $p \times p$ matrix $\Sigma > O$ containing $p(p+1)/2$ distinct elements, which can be uniquely represented as $\Lambda\Phi\Lambda' + \Psi$ where Λ has pr distinct elements, $\Phi > O$ has $r(r+1)/2$ distinct elements and Ψ is a diagonal matrix having p distinct elements? There is clearly no such matrices as can be inferred from (11.3.3). Note that an $r \times r$ arbitrary matrix A represents r^2 distinct elements. It can be observed from (11.3.3) that we can impose r^2 conditions on the parameters in Λ , Φ and Ψ . The question could also be posed as follows: Can the $p(p+1)/2$ distinct elements in Σ plus the r^2 elements in A (r^2 conditions) uniquely determine all the elements of Λ , Ψ and Φ ? Let us determine how many elements there are in total. Λ , Ψ and Φ have a total of $pr + p + r(r+1)/2$ elements while A and Σ have a total of $r^2 + p(p+1)/2$ elements. Hence, the difference, denoted by δ , is

$$\delta = \frac{p(p+1)}{2} + r^2 - \left[pr + \frac{r(r+1)}{2} + p \right] = \frac{1}{2}[(p-r)^2 - (p+r)]. \quad (11.3.4)$$

Observe that the right-hand side of (11.3.2) is not a linear function of Λ , Φ and Ψ . Thus, if $\delta > 0$, we can anticipate that existence and uniqueness will hold although these properties cannot be guaranteed, whereas if $\delta < 0$, then existence can be expected but uniqueness may be in question. Given (11.3.2), note that

$$\Sigma = \Psi + \Lambda\Phi\Lambda' \Rightarrow \Sigma - \Psi = \Lambda\Phi\Lambda'$$

where $\Lambda\Phi\Lambda'$ is positive semi-definite of rank r , since the $p \times r$, $r \leq p$, matrix Λ has full rank r and $\Phi > O$ (positive definite). Then, the existence question can also be stated as follows: Given a $p \times p$ real positive definite matrix $\Sigma > O$, can we find a diagonal matrix Ψ with positive diagonal elements such that $\Sigma - \Psi$ is a real positive semi-definite matrix of a specified rank r , which is expressible in the form BB' for some $p \times r$ matrix B of rank r where $r \leq p$? For the most part, the available results on this question of existence can be found in Anderson (2003) and Anderson and Rubin (1956). *If a set of parameters exist and if the model is uniquely determined, then we say that the model is identified, or*

alternatively, identifiable. The concept of identification or identifiability within the context of Factor Analysis has been studied by Ihara and Kano (1995), Wegge (1996), Allman et al. (2009) and Chen et al. (2020), among others.

Assuming that $\Phi = I$ will impose $r(r + 1)/2$ conditions. However, $r^2 = \frac{r(r+1)}{2} + \frac{r(r-1)}{2}$. Thus, we may impose $r(r - 1)/2$ additional conditions after requiring that $\Phi = I$. Observe that when $\Phi = I$, $\Lambda^* \Phi \Lambda'^* = \Lambda^* \Lambda'^* = \Lambda \Lambda^{-1} \Lambda'^{-1} \Lambda'$, and if this is equal to $\Lambda \Lambda'$ assuming that $\Phi = I$, this means that $(\Lambda' \Lambda)^{-1} = I$ or $\Lambda' \Lambda = I$, that is, Λ is an orthonormal matrix. Thus, under the condition $\Phi = I$, the arbitrary $r \times r$ matrix Λ becomes an orthonormal matrix. In this case, the transformation $Y = \Lambda A$ is an orthonormal transformation or a rotation of the coordinate axes. The following $r \times r$ symmetric matrix of $r(r + 1)/2$ distinct elements

$$\Delta = \Lambda' \Psi^{-1} \Lambda \quad (11.3.5)$$

is needed for solving estimation and hypothesis testing problems; accordingly, we can impose $r(r - 1)/2$ conditions by requiring Δ to be diagonal with distinct diagonal elements, that is, $\Delta = \text{diag}(\eta_1, \dots, \eta_r)$, $\eta_j > 0$, $j = 1, \dots, r$. This imposes $\frac{r(r+1)}{2} - r = \frac{r(r-1)}{2}$ conditions. Thus, for the model to be identifiable or for all the parameters in Λ , Φ , Ψ to be uniquely determined, we can impose the condition $\Phi = I$ and require that $\Delta = \Lambda' \Psi^{-1} \Lambda$ be diagonal with positive diagonal elements. These two conditions will provide $\frac{r(r+1)}{2} + \frac{r(r-1)}{2} = r^2$ restrictions on the model which will then be identified.

When $\Phi = I$, the main factors are orthogonal. If Φ is a diagonal matrix (including the identity matrix), the covariances are zeros and it is an orthogonal situation, in which case we say that the *main factors are orthogonal*. If Φ is not diagonal, then we say that the *main factors are oblique*.

One can also impose $r(r - 1)/2$ conditions on the $p \times r$ matrix Λ . Consider the first $r \times r$ block, that is, the leading $r \times r$ submatrix or the upper $r \times r$ block in the $p \times r$ matrix, which we will denote by B . Imposing the condition that this $r \times r$ block B is lower triangular will result in $r^2 - \frac{r(r+1)}{2} = \frac{r(r-1)}{2}$ conditions. Hence, $\Phi = I$ and the condition that this leading $r \times r$ block B is lower triangular will guarantee r^2 restrictions, and the model will then be identified. One can also take a preselected $r \times r$ matrix B_1 and then impose the condition that $B_1 B$ be lower triangular. This will, as well, produce $\frac{r(r-1)}{2}$ conditions. Thus, $\Phi = I$ and $B_1 B$ being lower triangular will ensure the identification of the model.

When we impose conditions on Φ and Ψ , the unknown covariance matrices must assume certain formats. Such conditions can be justified. However, could conditions be put on Λ , the factor loadings? Letting the first $r \times r$ block B in the $p \times r$ matrix Λ be lower triangular is tantamount to assuming that $\lambda_{12} = 0 = \lambda_{13} = \dots = \lambda_{1r}$ or, equivalently, that

f_2, \dots, f_r do not contribute to the model for determining x_1 in $X' = (x_1, x_2, \dots, x_p)$. Such restrictions are justified if we can design the experiment in such a way that x_1 depends on f_1 alone and not on f_2, \dots, f_r . In psychological tests, it is possible to design tests in such a way that only certain main factors affect the scores. Thus, in such instances, we are justified to utilize a triangular format such that, in general, there are no contributions from f_{i+1}, \dots, f_r , toward x_i or, equivalently, the factor loadings $\lambda_{i+1}, \dots, \lambda_{ir}$ equal zero for $i = 1, \dots, r - 1$. For example, suppose that the first r tests are designed in such a way that only f_1, \dots, f_i and no other factors contribute to x_i or, equivalently, $x_i = \mu_i + \lambda_{i1}f_1 + \dots + \lambda_{ii}f_i + e_i$, $i = 1, \dots, r$. We can also measure the contribution from f_i in λ_{ii} units or we can take $\lambda_{ii} = 1$. By taking $B = I_r$, we can impose r^2 conditions without requiring that $\Phi = I$. This means that the first r tests are specifically designed so that x_1 only has a one unit contribution from f_1 , x_2 only has a one unit contribution from f_2 , and so on, x_r receiving a one unit contribution from f_r . When B is taken to be diagonal, the factor loadings are $\lambda_{11}, \lambda_{22}, \dots, \lambda_{rr}$, respectively, so that only f_i contributes to x_i for $i = 1, \dots, r$. Accordingly, the following are certain model identification conditions:

- (1): $\Phi = I$ and $\Lambda'\Psi^{-1}\Lambda$ is diagonal with distinct diagonal elements;
- (2): $\Phi = I$ and the leading $r \times r$ submatrix B in the $p \times r$ matrix Λ is triangular;
- (3): $\Phi = I$ and B_1B is lower triangular where B_1 is a preselected matrix;
- (4): The leading $r \times r$ submatrix B in the $p \times r$ matrix Λ is an identity matrix.

Observe that when $r = p$, condition (4) corresponds to the design model considered in (11.2.3).

11.3.2. Scaling or units of measurement

A shortcoming of any analysis being based on a covariance matrix Σ is that the covariances depend on the units of measurement of the individual variables. Thus, modifying the units will affect the covariances. If we let y_i and y_j be two real scalar random variables with variances σ_{ii} and σ_{jj} and associated covariance σ_{ij} , the effect of scaling or changes in the measurement units may be eliminated by considering the variables $z_i = y_i/\sqrt{\sigma_{ii}}$ and $z_j = y_j/\sqrt{\sigma_{jj}}$ whose covariance $\text{Cov}(z_i, z_j) \equiv r_{ij}$ is actually the correlation between y_i and y_j , which is free of the units of measurement. Letting $Y' = (y_1, \dots, y_p)$ and $D = \text{diag}(\frac{1}{\sqrt{\sigma_{11}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}})$, consider $Z = DY$. We note that $\text{Cov}(Y) = \Sigma \Rightarrow \text{Cov}(Z) = D\Sigma D = R$ which is the correlation matrix associated with Y .

In psychological testing situations or referring to the model (11.3.1), when a test score x_j is multiplied by a scalar quantity c_j , then the factor loadings $\lambda_{j1}, \dots, \lambda_{jr}$, the error term e_j and the general effect μ_j are all multiplied by c_j , that is, $c_j x_j = c_j \mu_j + c_j (\lambda_{j1} f_1 + \dots + \lambda_{jr} f_r) + c_j e_j$. Let $\text{Cov}(X) = \Sigma = (\sigma_{ij})$, that is, $\text{Cov}(x_i, x_j) = \sigma_{ij}$, $X' = (x_1, \dots, x_p)$ and $D = \text{diag}(\frac{1}{\sqrt{\sigma_{11}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}})$. Consider the model

$$DX = DM + D\Lambda F + D\epsilon \Rightarrow \text{Cov}(DX) = D\Sigma D = D\Lambda\Phi\Lambda'D + D\Psi D. \quad (11.3.6)$$

If $X^* = DX$, $M^* = DM$, $\Lambda^* = D\Lambda$, and $\epsilon^* = D\epsilon$, then we obtain the following model and the resulting covariance matrix:

$$\begin{aligned} X^* = M^* + \Lambda^* F + \epsilon^* &\Rightarrow \Sigma^* = \text{Cov}(X^*) = \Lambda^* \text{Cov}(F) \Lambda^{*'} + \Psi^* \\ &\Rightarrow D\Sigma D = D\Lambda\Phi\Lambda'D + D\Psi D \\ &\Rightarrow R = \Lambda^* \Phi \Lambda^{*'} + \Psi^* \end{aligned} \quad (11.3.7)$$

where $R = (r_{ij})$ is the correlation matrix in X . An interesting point to be noted is that the identification conditions $\Phi = I$ and $\Lambda^{*'} \Psi^{*-1} \Lambda^*$ being diagonal become the following: $\Phi = I$ and $\Lambda^{*'} \Psi^{*-1} \Lambda^* = \Lambda' D D^{-1} \Psi^{-1} D^{-1} D \Lambda = \Lambda' \Psi^{-1} \Lambda$ which is diagonal, that is, $\Lambda' \Psi^{-1} \Lambda$ is invariant under scaling transformations on the model or under $X^* = DX$ and $\Psi^* = D\Psi D$.

11.4. Maximum Likelihood Estimators of the Parameters

A simple random sample of size n from the model $X = M + \Lambda F + \epsilon$ specified in (11.3.1) is understood to be constituted of independently and identically distributed (iid) X_j 's, $j = 1, \dots, n$, where

$$X_j = M + \Lambda F + E_j, \quad j = 1, \dots, n, \quad X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{bmatrix}, \quad E_j = \begin{bmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{pj} \end{bmatrix}, \quad (11.4.1)$$

and the X_j 's are iid. Let f_j and E_j be independently normally distributed. Let $E_j \sim N_p(O, \Psi)$ and $X_j \sim N_p(\mu, \Sigma)$, $\Sigma = \Lambda\Phi\Lambda' + \Psi$ where $\Phi = \text{Cov}(F) > O$, $\Psi = \text{Cov}(\epsilon) > O$ and $\Sigma > O$, Ψ being a diagonal matrix with positive diagonal elements. Then, the likelihood function is the following:

$$\begin{aligned} L &= \prod_{j=1}^n \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_j - M)' \Sigma^{-1} (X_j - M)} \\ &= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{j=1}^n (X_j - M)' \Sigma^{-1} (X_j - M)}. \end{aligned} \quad (11.4.2)$$

The sample matrix will be denoted by a boldface $\mathbf{X} = (X_1, \dots, X_n)$. Let J be the $n \times 1$ vector of unities, that is, $J = (1, 1, \dots, 1)'$. Then,

$$\mathbf{X} = (X_1, \dots, X_n) = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix}$$

$$\Rightarrow \frac{1}{n}\mathbf{X}J = \begin{bmatrix} \frac{1}{n}(\sum_{j=1}^n x_{1j}) \\ \vdots \\ \frac{1}{n}(\sum_{j=1}^n x_{pj}) \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \bar{\mathbf{X}}$$

where $\bar{\mathbf{X}}$ is the sample average vector or the sample mean vector. Let the boldface $\bar{\mathbf{X}}$ be the $p \times n$ matrix $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)$. Then,

$$(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = S = (s_{ij}), \quad s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j), \quad (11.4.3)$$

where S is the sample sum of products matrix or the “corrected” sample sum of squares and cross products matrix. Note that

$$\begin{aligned} \bar{\mathbf{X}} &= \frac{1}{n}\mathbf{X}J \\ \Rightarrow \bar{\mathbf{X}} &= (\bar{X}_1, \dots, \bar{X}_p) = \mathbf{X}\left(\frac{1}{n}JJ'\right) \\ \Rightarrow \mathbf{X} - \bar{\mathbf{X}} &= \mathbf{X}\left(I - \frac{1}{n}JJ'\right). \end{aligned}$$

Thus,

$$S = \mathbf{X}\left(I - \frac{1}{n}JJ'\right)\left(I - \frac{1}{n}JJ'\right)'\mathbf{X}' = \mathbf{X}\left(I - \frac{1}{n}JJ'\right)\mathbf{X}'. \quad (11.4.4)$$

Since $(X_j - M)' \Sigma^{-1} (X_j - M)$ is a real scalar quantity, we have the following:

$$\begin{aligned}
\sum_{j=1}^n (X_j - M)' \Sigma^{-1} (X_j - M) &= \sum_{j=1}^n \text{tr}(X_j - M)' \Sigma^{-1} (X_j - M) \\
&= \sum_{j=1}^n \text{tr}[\Sigma^{-1} (X_j - M)(X_j - M)'] \\
&= \text{tr}[\Sigma^{-1} \sum_{j=1}^n (X_j - \bar{X} + \bar{X} - M)(X_j - \bar{X} + \bar{X} - M)'] \\
&= \text{tr}[\Sigma^{-1} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})'] \\
&\quad + n \text{tr}[\Sigma^{-1} (\bar{X} - M)(\bar{X} - M)'] \\
&= \text{tr}(\Sigma^{-1} S) + n(\bar{X} - M)' \Sigma^{-1} (\bar{X} - M). \tag{11.4.5}
\end{aligned}$$

Hence,

$$L = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\frac{1}{2} \{\text{tr}(\Sigma^{-1} S) + n(\bar{X} - M)' \Sigma^{-1} (\bar{X} - M)\}}. \tag{11.4.6}$$

Differentiating (11.4.6) with respect to M , equating the result to a null vector and solving, we obtain an estimator for M , denoted by \hat{M} , as $\hat{M} = \bar{X}$. Then, $\ln L$ evaluated at $M = \bar{X}$ is

$$\begin{aligned}
\ln L &= -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}(\Sigma^{-1} S) \\
&= -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Lambda \Phi \Lambda' + \Psi| - \frac{1}{2} \text{tr}[(\Lambda \Phi \Lambda' + \Psi)^{-1} S]. \tag{11.4.7}
\end{aligned}$$

11.4.1. Maximum likelihood estimators under identification conditions

The derivations in the following sections parallel to those found in Mathai (2021). One of the conditions for identification of the model is $\Phi = I$ and $\Lambda' \Psi^{-1} \Lambda$ being a diagonal matrix with positive diagonal elements. We will examine the maximum likelihood estimators (MLE)/maximum likelihood estimates (MLE) under this identification condition. In this case, it follows from (11.4.7) that

$$\ln L = -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Lambda \Lambda' + \Psi| - \frac{1}{2} \text{tr}[(\Lambda \Lambda' + \Psi)^{-1} S]. \tag{11.4.8}$$

By expanding the following determinant in two different ways by applying properties of partitioned determinants that are stated in Sect. 1.3, we have the following identities:

$$\begin{vmatrix} \Psi & -\Lambda \\ \Lambda' & I_r \end{vmatrix} = |\Psi| |I + \Lambda' \Psi^{-1} \Lambda| = |\Psi + \Lambda \Lambda'|. \quad (11.4.9)$$

Hence, letting

$$\Delta = \Lambda' \Psi^{-1} \Lambda = \text{diag}(\delta'_1 \delta_1, \delta'_2 \delta_2, \dots, \delta'_r \delta_r),$$

we have

$$\begin{aligned} \ln |\Lambda \Lambda' + \Psi| &= \ln |\Psi| + \ln |I + \Lambda' \Psi^{-1} \Lambda| \\ &= \sum_{j=1}^p \ln \psi_{jj} + \sum_{j=1}^r \ln(1 + \delta'_j \delta_j), \end{aligned} \quad (11.4.10)$$

where $\delta'_j = \Lambda'_j \Psi^{-\frac{1}{2}}$, Λ_j is the j -th column of Λ and ψ_{jj} , $j = 1, \dots, p$, is the j -th diagonal element of the diagonal matrix Ψ , the identification condition being that $\Phi = I$ and $\Lambda' \Psi^{-1} \Lambda = \Delta = \text{diag}(\delta'_1 \delta_1, \dots, \delta'_r \delta_r)$. Accordingly, if we can write $\text{tr}(\Sigma^{-1} S) = \text{tr}[(\Lambda \Lambda' + \Psi)^{-1} S]$ in terms of ψ_{jj} , $j = 1, \dots, p$, and $\delta'_j \delta_j$, $j = 1, \dots, r$, then the likelihood equation can be directly evaluated from (11.4.8) and (11.4.10), and the estimators can be determined. The following result will be helpful in this connection.

Theorem 11.4.1. *Whenever $\Lambda \Lambda' + \Psi$ is nonsingular, which in this case, means real positive definite, the inverse is given by*

$$(\Lambda \Lambda' + \Psi)^{-1} = \Psi^{-1} - \Psi^{-1} \Lambda (\Delta + I)^{-1} \Lambda' \Psi^{-1} \quad (11.4.11)$$

where the Δ is defined in (11.4.10).

It can be readily verified that pre and post multiplications of $\Psi^{-1} - \Psi^{-1} \Lambda (\Delta + I)^{-1} \Lambda' \Psi^{-1}$ by $\Lambda \Lambda' + \Psi$ yield the identity matrix I_p .

11.4.2. Simplifications of $|\Sigma|$ and $\text{tr}(\Sigma^{-1} S)$

In light of (11.4.9) and (11.4.10), we have

$$\begin{aligned} |\Sigma| &= |\Lambda \Lambda' + \Psi| = |\Psi| |I + \Lambda' \Psi^{-1} \Lambda| \\ &= |\Psi| |I + \Delta| = \left\{ \prod_{j=1}^p \psi_{jj} \right\} \left\{ \prod_{j=1}^r (1 + \delta'_j \delta_j) \right\}. \end{aligned}$$

Now, observe the following: In $\Lambda(\Delta + I)^{-1} = \Lambda \text{diag}(\frac{1}{1+\delta'_1\delta_1}, \dots, \frac{1}{1+\delta'_r\delta_r})$, the j -th column of Λ is multiplied by $\frac{1}{1+\delta'_j\delta_j}$, $j = 1, \dots, r$, and

$$\Lambda(\Delta + I)^{-1}\Lambda' = \sum_{j=1}^r \frac{1}{1 + \delta'_j\delta_j} \Lambda_j \Lambda'_j$$

where Λ_j is the j -th column of Λ and the δ_j 's are specified in (11.4.10). Thus,

$$\ln |\Sigma| = \sum_{j=1}^p \ln \psi_{jj} + \sum_{j=1}^r \ln(1 + \delta'_j\delta_j) \quad (11.4.12)$$

and, on applying Theorem 11.4.1,

$$\begin{aligned} \text{tr}(\Sigma^{-1}S) &= \text{tr}[(\Lambda\Lambda' + \Psi)^{-1}S] = \text{tr}[(\Psi^{-1}S)] - \text{tr}[\Psi^{-1}\Lambda(\Delta + I)^{-1}\Lambda'\Psi^{-1}S] \\ &= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^r \frac{1}{1 + \delta'_j\delta_j} \text{tr}(\Lambda_j \Lambda'_j (\Psi^{-1}S\Psi^{-1})) \\ &= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^r \frac{1}{1 + \delta'_j\delta_j} \text{tr}(\Lambda'_j (\Psi^{-1}S\Psi^{-1}) \Lambda_j) \\ &= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^r \frac{1}{1 + \delta'_j\delta_j} \Lambda'_j (\Psi^{-1}S\Psi^{-1}) \Lambda_j \end{aligned} \quad (11.4.13)$$

where Λ_j is the j -th column of Λ , which follows by making use of the property $\text{tr}(AB) = \text{tr}(BA)$ and observing that $\Lambda'_j (\Psi^{-1}S\Psi^{-1}) \Lambda_j$ is a quadratic form.

11.4.3. Special case $\Psi = \sigma^2 I_p$

Letting $\Psi = \sigma^2 I$ where σ^2 is a real scalar, $\Psi^{-1} = \sigma^{-2} I_p = \theta I_p$ where $\theta = \sigma^{-2}$, and the log-likelihood function can be simplified as

$$\begin{aligned} \ln L &= -\frac{np}{2} \ln(2\pi) + \frac{np}{2} \ln \theta - \frac{n}{2} \sum_{j=1}^r \ln(1 + \delta'_j\delta_j) \\ &\quad - \frac{\theta}{2} \text{tr}(S) + \frac{\theta^2}{2} \sum_{j=1}^r \frac{1}{1 + \delta'_j\delta_j} \Lambda'_j S \Lambda_j \end{aligned}$$

where $1 + \delta'_j \delta_j = 1 + \theta \Lambda'_j \Lambda_j$ with Λ_j being the j -th column of Λ . Consider the equation

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln L = 0 \Rightarrow \\ \frac{np}{\theta} - n \sum_{j=1}^r \frac{\Lambda'_j \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \text{tr}(S) \\ + 2\theta \sum_{j=1}^r \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \theta^2 \sum_{j=1}^r \frac{\Lambda'_j \Lambda_j}{(1 + \theta \Lambda'_j \Lambda_j)^2} \Lambda'_j S \Lambda_j = 0. \end{aligned} \quad (11.4.14)$$

For a specific j , we have

$$\begin{aligned} \frac{\partial}{\partial \Lambda_j} \ln L = 0 \Rightarrow \\ -\frac{n}{2} \frac{2\theta \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} + \frac{\theta^2}{2} \frac{2S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \frac{\theta^2}{2} \frac{\Lambda'_j S \Lambda_j}{[1 + \theta \Lambda'_j \Lambda_j]^2} (2\theta) \Lambda_j = 0. \end{aligned} \quad (11.4.15)$$

Pre-multiplying (11.4.15) by Λ'_j yields

$$-\frac{n\theta \Lambda'_j \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} + \frac{\theta^2 \Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \frac{\theta^3 (\Lambda'_j S \Lambda_j) (\Lambda'_j \Lambda_j)}{(1 + \theta \Lambda'_j \Lambda_j)^2} = 0. \quad (11.4.16)$$

Now, on comparing (11.4.16) with (11.4.14) after multiplying (11.4.14) by θ , we have

$$np - \theta \text{tr}(S) + \theta^2 \sum_{j=1}^r \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} = 0. \quad (11.4.17)$$

Multiplying the left-hand side of (11.4.15) by $[1 + \theta \Lambda'_j \Lambda_j]^2 / \theta$ gives

$$-n(1 + \theta \Lambda'_j \Lambda_j) \Lambda_j + \theta(1 + \theta \Lambda'_j \Lambda_j) S \Lambda_j - \theta^2 (\Lambda'_j S \Lambda_j) \Lambda_j = 0. \quad (11.4.18)$$

Then, by pre-multiplying (11.4.18) by Λ'_j , we obtain

$$\begin{aligned} -n(1 + \theta \Lambda'_j \Lambda_j) \Lambda'_j \Lambda_j + \theta[(1 + \theta \Lambda'_j \Lambda_j) \Lambda'_j S \Lambda_j - \theta^2 (\Lambda'_j S \Lambda_j) \Lambda'_j \Lambda_j] = 0 \Rightarrow \\ \theta (\Lambda'_j S \Lambda_j) = n(1 + \theta \Lambda'_j \Lambda_j) (\Lambda'_j \Lambda_j), \end{aligned} \quad (11.4.19)$$

which provides the following representation of θ :

$$\theta = \frac{n \Lambda'_j \Lambda_j}{\Lambda'_j S \Lambda_j - n (\Lambda'_j \Lambda_j)^2} \text{ for } \Lambda'_j S \Lambda_j - n (\Lambda'_j \Lambda_j)^2 > 0 \quad (i)$$

since θ must be positive. Further, on substituting $\theta \Lambda_j' S \Lambda_j$ $\stackrel{(11.4.19)}{=} n(1 + \theta \Lambda_j' \Lambda_j)(\Lambda_j' \Lambda_j)$ in (11.4.18), we have

$$-n\Lambda_j + \theta[S - n(\Lambda_j' \Lambda_j)I]\Lambda_j = O, \quad (ii)$$

which, on replacing θ by the right-hand side of (i) yields

$$-(\Lambda_j' S \Lambda_j)\Lambda_j + (\Lambda_j' \Lambda_j)S\Lambda_j = O$$

or, equivalently,

$$\left[S - \frac{\Lambda_j' S \Lambda_j}{\Lambda_j' \Lambda_j} I \right] \Lambda_j = O \Rightarrow \quad (11.4.20)$$

$$[S - \lambda_j I]\Lambda_j = O \quad (11.4.21)$$

where $\lambda_j = \frac{\Lambda_j' S \Lambda_j}{\Lambda_j' \Lambda_j}$. Observe that Λ_j in (11.4.20) is an eigenvector of S for $j = 1, \dots, p$. Substituting the value of $\Lambda_j' S \Lambda_j$ from (11.4.19) into (11.4.17) gives the following estimate of θ :

$$\hat{\theta} = \frac{np}{\text{tr}(S) - n \sum_{j=1}^p \hat{\Lambda}_j' \hat{\Lambda}_j} \quad (11.4.22)$$

whenever the denominator is positive as θ is by definition positive. Now, in light of (11.4.18) and (11.4.19), we can also obtain the following result for each j :

$$\hat{\theta} = \frac{n\hat{\Lambda}_j' \hat{\Lambda}_j}{\hat{\Lambda}_j' S \hat{\Lambda}_j - n(\hat{\Lambda}_j' \hat{\Lambda}_j)^2}, \quad j = 1, \dots, p, \quad (11.4.23)$$

requiring again that the denominator be positive. In this case, $\hat{\Lambda}_j$ is an eigenvector of S for $j = 1, \dots, p$. Let us conveniently normalize the $\hat{\Lambda}_j$'s so that the denominator in (11.4.22) and (11.4.23) remain positive.

Thus, $\hat{\Lambda}_j$ is an eigenvector of S with the corresponding eigenvalue λ_j for each j , $j = 1, \dots, p$. Out of these, the first r of them, corresponding to the r largest eigenvalues, will also be estimates for the factor loadings Λ_j , $j = 1, \dots, r$. Observe that we can multiply $\hat{\Lambda}_j$ by any constant c_1 without affecting equations (11.4.20) or (11.4.21). This constant c_1 may become necessary to keep the denominators in (11.4.22) and (11.4.23) positive. Hence we have the following result:

Theorem 11.4.2. *The sum of all the eigenvalues of S from Eq. (11.4.20), including the estimates of the r factor loadings $\hat{\Lambda}_1, \dots, \hat{\Lambda}_p$, is given by*

$$\sum_{j=1}^p \frac{\hat{\Lambda}'_j S \hat{\Lambda}_j}{\hat{\Lambda}'_j \hat{\Lambda}_j} = \text{tr}(S). \quad (11.4.24)$$

It can be established that the representations of $\hat{\theta}$ given by (11.4.22) and (11.4.23) are one and the same. The equation giving rise to (11.4.23) is

$$\theta[\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2] = n\Lambda'_j \Lambda_j \text{ for each } j. \quad (iii)$$

Let us divide both sides of (iii) by $\Lambda'_j \Lambda_j$. Observe that $\frac{\Lambda'_j S \Lambda_j}{\Lambda'_j \Lambda_j} = \lambda_j$ is an eigenvalue of S for $j = 1, \dots, p$, treating Λ_j as an eigenvector of S . Now, taking the sum over $j = 1, \dots, p$, on both sides of (iii) after dividing by $\Lambda'_j \Lambda_j$, we have

$$\begin{aligned} \theta \left[\sum_{j=1}^p \lambda_j - n \sum_{j=1}^p (\Lambda'_j \Lambda_j) \right] &= np \Rightarrow \\ \theta \left[\text{tr}(S) - n \sum_{j=1}^p \Lambda'_j \Lambda_j \right] &= np, \end{aligned} \quad (iv)$$

which is Eq. (11.4.22). This proves the claim.

Hence the procedure is the following: Compute the eigenvalues and the corresponding eigenvectors of the sample sum of products matrix S . The estimates for the factor loadings, denoted by $\hat{\Lambda}_j$, are available from the eigenvectors $\hat{\Lambda}_j$ of S after appropriate normalization to make the denominators in (11.4.22) and (1.4.23) positive. Take the first r largest eigenvalues of S and then compute the corresponding eigenvectors to obtain estimates for all the factor loadings. This methodology is clearly related to that utilized in Principal Component Analysis, the estimates of the variances of the principal components being $\hat{\Lambda}'_j S \hat{\Lambda}_j / \hat{\Lambda}'_j \hat{\Lambda}_j$ for $j = 1, \dots, r$.

Verification

Does the representation of θ given in (11.4.22) and (11.4.23) satisfy the likelihood Eq. (11.4.14)? Since θ is estimated through Λ_j for each $j = 1, \dots, p$, we may replace θ in (11.4.14) by θ_j and insert the summation symbol. Equation (11.4.14) will then be

$$\begin{aligned}
& n \sum_j \frac{1}{\theta_j} - n \sum_j \frac{\Lambda_j' \Lambda_j}{1 + \theta_j \Lambda_j' \Lambda_j} - \text{tr}(S) \\
& + 2 \sum_j \theta_j \frac{\Lambda_j' S \Lambda_j}{1 + \theta_j \Lambda_j' \Lambda_j} - \sum_j \theta_j^2 \frac{\Lambda_j' \Lambda_j (\Lambda_j' S \Lambda_j)}{(1 + \theta_j \Lambda_j' \Lambda_j)^2} = 0. \quad (11.4.25)
\end{aligned}$$

Now, substituting the value of θ_j specified in (11.4.23) into (11.4.14), the left-hand side of (11.4.14) reduces to the following:

$$\begin{aligned}
& n \sum_j \frac{[\Lambda_j' S \Lambda_j - n(\Lambda_j' \Lambda_j)^2]}{n \Lambda_j' \Lambda_j} - n \sum_j \frac{\Lambda_j' \Lambda_j}{\Lambda_j' S \Lambda_j} [\Lambda_j' S \Lambda_j - n(\Lambda_j' \Lambda_j)^2] - \text{tr}(S) \\
& + 2 \sum_j n \Lambda_j' \Lambda_j - \sum_j \frac{\Lambda_j' \Lambda_j}{\Lambda_j' S \Lambda_j} (n \Lambda_j' \Lambda_j)^2 \\
& = \sum_j \frac{\Lambda_j' S \Lambda_j}{\Lambda_j' \Lambda_j} - \text{tr}(S) = 0,
\end{aligned}$$

owing to Theorem 11.4.2. Hence, Eq. (11.4.14) holds for the value of θ given in (11.4.23) and the value of Λ_j specified in (11.4.20).

Since the basic estimating equation for $\hat{\theta}$ arises from (11.4.23) as

$$\theta [\Lambda_j' S \Lambda_j - n(\Lambda_j' \Lambda_j)^2] = n \Lambda_j' \Lambda_j, \quad (v)$$

a combined estimate for θ can be secured. On dividing both sides of (v) by $\Lambda_j' \Lambda_j$ and summing up over j , $j = 1, \dots, p$, it is seen that the resulting estimate of θ agrees with that given in (11.4.22).

11.4.4. Maximum value of the exponent

We have the estimate $\hat{\theta}$ of θ provided in (11.4.23) at the estimated value $\hat{\Lambda}_j$ of Λ_j for each j , where $\hat{\Lambda}_j$ is an eigenvector of S resulting from (11.4.20). The exponent of the likelihood function is $-\frac{1}{2} \text{tr}(\Sigma^{-1} S)$ and, in the current context, $\Sigma = \Lambda \Phi \Lambda' + \Psi$, the identification conditions being that $\Phi = I_p$ and $\Lambda' \Psi^{-1} \Lambda$ be a diagonal matrix with positive diagonal elements. Under these conditions and for the special case $\Psi = \sigma^2 I_p$ with $\sigma^{-2} = \theta$, we have shown that the exponent in the log-likelihood function reduces to $-\frac{1}{2} \theta \text{tr}(S) + \frac{1}{2} \theta^2 \sum_{j=1}^r \frac{\Lambda_j' S \Lambda_j}{1 + \theta \Lambda_j' \Lambda_j}$. Now, consider $\theta \text{tr}(S) - \sum_{j=1}^r \theta^2 \frac{\Lambda_j' S \Lambda_j}{1 + \theta \Lambda_j' \Lambda_j} \equiv \delta$. Then,

$$\begin{aligned}
\delta &= \theta \operatorname{tr}(S) - \theta \sum_{j=1}^r \theta \frac{\Lambda_j' S \Lambda_j}{1 + \theta \Lambda_j' \Lambda_j} \\
&= \theta \left[\operatorname{tr}(S) - \sum_{j=1}^p \theta \frac{\Lambda_j' S \Lambda_j}{1 + \theta \Lambda_j' \Lambda_j} \right] \\
&= \theta \left[\operatorname{tr}(S) - \sum_{j=1}^p n \Lambda_j' \Lambda_j \right] \text{ from (11.4.19)} \\
&= np \text{ from (11.4.22)}.
\end{aligned}$$

Hence the result.

Example 11.4.1. Tests are conducted to evaluate x_1 : verbal-linguistic skills, x_2 : spatial visualization ability, and x_3 : mathematical abilities. Test scores on x_1 , x_2 , and x_3 are available. It is known that these abilities are governed by two intellectual faculties that will be identified as f_1 and f_2 , and that linear functions of f_1 and f_2 are contributing to x_1 , x_2 , x_3 . These coefficients in the linear functions, known as factor loadings, are unknown. Let $\Lambda = (\lambda_{ij})$ be the matrix of factor loadings. Then, we have the model

$$\begin{aligned}
x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \mu_1 + e_1 \\
x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \mu_2 + e_2 \\
x_3 &= \lambda_{31}f_1 + \lambda_{32}f_2 + \mu_3 + e_3.
\end{aligned}$$

Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

where M is some general effect, ϵ is the error vector or the sum total of the contributions from unknown factors, F represents the vector of contributing factors and Λ , the levels of the contributions. Let $\operatorname{Cov}(X) = \Sigma$, $\operatorname{Cov}(F) = \Phi$ and $\operatorname{Cov}(\epsilon) = \Psi$. Under the assumptions $\Phi = I$, $\Psi = \sigma^2 I$ where σ^2 is a real scalar quantity, and I is the identity matrix, and $\Lambda' \Psi^{-1} \Lambda$ is diagonal, estimate the factor loadings λ_{ij} 's and σ^2 in Ψ . A battery of tests are conducted on a random sample of six individuals and the following are the data, where our notations are \mathbf{X} : the matrix of sample values, $\bar{\mathbf{X}}$: the sample average, $\bar{\mathbf{X}}$: the matrix of sample averages, and S : the sample sum of products matrix. So, letting

$$\mathbf{X} = [X_1, X_2, \dots, X_6] = \begin{bmatrix} 4 & 2 & 4 & 2 & 4 & 2 \\ 4 & 2 & 1 & 2 & 1 & 2 \\ 2 & 5 & 3 & 3 & 3 & 2 \end{bmatrix},$$

estimate the factor loadings λ_{ij} 's and the variance σ^2 .

Solution 11.4.1. We begin with the computations of the various quantities required to arrive at a solution. Observe that since the matrix of factor loadings A is 3×2 and a random sample of 6 observation vectors is available, $n = 6$, $p = 3$ and $r = 2$ in our notation.

In this case,

$$\bar{X} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \quad \bar{\mathbf{X}} = [\bar{X}, \bar{X}, \dots, \bar{X}],$$

$$\mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 2 & 0 & -1 & 0 & -1 & 0 \\ -1 & 2 & 0 & 0 & 0 & -1 \end{bmatrix},$$

$$[\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]' = S = \begin{bmatrix} 6 & 0 & -2 \\ 0 & 6 & -2 \\ -2 & -2 & 6 \end{bmatrix}.$$

An estimator/estimate of Σ is $\hat{\Sigma} = \frac{S}{n}$ whereas an unbiased estimator/estimate of Σ is $\frac{S}{n-1}$. An eigenvalue of $\frac{S}{\alpha}$ is $\frac{1}{\alpha}$ times the corresponding eigenvalue of S . Moreover, constant multiples of eigenvectors are also eigenvectors for a given eigenvalue. Accordingly, we will work with S instead of $\hat{\Sigma}$ or an unbiased estimate of Σ . Since

$$\begin{vmatrix} 6 - \lambda & 0 & -2 \\ 0 & 6 - \lambda & -2 \\ -2 & -2 & 6 - \lambda \end{vmatrix} = 0 \Rightarrow (6 - \lambda)(\lambda^2 - 12\lambda + 28) = 0,$$

the eigenvalues are $\lambda_1 = 6 + \sqrt{8}$, $\lambda_2 = 6$, $\lambda_3 = 6 - \sqrt{8}$, the two largest ones being $\lambda_1 = 6 + \sqrt{8}$ and $\lambda_2 = 6$. Let us evaluate the eigenvectors U_1 , U_2 and U_3 associated with these three eigenvalues. An eigenvector U_1 corresponding to $\lambda_1 = 6 + \sqrt{8}$ will be a solution of the equation

$$\begin{bmatrix} -\sqrt{8} & 0 & -2 \\ 0 & -\sqrt{8} & -2 \\ -2 & -2 & -\sqrt{8} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow U_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 6$, the equation to be solved is

$$\begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & -2 \\ -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow U_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

As for the eigenvalue $\lambda_3 = 6 - \sqrt{8}$, it is seen from the derivation of U_1 that $U_3' = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1]$. Let us now examine the denominator of $\hat{\theta}$ in (11.4.22). Observe that $U_1'U_1 = 2$, $U_2'U_2 = 2$, $U_3'U_3 = 2$ and $n \sum_{j=1}^p U_j'U_j = 6(2+2+2) = 36$. However, $\text{tr}(S) = (6 + \sqrt{8}) + (6) + (6 - \sqrt{8}) = 18$. So, let us multiply each vector by $\frac{1}{\sqrt{3}}$ so that $n(\sum_{j=1}^p U_j'U_j = 6(\frac{2}{3} + \frac{2}{3} + \frac{2}{3}) = 12$ and $\text{tr}(S) - n \sum_{j=1}^p U_j'U_j = 18 - 12 > 0$. Thus, the estimate of θ is given by

$$\hat{\theta} = \frac{np}{\text{tr}(S) - n \sum_{j=1}^p U_j'U_j} = \frac{(6)(3)}{18 - 12} = 3 \Rightarrow \hat{\sigma}^2 = \frac{1}{3}$$

In light of (11.4.20), the factor loadings are estimated by U_1 and U_2 scaled by $\frac{1}{\sqrt{3}}$. Hence, the estimates of the factor loadings, denoted with a hat, are the following: $\hat{\lambda}_{11} = (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{2}}) = -\frac{1}{\sqrt{6}}$, $\hat{\lambda}_{21} = (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{2}}) = -\frac{1}{\sqrt{6}}$, $\hat{\lambda}_{31} = (\frac{1}{\sqrt{3}})(1) = \frac{1}{\sqrt{3}}$, $\hat{\lambda}_{12} = (\frac{1}{\sqrt{3}})(1) = \frac{1}{\sqrt{3}}$, $\hat{\lambda}_{22} = (\frac{1}{\sqrt{3}})(-1) = -\frac{1}{\sqrt{3}}$, $\hat{\lambda}_{32} = 0$.

11.5. General Case

Let

$$\delta_j = \begin{bmatrix} \theta_1 \lambda_{1j} \\ \theta_2 \lambda_{2j} \\ \vdots \\ \theta_p \lambda_{pj} \end{bmatrix}, \quad \delta_j' \delta_j = \Lambda_j' \Theta^2 \Lambda_j, \quad \Lambda_j = \begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \vdots \\ \lambda_{pj} \end{bmatrix},$$

$$\Psi^{-1} = \Theta^2 = \begin{bmatrix} \theta_1^2 & 0 & \dots & 0 \\ 0 & \theta_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_p^2 \end{bmatrix} \quad \text{and} \quad |I + \Lambda' \Theta^2 \Lambda| = \prod_{j=1}^r (1 + \delta_j' \delta_j).$$

We will take δ_j , $j = 1, \dots, r$, and $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_p)$ as the parameters. Expressed in terms of the δ_j 's and Θ , the log-likelihood function is the following:

$$\begin{aligned} \ln L = & -\frac{np}{2} \ln(2\pi) + n \sum_{j=1}^p \ln \theta_j - \frac{n}{2} \sum_{j=1}^r \ln(1 + \delta_j' \delta_j) \\ & - \frac{1}{2} \text{tr}(\Theta^2 S) + \frac{1}{2} \sum_{j=1}^r \frac{1}{1 + \delta_j' \delta_j} (\delta_j' \Theta S \Theta \delta_j). \end{aligned} \quad (11.5.1)$$

Let us take δ_j , $j = 1, \dots, r$, and Θ as the parameters. Differentiating $\ln L$ partially with respect to the vector δ_j , for a specific j , and equating the result to a null vector, we have the following (referring to Chap. 1 for vector/matrix derivatives):

$$-\frac{n}{2} \frac{2\delta_j}{1 + \delta'_j \delta_j} - \frac{1}{2} (\delta'_j \Theta S \Theta \delta_j) \frac{2\delta_j}{(1 + \delta'_j \delta_j)^2} + \frac{1}{2} \frac{2 \Theta S \Theta \delta_j}{1 + \delta'_j \delta_j} = O, \quad (i)$$

which multiplied by $1 + \delta'_j \delta_j > 0$, yields

$$-n\delta_j - \frac{(\delta'_j \Theta S \Theta \delta_j)}{1 + \delta'_j \delta_j} \delta_j + (\Theta S \Theta) \delta_j = O. \quad (11.5.2)$$

On premultiplying (11.5.2) by δ'_j and then by $1 + \delta'_j \delta_j$, and simplifying, we then have

$$\begin{aligned} -n(1 + \delta'_j \delta_j)(\delta'_j \delta_j) + (\delta'_j \Theta S \Theta \delta_j) &= O \Rightarrow \\ \frac{\delta'_j \Theta S \Theta \delta_j}{1 + \delta'_j \delta_j} &= n\delta'_j \delta_j. \end{aligned} \quad (11.5.3)$$

Let us differentiate $\ln L$ as given in (11.5.1) partially with respect to θ_j for a specific j such as $j = 1$. Then,

$$\frac{n}{\theta_j} - \frac{1}{2} 2\theta_j s_{jj} + \frac{1}{2} \sum_{j=1}^r \frac{1}{1 + \delta'_j \delta_j} \frac{\partial}{\partial \theta_j} (\delta'_j \Theta S \Theta \delta_j),$$

where

$$\begin{aligned} \frac{\partial}{\partial \theta_j} (\delta'_j \Theta S \Theta \delta_j) &= \delta'_j \left[\frac{\partial}{\partial \theta_j} \Theta \right] S \Theta \delta_j + \delta'_j \Theta S \left[\frac{\partial}{\partial \theta_j} \Theta \right] \delta_j, \\ \frac{\partial}{\partial \theta_1} \Theta &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \Rightarrow \theta_1 \frac{\partial}{\partial \theta_1} \Theta = \begin{bmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \end{aligned}$$

so that

$$\left[\theta_1 \frac{\partial}{\partial \theta_1} + \dots + \theta_p \frac{\partial}{\partial \theta_p} \right] \Theta = \Theta. \quad (ii)$$

Hence,

$$\left[\sum_{j=1}^p \theta_j \frac{\partial}{\partial \theta_j} \right] \sum_{j=1}^r \frac{\delta'_j \Theta S \Theta \delta_j}{1 + \delta'_j \delta_j} = 2 \sum_{j=1}^r \frac{\delta'_j \Theta S \Theta \delta_j}{1 + \delta'_j \delta_j},$$

and then,

$$\left[\sum_{j=1}^p \theta_j \frac{\partial}{\partial \theta_j} \right] L = 0 \Rightarrow$$

$$np - \sum_{j=1}^p \theta_j^2 s_{jj} + \sum_{j=1}^r \frac{\delta'_j \Theta S \Theta \delta_j}{1 + \delta'_j \delta_j} = 0. \quad (11.5.4)$$

However, given (11.5.3), we have $n\delta'_j \delta_j = \delta'_j \Theta S \Theta \delta_j / (1 + \delta'_j \delta_j)$, $j = 1, \dots, r$, and therefore (11.5.4) can be expressed as

$$np - \sum_{j=1}^p \theta_j^2 s_{jj} + \sum_{j=1}^r n\delta'_j \delta_j = 0. \quad (iii)$$

Letting

$$c = \frac{1}{p} \sum_{j=1}^r \delta'_j \delta_j, \quad (11.5.5)$$

equation(iii) can be written as

$$\sum_{j=1}^p [n(1+c) - \theta_j^2 s_{jj}] = 0, \quad c = 0, \quad j = r+1, \dots, p,$$

so that a solution for θ_j is

$$\hat{\theta}_j^2 = \frac{n(1+c)}{s_{jj}} \quad \text{or} \quad \hat{\sigma}_j^2 = \frac{s_{jj}}{n(1+c)}, \quad (11.5.6)$$

with the proviso that $c = 0$ for $\hat{\theta}_j^2$ and $\hat{\sigma}_j^2$, $j = r+1, \dots, p$. Then, an estimate of $\Theta S \Theta$ is given by

$$\hat{\Theta} S \hat{\Theta} = n(1+c) \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix} S \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{s_{pp}}} \end{bmatrix}$$

$$= n(1+c)R \quad (11.5.7)$$

where R is the sample correlation matrix, and on applying the identities (11.5.3) and (11.5.7), (11.5.2) becomes

$$-n\delta_j - n(\delta_j'\delta_j)\delta_j + (n(1+c)R)\delta_j = 0 \Rightarrow$$

$$\left[R - \frac{1 + \hat{\delta}_j'\hat{\delta}_j}{1+c} I \right] \delta_j = O. \quad (11.5.8)$$

This shows that δ_j is an eigenvector of R . If ν_j is an eigenvalue of R , then the largest r eigenvalues are of the form

$$\nu_j = \frac{1 + \hat{\delta}_j'\hat{\delta}_j}{1+c}, \quad j = 1, \dots, r, \quad (11.5.9)$$

and the remaining ones are ν_{r+1}, \dots, ν_p , where c is as specified in (11.5.5). Thus, the procedure is the following: Compute the eigenvalues ν_j , $j = 1, \dots, p$, of R and determine the corresponding eigenvectors, denoted by δ_j , $j = 1, \dots, p$. The first r of them which correspond to the r largest ν_j 's, are $\hat{\delta}_j = \hat{\Theta} \hat{\Lambda}_j \Rightarrow \hat{\Lambda}_j = \hat{\Theta}^{-1} \hat{\delta}_j$, $j = 1, \dots, r$. Let

$$\hat{\delta}_j = \begin{bmatrix} \hat{\delta}_{1j} \\ \hat{\delta}_{2j} \\ \vdots \\ \hat{\delta}_{pj} \end{bmatrix}, \quad \hat{\Lambda}_j = \begin{bmatrix} \hat{\lambda}_{1j} \\ \hat{\lambda}_{2j} \\ \vdots \\ \hat{\lambda}_{pj} \end{bmatrix} \quad \text{and} \quad \hat{\Theta}^2 = \begin{bmatrix} \frac{n(1+c)}{s_{11}} & 0 & \dots & 0 \\ 0 & \frac{n(1+c)}{s_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{n(1+c)}{s_{pp}} \end{bmatrix}. \quad (11.5.10)$$

Then,

$$\hat{\lambda}_{ij} = \frac{\sqrt{s_{jj}}}{\sqrt{n(1+c)}} \hat{\delta}_{ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, r, \quad (11.5.11)$$

and $\hat{\Theta}$ is available from (11.5.10). All the model parameters have now been estimated.

11.5.1. The exponent in the likelihood function

Given the MLE's of the parameters which are available from (11.5.6), (11.5.8), (11.5.10) and (11.5.11), what will be the maximum value of the likelihood function? Let us examine its exponent:

$$-\frac{1}{2} \text{tr}(\hat{\Theta}^2 S) + \frac{n}{2} \sum_{j=1}^r \hat{\delta}_j' \hat{\delta}_j = -\frac{1}{2} n(1+c) \text{tr}(R) + \frac{n}{2} pc$$

$$= -\frac{1}{2} n(1+c)p + \frac{1}{2} npc = -\frac{np}{2},$$

that is, the same value of the exponent that is obtained under a general Σ . The estimates were derived under the assumptions that $\Sigma = \Psi + \Lambda\Lambda'$, $\Lambda'\Psi^{-1}\Lambda$ is diagonal with the diagonal elements $\delta'_j\delta_j$, $j = 1, \dots, r$, and $\Psi = \Theta^{-2}$.

Example 11.5.1. Using the data set provided in Example 11.4.1, estimate the factor loadings and the diagonal elements of $\text{Cov}(\epsilon) = \Psi = \text{diag}(\psi_{11}, \dots, \psi_{pp})$. In this example, $p = 3$, $n = 6$, $r = 2$.

Solution 11.5.1. We will adopt the same notations and make use of some of the computational results already obtained in the previous solution. First, we need to compute the eigenvalues of the sample correlation matrix R . The sample sum of products matrix S is given by

$$S = \begin{bmatrix} 6 & 0 & -2 \\ 0 & 6 & -2 \\ -2 & -2 & 6 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & 0 & -\frac{2}{6} \\ 0 & 1 & -\frac{2}{6} \\ -\frac{2}{6} & -\frac{2}{6} & 1 \end{bmatrix} = \frac{1}{6}S.$$

Hence, the eigenvalues of R are $\frac{1}{6}$ times the eigenvalues of S , that is, $v_1 = \frac{1}{6}(6 + \sqrt{8}) = 1 + \frac{\sqrt{2}}{3}$, $v_2 = \frac{1}{6}(6) = 1$ and $v_3 = 1 - \frac{\sqrt{2}}{3}$. Since $\frac{1}{6}$ will be canceled when determining the eigenvectors, the eigenvectors of S will coincide with those of R . They are the following, denoted again by δ_j , $j = 1, 2, 3$:

$$\delta_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \delta_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}.$$

Therefore, $\delta'_1\delta_1 = 2$, $\delta'_2\delta_2 = 2$, $\delta'_3\delta_3 = 2$ and c as defined in (11.5.5) is $c = \frac{1}{3}(2 + 2) = \frac{4}{3} \Rightarrow n(1 + c) = 6(1 + \frac{4}{3}) = 14$. Then, in light of (11.5.10), the estimates of ψ_{jj} , $j = 1, 2, 3$, are available as $\hat{\psi}_{jj} = \hat{\theta}_j^{-2} = \frac{s_{jj}}{n(1+c)}$ or $\hat{\psi}_{11} = \hat{\theta}_1^{-2} = \frac{6}{14} = \frac{3}{7} = \hat{\psi}_{22} = \hat{\psi}_{33}$, denoting the estimates with a hat. Therefore, the diagonal matrix $\hat{\Psi} = \text{diag}(\frac{3}{7}, \frac{3}{7}, \frac{3}{7})$. Hence, the matrix $\Theta^{-1} = \Psi^{\frac{1}{2}}$ is estimated by $\hat{\Theta}^{-1} = \text{diag}(\frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}})$. From (11.5.11), $\hat{\Lambda}_j = \hat{\Theta}^{-1}\delta_j = \text{diag}(\frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}})\delta_j$, that is, δ_j is pre-multiplied by $\frac{\sqrt{3}}{\sqrt{7}}$. Therefore, the estimates of the factor loadings are: $\hat{\lambda}_{11} = -(\frac{\sqrt{3}}{\sqrt{7}})(\frac{1}{\sqrt{2}}) = \hat{\lambda}_{21} = -\hat{\lambda}_{13} = -\hat{\lambda}_{23}$, $\hat{\lambda}_{31} = \frac{\sqrt{3}}{\sqrt{7}} = \hat{\lambda}_{33} = \hat{\lambda}_{12} = -\hat{\lambda}_{22}$, and $\hat{\lambda}_{32} = 0$.

11.6. Tests of Hypotheses

The usual test in connection with the current topic consists of assessing identifiability, that is, testing the hypothesis H_0 that the population covariance matrix $\Sigma > O$ can be represented as $\Sigma = \Lambda\Phi\Lambda' + \Psi$ when $\Phi = I$, $\Lambda'\Psi^{-1}\Lambda$ is a diagonal matrix with positive diagonal elements, $\Psi > O$ is a diagonal matrix and $\Lambda = (\lambda_{ij})$ is a $p \times r$, $r \leq p$, matrix of full rank r , whose elements are the factor loadings. That is,

$$H_0 : \Sigma = \Lambda\Lambda' + \Psi. \quad (11.6.1)$$

In this instance, a crucial aspect of the hypothesis H_0 consisting of determining whether “the model fits”, is that the number r be designated since the other quantities n , the sample size, and p , the order of the observation vector, are preassigned. Thus, the phrase “model fits” means that for a given r , Σ can be expressed in the form $\Sigma = \Psi + \Lambda\Lambda'$, in addition to satisfying the identification conditions. The assumed model has the representation: $X = M + \Lambda F + \epsilon$ where $X' = (x_1, \dots, x_p)$ stands for the $p \times 1$ vector of observed scores on p tests or p batteries of tests, M is a $p \times 1$ vector of general effect, F is an $r \times 1$ vector of unknown factors, $\Lambda = (\lambda_{ij})$ is the unknown $p \times r$ matrix of factor loadings and ϵ is the $p \times 1$ error vector. When ϵ and F are uncorrelated, the covariance matrix of X is given by

$$\Sigma = \Lambda\Phi\Lambda' + \Psi$$

where $\Phi = \text{Cov}(F) > O$ and $\Psi = \text{Cov}(\epsilon) > O$ with Φ being $r \times r$ and Ψ being $p \times p$ and diagonal. A simple random sample from X will be taken to mean a sample of independently and identically distributed (iid) $p \times 1$ vectors $X'_j = (x_{1j}, x_{2j}, \dots, x_{pj})$, $j = 1, \dots, n$, with n denoting the sample size. The sample sum of products matrix or “corrected” sample sum of squares and cross products matrix is $S = (s_{ij})$, $s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$, where, for example, the average of the x_i 's comprising the i -th row of $\mathbf{X} = [X_1, \dots, X_n]$, namely, \bar{x}_i , is $\bar{x}_i = \sum_{k=1}^n x_{ik}/n$. If ϵ and F are independently normally distributed, then the likelihood ratio criterion or λ -criterion is

$$\lambda = \frac{\max_{H_0} L}{\max L} = \frac{|\hat{\Sigma}|^{\frac{n}{2}}}{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|^{\frac{n}{2}}} \Rightarrow w = \lambda^{\frac{2}{n}} = \frac{|\hat{\Sigma}|}{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|} \quad (11.6.2)$$

where $\hat{\Sigma} = \frac{1}{n}S$ and the covariance matrix $\Sigma = \Lambda\Lambda' + \Psi$ under H_0 , with $\Phi = \text{Cov}(F)$ assumed to be an identity matrix and the $r \times r$ matrix $\Lambda'\Psi^{-1}\Lambda = \text{diag}(\delta'_1\delta_1, \dots, \delta'_r\delta_r)$ having positive diagonal elements $\delta'_j\delta_j$, $j = 1, \dots, r$. Referring to Sect. 11.4.2, we have

$$|\Lambda\Lambda' + \Psi| = |\Psi| |\Lambda'\Psi^{-1}\Lambda + I| \quad (11.6.3)$$

and $1 + \delta'_j \delta_j = 1 + \Lambda'_j \Psi^{-1} \Lambda_j = 1 + \Lambda'_j \Theta^2 \Lambda_j$ where $\delta_j = \Psi^{-\frac{1}{2}} \Lambda_j = \Theta \Lambda_j$ and Λ_j is the j -th column of Λ for $j = 1, \dots, r$. It was shown in (11.5.8) that δ_j is an eigenvector of the sample correlation matrix R and

$$\prod_{j=1}^r (1 + \delta'_j \delta_j) = |\Lambda' \Psi^{-1} \Lambda + I|.$$

However, in view of the discussion following (11.5.8), an eigenvalue of R is of the form $v_j = \frac{1 + \delta'_j \delta_j}{(1+c)}$, $j = 1, \dots, r$. Let v_1, \dots, v_p be the eigenvalues of R and let the largest r of them be v_1, \dots, v_r . It also follows from (11.5.8) that $\hat{\Theta}^2 = n(1+c)\text{diag}(\frac{1}{s_{11}}, \dots, \frac{1}{s_{pp}})$ with $\Theta = \Psi^{-\frac{1}{2}}$. Thus,

$$\begin{aligned} \frac{|\hat{\Sigma}|}{|\hat{\Psi}|} &= \left| \hat{\Theta} \left(\frac{S}{n} \right) \hat{\Theta} \right| = |(1+c)R| = \left\{ \prod_{j=1}^r (1+c)v_j \right\} (1+0)^{p-r} v_{r+1} \cdots v_p \\ &= \left\{ \prod_{j=1}^r (1 + \delta'_j \delta_j) \right\} v_{r+1} \cdots v_p \\ &\Rightarrow \frac{|\frac{1}{n}S|}{|\hat{\Psi}| |\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} + I|} = \frac{\{\prod_{j=1}^r (1 + \delta'_j \delta_j)\} v_{r+1} \cdots v_p}{\{\prod_{j=1}^r (1 + \delta'_j \delta_j)\}} \\ &= v_{r+1} \cdots v_p = w = \lambda^{\frac{2}{n}}. \end{aligned} \tag{11.6.4}$$

Hence, we reject the null hypothesis for small values of the product $v_{r+1} \cdots v_p$, that is, the product of the smallest $p - r$ eigenvalues of the sample correlation matrix R . In order to evaluate critical points, one would require the null distribution of the product of the eigenvalues, $v_{r+1} \cdots v_p$, which is difficult to determine for a general p . How can rejecting the null hypothesis that the “model fits” be interpreted? Since, in the whole structure, the decisive quantity is r , we are actually rejecting the hypothesis that a given r is the number of main factors contributing to the observations. Hence, we may seek a larger or smaller r , keeping the structure unchanged and testing the same hypothesis again until the hypothesis is not rejected. We may then assume that the r specified at the last stage is the number of main factors contributing to the observation or we may assert that, with that particular r , there is evidence that the model fits.

We will now determine conditions ensuring that the likelihood ratio criterion λ be less than or equal to one. While, assuming that $\Lambda' \Psi^{-1} \Lambda$ is diagonal, the left-hand side of the deciding equation, $\Sigma = \Lambda \Lambda' + \Psi$, has $p(p+1)/2$ parameters, there are $p r + p - r(r-1)/2$

conditions on the right-hand side where $r(r - 1)/2$ arises from the diagonality condition. The difference is then

$$\frac{p(p + 1)}{2} - \left[pr + p - \frac{r(r - 1)}{2} \right] = \frac{1}{2}[(p - r)^2 - (p + r)] \equiv \rho. \quad (11.6.5)$$

This ρ depends upon the parameters p and r , whereas λ depends upon p, r and c . Thus, λ may not be ≤ 1 . In order to make $\lambda \leq 1$, we can make c close to 0 by multiplying the $\hat{\delta}_j$'s by a constant, observing that this is always possible because the $\hat{\delta}_j$'s are the eigenvectors of R . By selecting a constant m and taking the new $\hat{\delta}_j$ as $\frac{1}{\sqrt{m}}\delta_j$, c can be made close to 0 and λ will be ≤ 1 , so that rejecting the null hypothesis for small values of λ will make sense. It may so happen that there will not be any parameter left to be restricted by the hypothesis H_0 that "model fits". The quantity ρ appearing in (11.6.5) could then be ≤ 0 , and in such an instance, the hypothesis would not make sense and could not be tested.

The density of the sample correlation matrix R is provided in Example 1.25 of Mathai (1997, p. 58). Denoting this density by $f(R)$, it is the following for the population covariance matrix Σ in a parent $N_p(\mu, \Sigma)$ population with Σ being a positive definite diagonal matrix, as was assumed in Sect. 11.6:

$$f(R) = \frac{[\Gamma(\frac{m}{2})]^p}{\Gamma_p(\frac{m}{2})} |R|^{\frac{m-p-1}{2}}, \quad R > O, \quad m = n - 1, \quad n > p, \quad (11.6.6)$$

and zero elsewhere, where n is the sample size.

11.6.1. Asymptotic distribution of the likelihood ratio statistic

For a large sample size n , $-2 \ln \lambda$ is approximately distributed as a chisquare random variable having k degrees of freedom where λ is the likelihood ratio criterion and k is the number of parameters restricted by the hypothesis H_0 . This approximation holds whenever the sample size n is large and $k \geq 1$. With ρ as defined in (11.6.5), we have

$$k = \rho = \frac{1}{2}[(p - r)^2 - (p + r)]. \quad (11.6.7)$$

However, $p - r = 1$ and $p + r = 5$ in the illustrative example, so that $k = -2$. Accordingly, even if the sample size n were large, this asymptotic result would not be applicable.

11.6.2. How to decide on the number r of main factors?

The structure of the population covariance matrix Σ under the model $X_j = M + \Lambda F + E_j$, $j = 1, \dots, n$, is

$$\Sigma = \Lambda \Phi \Lambda' + \Psi \Rightarrow \Sigma = \Lambda \Lambda' + \Psi \quad \text{for } \Phi = I, \quad (11.6.8)$$

where it is assumed that E_j and F are uncorrelated, $\Sigma = \text{Cov}(X_j) > O$ is $p \times p$, $\Phi = \text{Cov}(F) = I$, the $r \times r$ identity matrix, $\Psi = \text{Cov}(E_j)$ is a $p \times p$ diagonal matrix and $\Lambda = (\lambda_{ij})$ is a full rank $p \times r$, $r \leq p$, matrix whose elements are the factor loadings. Under the orthogonal factor model, $\Phi = I$. Moreover, to ensure the identification of the model, we assume that $\Lambda' \Psi^{-1} \Lambda$ is a diagonal matrix. Before initiating any data analysis, we have to assign a value to r on the basis of the data set at hand in order to set up the model. Thus, the matter of initially setting the number of main factors has to be addressed. Given

$$R = \Lambda \Lambda' + \Psi \quad (11.6.9)$$

where $\Lambda = (\lambda_{ij})$ is a $p \times r$ matrix and Ψ is a $p \times p$ diagonal matrix, does a solution that is expressible in terms of the elements of R (or those of S if S is used), exist for all λ_{ij} 's and ψ_{jj} 's? In general

$$R = \lambda_1 U_1 U_1' + \cdots + \lambda_p U_p U_p' \quad (11.6.10)$$

where the λ_j 's are the eigenvalues of R and the U_j 's are the corresponding normalized eigenvectors. Observe that $U_j U_j'$ is $p \times p$ whereas $U_j' U_j = 1$, $j = 1, \dots, p$. If $r = p$, then a solution always exists for (11.6.9). When taking $\Psi = O$, we can always let $R = B B'$ for some $p \times p$ matrix B , which can be achieved for example via a triangular decomposition. Accordingly, the relevant aspects are $r < p$ and the diagonal elements in Ψ , namely, the ψ_{jj} 's being positive. Can we then solve for all the λ_{ij} 's and ψ_{jj} 's involved in (11.6.9) in terms of the elements in R ? The answer is that a solution exists, but only when certain conditions are satisfied. Our objective is to select a value of r that is as small as possible and then, to obtain a solution to (11.6.9) in terms of the elements in R .

The analysis is to be carried out by utilizing either the sample sum of products matrix S or the sample correlation matrix R . The following are some of the guidelines for selecting r in order to set up the model.

- (1): Compute all the eigenvalues of R (or S). Let r be the number of eigenvalues ≥ 1 if the sample correlation matrix R is used. If S is used, then determine all the eigenvalues, calculate the average of these eigenvalues, and count the number of eigenvalues that are greater than or equal to this average. Take that number to be r .
- (2): Carry out a Principal Component Analysis on R (or S). If S is used, ensure that the units of measurements are not creating discrepancies. Compute the variances of these Principal Components, which are the eigenvalues of R (or S). Let λ_j , $j = 1, \dots, p$, denote these eigenvalues. Compute the ratios

$$\frac{\lambda_1 + \cdots + \lambda_m}{\lambda_1 + \cdots + \lambda_p}, \quad m = 1, 2, \dots,$$

and stop with that m for which the desired fraction of the total variation in the data is accounted for. Take that m as r . When implementing the principal component approach, the factor loadings λ_{ij} 's and the ψ_{jj} 's can be estimated as follows: From (11.6.10), write

$$R = A + B \text{ with } A = \sum_{j=1}^r \lambda_j U_j U_j' \text{ and } B = \sum_{j=r+1}^p \lambda_j U_j U_j',$$

where A can be expressed as VV' with $V = [\sqrt{\lambda_1}U_1, \dots, \sqrt{\lambda_r}U_r]$. Then, V is taken as an approximate estimate of Λ or as $\hat{\Lambda}$. Observe that $\lambda_j > 0$, $j = 1, \dots, p$. The sum over j of the i -th diagonal elements of $\lambda_j U_j U_j'$, $j = r+1, \dots, p$, will provide an estimate for ψ_{ii} , $i = 1, \dots, p$. These estimates can also be obtained as follows: Consider the estimate of σ_{ii} denoted by $\hat{\sigma}_{ii}$ which is equal to the sum of all the i -th diagonal elements in $A + B$; it will be 1 if R is used and $\hat{\sigma}_{ii}$ if S is utilized in the analysis; then, $\hat{\psi}_{ii} = \hat{\sigma}_{ii} - \sum_{j=1}^r \hat{\lambda}_{ij}$ and $\hat{\psi}_{ii}$ is now the sum of the i -th diagonal elements in B .

(iii): Consider the individual correlations in the sample correlation matrix R . Identify the largest ones in absolute value. If the largest ones occur at the (1,3)-th and (2,3)-th positions, then the factor f_3 will be deemed influential. Start with $r = 1$ (factor f_3) and carry out the analysis. Then, assess the proportion of the total variation accounted for by $\hat{\sigma}_{33}$. Should the proportion not be satisfactory, we may continue with $r = 2$. If the (2,3)-th position value is larger in absolute value than the value at the (1,3)-th position, then f_2 may be the next significant factor. Compute $\hat{\sigma}_{33} + \hat{\sigma}_{22}$ and determine the proportion to the total variation. If the resulting model is rejected, then take $r = 3$, and continue in this fashion until an acceptable proportion of the total variation is accounted for.

(iv): The maximum likelihood method. With this approach, we begin with a preselected r and test the hypothesis that, when comprising r factors, the model fits. If the hypothesis is rejected, then we let number of influential factors be $r - 1$ or $r + 1$ and continue the process of testing and deciding until the hypothesis is not rejected. That final r is to be taken as the number of main factors contributing towards the observations. The initial value of r may be determined by employing one of the methods described in (i) or (ii) or (iii).

Exercises

11.1. For the following data, where the 6 columns of the matrix represent the 6 observation vectors, verify whether $r = 2$ provides a good fit to the data. The proposed model is the Factor Analysis model $X = M + \Lambda F + \epsilon$, $F' = (f_1, f_2)$, Λ is 3×2 and of rank 2, $\text{Cov}(\epsilon) = \Psi$ is diagonal, $\text{Cov}(F) = \Phi = I$, and $\text{Cov}(X) = \Sigma > O$. The data set is

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 2 & 2 & 0 \\ -1 & -1 & 0 & -1 & -1 & -2 \end{bmatrix}.$$

11.2. For the model $X = M + \Lambda F + \epsilon$ with the conditions as specified in Exercise 11.1, verify whether the model with $r = 2$ or $r = 3$ gives a good fit on the basis of the following data, where the columns in the matrix represent five observation vectors:

$$\begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ -1 & 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 0 \end{bmatrix}.$$

11.3. Do a Principal Component Analysis in Exercise 11.1 to assess what percentage of the total variation in the data is accounted for by $r = 2$.

11.4. Do a Principal Component Analysis in Exercise 11.2 to determine what percentages of the total variation in the data are accounted for by $r = 2$ and $r = 3$.

11.5. Even though the sample sizes are not large, perform tests based on the asymptotic chisquare to assess whether the two tests there agree with the findings in Exercises 11.1 and 11.2.

11.6. Four model identification conditions are stated at the end of Sect 11.3.1. Develop λ -criteria under the conditions stated in (i): case (2); (ii): case (3), selecting your own B_1 ; (iii): case (4).

References

- E. S. Allman, C. Matias and J. A. Rhodes (2009): Identifiability of parameters in latent structure models with many observed variables, *The Annals of Statistics*, **37**, 3099–3132
- T. W. Anderson (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd ed., Wiley-Interscience, New Jersey.
- T. W. Anderson and H. Rubin (1956): Statistical inference in factor analysis, Proc. 3rd Berkeley Symposium on Mathematical Statistics and Probability, Vol. **5**, 111–150.
- Y. Chen, X. Li, and S. Zhang (2020): Structured latent factor analysis for large-scale data: identifiability, estimability, and their implications, *Journal of the American Statistical Association*, **115**, 1756–1770.

M. Ihara and Y. Kano (1995): Identifiability of full, marginal, and conditional factor analysis models, *Statistics & Probability Letters*, **23**, 343–350.

A. M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A.M. Mathai (2021): Factor analysis revisited, *C.R. Rao, 100th Birth Anniversary Volume, Statistical Theory and Practice*, <https://doi.org/10.1007/s42519-020-00160-1>

L. L. Wegge (1996): Local identifiability of the factor analysis and measurement error model parameter, *Journal of Econometrics*, **70**, 351–382.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 12

Classification Problems



12.1. Introduction

We will use the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

Historically, classification problems arose in anthropological studies. By taking a set of measurements on skeletal remains, anthropologists wanted to classify them as belonging to a certain racial group such as being of African or European origin. The measurements might have been of the following type: $x_1 =$ width of the skull, $x_2 =$ volume of the skull, $x_3 =$ length of the thigh bone, $x_4 =$ width of the pelvis, and so on. Let the measurements be represented by a $p \times 1$ vector X , with $X' = (x_1, \dots, x_p)$ where a prime denotes the transpose. Nowadays, classification procedures are employed in all types of problems occurring in various contexts. For example, consider the situation of a battery of tests in an entrance examination to admit students into a professional program such as medical sciences, law studies, engineering science or management studies. Based on the $p \times 1$

vector of test scores, a statistician would like to classify an applicant as to whether or not he/she belongs to the group of applicants who will successfully complete a given program. This is a 2-group situation. If a third category is added such as those who are expected to complete the program with flying colors, this will become a 3-group situation. In general, one will have a k -group situation when an individual is classified into one of k classes.

Let us begin with the 2-group situation. The problem consists of classifying the $p \times 1$ vector X into one of two, groups, classes or categories. Let the categories be denoted by population π_1 and population π_2 . This means X will either belong to π_1 or to π_2 , no other options being considered. The $p \times 1$ vector X may be taken as a point in a p -space R_p or p -dimensional Euclidean space \mathfrak{R}^p . In a two-group situation when it is decided that the candidate either belongs to the population π_1 or the population π_2 , two subspaces A_1 and A_2 within the p -space R_p are determined: $A_1 \subset R_p$ and $A_2 \subset R_p$, with $A_1 \cap A_2 = O$ (the empty set) or a decision rule can be symbolically written as $A = (A_1, A_2)$. If X falls in A_1 , the candidate is classified into π_1 and if X falls in A_2 , then the candidate is classified into π_2 . In other words, $X \in A_1$ means the individual is classified into population π_1 and $X \in A_2$ means that the individual is classified into population π_2 . The regions A_1 and A_2 or the rule $A = (A_1, A_2)$ are not known beforehand. These are to be determined by employing certain decision rules. Criteria for determining A_1 and A_2 will be subsequently put forward. Let us now consider the consequences. When a decision is made to classify X as coming from π_1 , either the decision is correct or the decision is erroneous. If the population is actually π_1 and the decision rule classifies X into π_1 , then the decision is correct. If X is classified into π_2 when in reality the population is π_1 , then a mistake has been committed or a misclassification occurred. Misclassification will involve penalties, costs or losses. Let such a penalty, cost or loss of classifying an individual into group i when he/she actually belongs to group j , be denoted by $C(i|j)$. In a 2-group situation, i and j can only equal 1 or 2. That is, $C(1|2) > 0$ and $C(2|1) > 0$ are the costs of misclassifying, whereas $C(1|1) = 0$ and $C(2|2) = 0$ since there is no cost or penalty associated with correct decisions. The following table summarizes this discussion:

Table 12.1: Cost of misclassification $C(i|j)$

		Statistician's decision to classify into	
		π_1	π_2
Population	π_1	0	$C(2 1)$
In reality	π_2	$C(1 2)$	0

12.2. Probabilities of Classification

The vector random variable corresponding to the observation vector X may have its own probability/density function. The real scalar variables as well as the observations on them will be denoted by the lower-case letters x_1, \dots, x_p . When dealing with the probability/density function of X , X is taken as vector random variable, whereas when looked upon as a point in the p -space, R_p , X is deemed to be an observation vector. The $p \times 1$ vector X may have a probability/density function $P(X)$. In a 2-group or two classes situation, $P(X)$ is either $P_1(X)$, the population density of π_1 or $P_2(X)$, the population density of π_2 . For convenience, it will be assumed that X of the continuous type, the derivations in the discrete case being analogous. In the 2-group situation, $P(X)$ can only be $P_1(X)$ or $P_2(X)$. What is then the probability of achieving a correct classification under the rule $A = (A_1, A_2)$? If the sample point X falls in A_1 , we classify the candidate as belonging to π_1 , and if the true population is also π_1 , then a correct decision is made. In that instance, the corresponding probability is

$$Pr\{1|1, A\} = \int_{A_1} P_1(X) dX \quad (12.2.1)$$

where $dX = dx_1 \wedge dx_2 \wedge \dots \wedge dx_p$, $A = (A_1, A_2)$ denoting one decision rule or one given set of subspaces of the p -space R_p . The probability of misclassification in this case is

$$Pr\{2|1, A\} = \int_{A_2} P_1(X) dX. \quad (12.2.2)$$

Similarly, the probabilities of correctly selecting and misclassifying $P_2(X)$ are respectively given by

$$Pr\{2|2, A\} = \int_{A_2} P_2(X) dX \quad (12.2.3)$$

and

$$Pr\{1|2, A\} = \int_{A_1} P_2(X) dX. \quad (12.2.4)$$

In a Bayesian setting, there is a prior probability q_1 of selecting the population π_1 and q_2 of selecting the population π_2 , with $q_1 + q_2 = 1$. Then, what will be the probability of drawing an observation from π_1 and misclassifying it as belonging to π_2 ? It is $q_1 \times Pr\{2|1, A\} = q_1 \int_{A_2} P_1(X) dX$ and, similarly, the probability of drawing an observation from π_2 and misclassifying it as coming from π_1 is $q_2 \times Pr\{1|2, A\} = q_2 \int_{A_1} P_2(X)$, with the respective costs of misclassifications being $C(2|1) = C(2|1, A)$ and $C(1|2) = C(1|2, A)$. What is

then the expected cost of misclassification? It is the sum of the costs multiplied by the corresponding probabilities. Thus,

$$\text{the expected cost} = q_1 C(2|1)Pr\{2|1, A\} + q_2 C(1|2)Pr\{1|2, A\}. \quad (12.2.5)$$

So, an advantageous criterion to rely on, when setting up A_1 and A_2 would consist in minimizing the expected cost as given in (12.2.5). A rule could be devised for determining A_1 and A_2 accordingly. In this regard, this actually corresponds to Bayes' rule. How can one interpret this expected cost? For example, in the case of admitting students to a particular program of study based on a vector X of test scores, it is the cost of admitting potentially incompetent students or students who would not have successfully completed the program of study and training them, plus the projected cost of losing good students who would have successfully completed the program of study.

If prior probabilities q_1 and q_2 are not involved, then the expected cost of misclassifying an observation from π_1 as coming from π_2 is

$$C(2|1)Pr\{2|1, A\} \equiv E_1(A), \quad (12.2.6)$$

and the expected cost of misclassifying an observation from π_2 as coming from π_1 is

$$C(1|2)Pr\{1|2, A\} \equiv E_2(A). \quad (12.2.7)$$

We would like to have $E_1(A)$ and $E_2(A)$ as small as possible. In this case, a procedure, rule or criterion $A = (A_1, A_2)$ corresponds to determining suitable subspaces A_1 and A_2 in the p -space R_p . If there is another procedure $A^{(j)} = (A_1^{(j)}, A_2^{(j)})$ such that $E_1(A) \leq E_1(A^{(j)})$ and $E_2(A) \leq E_2(A^{(j)})$, then procedure A is said to be as good as $A^{(j)}$, and if at least one of the inequalities above is a strict inequality, that is $<$, then A is preferable to $A^{(j)}$. If procedure A is preferable to all other available procedures $A^{(j)}$, $j = 1, 2, \dots$, A is said to be *admissible*. We are seeking an admissible class $\{A\}$ of procedures.

12.3. Two Populations with Known Distributions

Let π_1 and π_2 be the two populations. Let $P_1(X)$ and $P_2(X)$ be the known p -variate probability/density functions associated with π_1 and π_2 , respectively. That is, $P_1(X)$ and $P_2(X)$ are two p -variate probability/density functions which are fully known in the sense that all their parameters are known in addition to their functional forms. Consider the Bayesian situation where it is assumed that the prior probabilities q_1 and q_2 of selecting π_1 and π_2 , respectively, are known. Suppose that a particular p -vector X is at hand. What

is the probability that this given X is an observation from π_1 ? This probability is $q_1 P_1(X)$ if X is discrete or $q_1 P_1(X)dX$ if X is continuous. What is the probability that the given vector X is an observation vector either from π_1 or from π_2 ? This probability is $q_1 P_1(X) + q_2 P_2(X)$ or $[q_1 P_1(X) + q_2 P_2(X)]dX$. What is then the probability that the vector X at hand is from $P_1(X)$, given that it is an observation vector from π_1 or π_2 ? As this is a conditional statement, it is given by the following in the discrete or continuous case:

$$\frac{q_1 P_1(X)}{q_1 P_1(X) + q_2 P_2(X)} \quad \text{or} \quad \frac{q_1 P_1(X)dX}{[q_1 P_1(X) + q_2 P_2(X)]dX} = \frac{q_1 P_1(X)}{q_1 P_1(X) + q_2 P_2(X)} \quad (12.3.1)$$

where dX , which is the wedge product of differentials and positive in this case, cancels out. If the conditional probability that a given X is an observation from π_1 is larger than or equal to the conditional probability that the given vector X is an observation from π_2 and if we assign X to π_1 , then the chance of misclassification is reduced. Our main objective is to minimize the probability of misclassification and then come up with a decision rule. This statement is equivalent to the following: If

$$\frac{q_1 P_1(X)}{q_1 P_1(X) + q_1 P_2(X)} \geq \frac{q_2 P_2(X)}{q_1 P_1(X) + q_2 P_2(X)} \Rightarrow q_1 P_1(X) \geq q_2 P_2(X) \quad (12.3.2)$$

then we assign X to π_1 , meaning that our subspace A_1 is specified by the following rule:

$$\begin{aligned} A_1 : q_1 P_1(X) \geq q_2 P_2(X) &\Rightarrow \frac{P_1(X)}{P_2(X)} \geq \frac{q_2}{q_1} \\ A_2 : q_1 P_1(X) < q_2 P_2(X) &\Rightarrow \frac{P_1(X)}{P_2(X)} < \frac{q_2}{q_1}. \end{aligned} \quad (12.3.3)$$

Note that if $q_1 P_1(X) = q_2 P_2(X)$, then X can be assigned to either π_1 or π_2 ; however, we have assigned it to π_1 for convenience. Observe that, it is assumed that $q_1 P_1(X) + q_2 P_2(X) \neq 0$, $q_1 > 0$, $q_2 > 0$ and $q_1 + q_2 = 1$ in (12.3.2). The conditional statement made in (12.3.2), which can also be written as

$$\frac{q_i P_i(X)}{q_1 P_1(X) + q_2 P_2(X)} = \frac{\eta_i P_i(X)}{\eta_1 P_1(X) + \eta_2 P_2(X)}, \quad \eta_i > 0, \quad \eta_1 + \eta_2 = \eta > 0, \quad \frac{\eta_i}{\eta} = q_i, \quad i = 1, 2,$$

holds for some weight functions η_i , $i = 1, 2$.

If the observation is from $\pi_1 : P_1(X)$, then the expected cost of misclassification is $q_1 P_1(X)C(2|1) + q_2 P_2(X)C(2|2) = q_1 P_1(X)C(2|1)$ since $C(i|i) = 0$, $i = 1, 2$. Similarly, the expected cost of misclassifying of the observation X from $\pi_2 : P_2(X)$ is $q_2 P_2(X)C(1|2)$. If $P_1(X)$ is our preferred distribution, then we would like the associated expected cost of misclassification to be the lesser one, that is,

$$\begin{aligned} q_1 P_1(X)C(2|1) < q_2 P_2(X)C(1|2) \text{ in } A_2 &\Rightarrow \\ \frac{P_1(X)}{P_2(X)} < \frac{q_2 C(1|2)}{q_1 C(2|1)} \text{ in } A_2 \text{ or} \\ \frac{P_1(X)}{P_2(X)} &\geq \frac{q_2 C(1|2)}{q_1 C(2|1)} \text{ in } A_1, \end{aligned} \quad (12.3.4)$$

which is the same rule as in (12.3.3) where q_1 is replaced by $q_1 C(2|1)$ and q_2 , by $q_2 C(1|2)$.

12.3.1. Best procedure

It can be established that the procedure $A = (A_1, A_2)$ in (12.3.3) is the best one for minimizing the probability of misclassification. To this end, consider any other procedure $A^{(j)} = (A_1^{(j)}, A_2^{(j)})$, $j = 1, 2, \dots$. The probability of misclassification under the procedure $A^{(j)}$ is the following:

$$\begin{aligned} q_1 \int_{A_2^{(j)}} P_1(X) dX + q_2 \int_{A_1^{(j)}} P_2(X) dX \\ = \int_{A_2^{(j)}} [q_1 P_1(X) - q_2 P_2(X)] dX + q_2 \int_{A_1^{(j)} \cup A_2^{(j)}} P_2(X) dX. \end{aligned} \quad (12.3.5)$$

If $A_1^{(j)} \cup A_2^{(j)} = R_p$, then $\int_{A_1^{(j)} \cup A_2^{(j)}} P_2(X) dX = 1$; it is otherwise a given positive constant. However, $q_1 P_1(X) - q_2 P_2(X)$ can be negative, zero or positive, whereas the left-hand side of (12.3.5) is a positive probability. Accordingly, the left-hand side is minimum if

$$q_1 P_1(X) - q_2 P_2(X) < 0 \Rightarrow \frac{P_1(X)}{P_2(X)} < \frac{q_2}{q_1}, \quad (i)$$

which actually is the rejection region A_2 of the procedure $A = (A_1, A_2)$. Hence, the procedure $A = (A_1, A_2)$ minimizes the probabilities of misclassification; in other words, it is the best procedure. If cost functions are also involved, then (i) becomes the following:

$$\frac{P_1(X)}{P_2(X)} < \frac{C(1|2) q_2}{C(2|1) q_1}. \quad (ii)$$

The region where $q_1 P_1(X) - q_2 P_2(X) = 0$ or $q_1 C(2|1)P_1(X) - q_2 C(1|2)P_2(X) = 0$ need not be empty and the probability over this set need not be zero. If

$$Pr \left\{ \frac{P_1(X)}{P_2(X)} = \frac{q_2 C(1|2)}{q_1 C(2|1)} \middle| \pi_i \right\} = 0, \quad i = 1, 2, \quad (12.3.6)$$

it can also be shown that the above Bayes procedure $A = (A_1, A_2)$ is unique. This is stated as a theorem:

Theorem 12.3.1. *Let q_1 be the prior probability of drawing an observation X from the population π_1 with probability/density function $P_1(X)$ and let q_2 be the prior probability of selecting an observation X from the population π_2 with probability/density function $P_2(X)$. Let the cost or loss associated with misclassifying an observation from π_1 as coming from π_2 be $C(2|1)$ and the cost of misclassifying an observation from π_2 as originating from π_1 be $C(1|2)$. Letting*

$$Pr \left\{ \frac{P_1(X)}{P_2(X)} = \frac{C(1|2) q_2}{C(2|1) q_1} \middle| \pi_i \right\} = 0, \quad i = 1, 2,$$

the classification rule given by $A = (A_1, A_2)$ of (12.3.4) is unique and best in the sense that it minimizes the probabilities of misclassification.

Example 12.3.1. Let π_1 and π_2 be two univariate exponential populations whose parameters are θ_1 and θ_2 with $\theta_1 \neq \theta_2$. Let the prior probability of drawing an observation from π_1 be $q_1 = \frac{1}{2}$ and that of selecting an observation from π_2 be $q_2 = \frac{1}{2}$. Let the costs or loss associated with misclassifications be $C(2|1) = C(1|2)$. Compute the regions and probabilities of misclassification if (1): a single observation x is drawn; (2): iid observations x_1, \dots, x_n are drawn.

Solution 12.3.1.(1). In this case, one observation is drawn and the populations are

$$P_i(x) = \frac{1}{\theta_i} e^{-\frac{x}{\theta_i}}, \quad x \geq 0, \quad \theta_i > 0, \quad i = 1, 2.$$

Consider the following inequality on the support of the density:

$$\frac{P_1(x)}{P_2(x)} \geq \frac{C(1|2) q_2}{C(2|1) q_1} = 1,$$

or equivalently,

$$\frac{\theta_2}{\theta_1} e^{-x(\frac{1}{\theta_1} - \frac{1}{\theta_2})} \geq 1 \Rightarrow e^{-x(\frac{1}{\theta_1} - \frac{1}{\theta_2})} \geq \frac{\theta_1}{\theta_2}.$$

On taking logarithms, we have

$$\begin{aligned} -x\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right) \geq \ln \frac{\theta_1}{\theta_2} &\Rightarrow x\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right) \geq \ln \frac{\theta_1}{\theta_2} \\ &\Rightarrow x \geq \frac{\theta_1\theta_2}{\theta_1 - \theta_2} \ln \frac{\theta_1}{\theta_2} \text{ for } \theta_1 > \theta_2. \end{aligned}$$

Letting $\theta_1 > \theta_2$, the steps in the case $\theta_1 < \theta_2$ being parallel, we have

$$x \geq k, \quad k = \frac{\theta_1\theta_2}{\theta_1 - \theta_2} \ln \frac{\theta_1}{\theta_2}.$$

Accordingly,

$$A_1 : x \geq k \text{ and } A_2 : x < k.$$

The probabilities of misclassification are:

$$\begin{aligned} P(2|1) &= \int_{A_2} P_1(x) dx = \int_{x=0}^k \frac{1}{\theta_1} e^{-\frac{x}{\theta_1}} dx = 1 - e^{-\frac{k}{\theta_1}} \\ P(1|2) &= \int_{x=k}^{\infty} \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}} dx = e^{-\frac{k}{\theta_2}}. \end{aligned}$$

Solution 12.3.1.(2). In this case, $X' = (x_1, \dots, x_n)$ and

$$P_i(X) = \prod_{j=1}^n \frac{1}{\theta_i} e^{-\frac{x_j}{\theta_i}} = \frac{1}{\theta_i^n} e^{-\frac{u}{\theta_i}}, \quad i = 1, 2,$$

where $u = \sum_{j=1}^n x_j$ is gamma distributed with the parameters (n, θ_i) , $i = 1, 2$. The density of u is then given by

$$g_i(u) = \frac{1}{\theta_i^n \Gamma(n)} u^{n-1} e^{-\frac{u}{\theta_i}}, \quad i = 1, 2.$$

Proceeding as above, for $\theta_1 > \theta_2$, $A_1 : u \geq k_1$ and $A_2 : u < k_1$, $k_1 = \frac{\theta_1\theta_2}{\theta_1 - \theta_2} \ln\left[\frac{\theta_1}{\theta_2}\right]^n = nk$ where k is as given in Solution 12.3.1(1). Consequently, the probabilities of misclassification are as follows:

$$\begin{aligned} P(2|1) &= \int_{u=0}^{k_1} \frac{u^{n-1}}{\theta_1^n \Gamma(n)} e^{-\frac{u}{\theta_1}} du = \int_0^{\frac{k_1}{\theta_1}} \frac{u^{n-1}}{\Gamma(n)} e^{-u} du \\ P(1|2) &= \int_{k_1}^{\infty} \frac{u^{n-1}}{\theta_2^n \Gamma(n)} e^{-\frac{u}{\theta_2}} du = \int_{\frac{k_1}{\theta_2}}^{\infty} \frac{u^{n-1}}{\Gamma(n)} e^{-u} du \end{aligned}$$

where the integrals can be expressed in terms of incomplete gamma functions or determined by using integration by parts.

Example 12.3.2. Assume that no prior probabilities or costs are involved. Suppose that in a certain clinic, the waiting time before a customer is attended to, depends upon the manager on duty. If manager M_1 is on duty, the expected waiting time is 10 minutes, and if manager M_2 is on duty, the expected waiting time is 5 minutes. Assume that the waiting times are exponentially distributed with expected waiting time equal to θ_i , $i = 1, 2$. On a particular day (1): a customer had to wait 6 minutes before she was attended to, (2): three customers had to wait 6, 6 and 8 minutes, respectively. Who between M_1 and M_2 was likely to be on duty on that day?

Solution 12.3.2.(1). In this case, $\theta_1 = 10$, $\theta_2 = 5$ and the populations are exponential with parameters θ_1 and θ_2 , respectively. Thus, $k = \frac{\theta_1\theta_2}{\theta_1-\theta_2} \ln \frac{\theta_1}{\theta_2} = \frac{(10)(5)}{10-5} \ln \frac{10}{5} = 10 \ln 2$, $\frac{k}{\theta_1} = \frac{10 \ln 2}{10} = \ln 2$, $\frac{k}{\theta_2} = 2 \ln 2 = \ln 4$, $e^{-\frac{k}{\theta_1}} = e^{-\ln 2} = \frac{1}{2} = 0.5$, and $e^{-\frac{k}{\theta_2}} = e^{-\ln 4} = \frac{1}{4} = 0.25$. In (1): the observed value of $x = 6 < 10(\ln 2) = 10(0.69314718056) \approx 6.9315$. Accordingly, we classify x to M_2 , that is, the manager M_2 was likely to be on duty. Thus,

$$\begin{aligned} P(2|2, A) &= \text{The probability of making a correct decision} \\ &= \int_{x < k} P_2(x) dx = \int_0^k P_2(x) dx = \int_0^k \frac{1}{5} e^{-\frac{x}{5}} dx \\ &= 1 - e^{-\ln 4} = 1 - \frac{1}{4} = 0.75; \end{aligned}$$

$$\begin{aligned} P(2|1, A) &= \text{Probability of misclassification or making an incorrect decision} \\ &= \int_0^k P_1(x) dx = \int_0^k \frac{1}{10} e^{-\frac{x}{10}} dx = 1 - e^{-\frac{k}{10}} = 1 - e^{-\ln 2} = \frac{1}{2} = 0.5. \end{aligned}$$

Solution 12.3.2.(2). Here, $u = 6 + 6 + 8 = 20$, $n = 3$ and $k_1 = \frac{\theta_1\theta_2}{\theta_1-\theta_2} n \ln \frac{\theta_1}{\theta_2} = \frac{(10)(5)}{10-5} 3 \ln \frac{10}{5} = 30 \ln 2$. Since $30 \ln 2 \approx 20.795$ and the observed value of u is 20, $u < k_1$, and we assign the sample to π_2 or to $P_2(X)$ or M_2 , with $\frac{k_1}{\theta_2} = \frac{30 \ln 2}{5} = 6 \ln 2$ and $\frac{k_1}{\theta_1} = \frac{30 \ln 2}{10} = 3 \ln 2$. Thus,

$$\begin{aligned} P(2|2, A) &= \text{Probability of making a correct classification decision} \\ &= Pr\{u < k_1 | P_2(X)\} = \int_0^{k_1} \frac{u^{n-1}}{\theta_2^n \Gamma(n)} e^{-\frac{u}{\theta_2}} du \\ &= \int_0^{6 \ln 2} \frac{v^2 e^{-v}}{\Gamma(3)} dv, \quad \text{with } \Gamma(3) = 2! = 2. \end{aligned}$$

Integrating by parts,

$$\int v^2 e^{-v} dv = -[v^2 + 2v + 2]e^{-v}.$$

Then,

$$\begin{aligned} \frac{1}{2} \int_0^{6 \ln 2} v^2 e^{-v} dv &= -\{[2v^2/2 + v + 1]e^{-v}\}_0^{6 \ln 2} = 1 - \frac{1}{64}[(6 \ln 2)^2/2 + (6 \ln 2) + 1] \\ &\approx 1 - \frac{1}{64}[13.797] \approx 0.785, \text{ and} \end{aligned}$$

$P(2|1, A)$ = Probability of misclassification

$$\begin{aligned} &= \int_0^{k_1} P_1(X) dX = \frac{1}{2} \int_0^{3 \ln 2} v^2 e^{-v} dv = -[\frac{1}{2}v^2 + v + 1]e^{-v} \Big|_0^{3 \ln 2} \\ &= 1 - \frac{1}{2^3}[(3 \ln 2)^2/2 + (3 \ln 2) + 1] \approx 0.485. \end{aligned}$$

Example 12.3.3. Let the two populations π_1 and π_2 be univariate normal with mean values μ_1 and μ_2 , respectively, and the same variance σ^2 , that is, $P_1(x) : N_1(\mu_1, \sigma^2)$ and $P_2(x) : N_1(\mu_2, \sigma^2)$. Let the prior probabilities of drawing an observation from these populations be $q_1 = \frac{1}{2}$ and $q_2 = \frac{1}{2}$, respectively, and the costs or loss involved with misclassification be $C(1|2) = C(2|1)$. Determine the regions of misclassification and the corresponding probabilities of misclassification if (1): a single observation x is available; (2): iid observations x_1, \dots, x_n are available, from π_1 or π_2 .

Solution 12.3.3.(1). If one observation is available,

$$P_i(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad -\infty < \mu_i < \infty, \quad \sigma > 0.$$

Consider regions

$$\begin{aligned} A_1 : \frac{P_1(x)}{P_2(x)} \geq \frac{C(1|2)q_2}{C(2|1)q_1} = 1 &\Rightarrow e^{-\frac{1}{2\sigma^2}[(x-\mu_1)^2 - (x-\mu_2)^2]} \geq 1 \\ &\Rightarrow -\left[\frac{1}{2\sigma^2}[(x-\mu_1)^2 - (x-\mu_2)^2]\right] \geq 0. \end{aligned}$$

Now, note that

$$\begin{aligned} -[(x-\mu_1)^2 - (x-\mu_2)^2] &= 2x(\mu_1 - \mu_2) - (\mu_1^2 - \mu_2^2) \geq 0 \Rightarrow \\ x &\geq \frac{1}{2} \frac{(\mu_1^2 - \mu_2^2)}{(\mu_1 - \mu_2)} = \frac{1}{2}(\mu_1 + \mu_2) \text{ for } \mu_1 > \mu_2 \Rightarrow \\ A_1 : x &\geq \frac{1}{2}(\mu_1 + \mu_2) \text{ and } A_2 : x < \frac{1}{2}(\mu_1 + \mu_2). \end{aligned}$$

The probabilities of misclassification are the following for $k = \frac{1}{2}(\mu_1 + \mu_2)$:

$$P(2|1) = \int_{-\infty}^k \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} dx = \Phi\left(\frac{k-\mu_1}{\sigma}\right)$$

$$P(1|2) = \int_k^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}} dx = 1 - \Phi\left(\frac{k-\mu_2}{\sigma}\right)$$

where $\Phi(\cdot)$ is the distribution function of a univariate standard normal density and $k = \frac{1}{2}(\mu_1 + \mu_2)$.

Solution 12.3.3(2). In this case, x_1, \dots, x_n are iid and $X' = (x_1, \dots, x_n)$. The multivariate densities are

$$P_i(X) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu_i)^2} = \frac{e^{-\frac{1}{2\sigma^2} (\sum_{j=1}^n (x_j - \bar{x})^2 + n(\bar{x} - \mu_i)^2)}}{\sigma^n (\sqrt{2\pi})^n}, \quad i = 1, 2,$$

where $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$. Hence for $\mu_1 > \mu_2$,

$$A_1 : \frac{P_1(X)}{P_2(X)} \geq 1 \Rightarrow e^{-\frac{n}{2\sigma^2} [(\bar{x} - \mu_1)^2 - (\bar{x} - \mu_2)^2]} \geq 1.$$

Taking logarithms and simplifying, we have

$$-\frac{n}{2\sigma^2} [(\bar{x} - \mu_1)^2 - (\bar{x} - \mu_2)^2] \geq 0 \Rightarrow$$

$$\bar{x} \geq \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{1}{2}(\mu_1 + \mu_2) \text{ for } \mu_1 > \mu_2$$

where

$$\bar{x} \sim N_1\left(\mu_i, \frac{\sigma^2}{n}\right), \quad i = 1, 2.$$

Therefore the probabilities of misclassification are the following:

$$P(2|1) = \int_{-\infty}^k \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2} (\bar{x} - \mu_1)^2} d\bar{x} = \Phi\left(\frac{\sqrt{n}(\mu_2 - \mu_1)}{2\sigma}\right)$$

$$P(1|2) = \int_k^{\infty} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2} (\bar{x} - \mu_2)^2} d\bar{x} = 1 - \Phi\left(\frac{\sqrt{n}(\mu_1 - \mu_2)}{2\sigma}\right)$$

where $k = \frac{1}{2}(\mu_1 + \mu_2)$ and $\Phi(\cdot)$ is the distribution function of a univariate standard normal random variable.

Example 12.3.4. Assume that no prior probabilities or costs are involved. A tuber crop called tapioca is planted by farmers. While farmer F_1 applies a standard fertilizer to the soil to enhance the growth of the tapioca plants, farmer F_2 does not apply any fertilizer and let the plants grow naturally. At harvest time, a tapioca plant is pulled up with all its tubers attached to the bottom of the stem. The upper part of the stem is cut off and the lower part with its tubers is put out for sale. Tuber yield per plant, x , is measured by weighing the lower part of the stem with the tubers attached. It is known from past experience that x is normally distributed with mean value $\mu_1 = 5$ and variance $\sigma^2 = 1$ for F_1 type farms, that is, $x \sim N_1(\mu_1 = 5, \sigma^2 = 1)|F_1$ and that for F_2 type farms, $x \sim N_1(\mu_2 = 3, \sigma^2 = 1)|F_2$, the weights being measured in kilograms. A road-side vendor is selling tapioca and his collection is either from F_1 type farms or F_2 type farms, but not both. A customer picked (1): one stem with its tubers attached weighing 4.2 kg (2) a random sample of four stems respectively weighing 6, 4, 3 and 5 kg. To which type of farms will you classify the observations in (1) and (2)?

Solution 12.3.4. (1). The decision is based on $k = \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}(5 + 3) = 4$. In this case, the decision rule $A = (A_1, A_2)$ is such that $A_1 : x \geq k$ and $A_2 : x < k$ for $\mu_1 > \mu_2$. Note that $\frac{k-\mu_1}{\sigma} = k - \mu_1 = 4 - 5 = -1$ and $\frac{k-\mu_2}{\sigma} = (4 - 3) = 1$. As the observed x is $4.2 > 4 = k$, we classify x into $P_1(X) : N_1(\mu_1, 1)$. Moreover,

$P(1|1, A) =$ Probability of making a correct classification decision

$$\begin{aligned} &= Pr\{x \geq k | P_1(x)\} = \int_k^{\infty} \frac{e^{-\frac{1}{2}(x-\mu_1)^2}}{\sqrt{(2\pi)}} dx \\ &= \int_{-1}^{\infty} \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du = 0.5 + \int_0^1 \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du \approx 0.84, \end{aligned}$$

and

$P(1|2, A) =$ Probability of misclassification

$$= Pr\{x \geq k | P_2(x)\} = \int_k^{\infty} \frac{e^{-\frac{1}{2}(x-\mu_2)^2}}{\sqrt{(2\pi)}} dx = \int_1^{\infty} \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du \approx 0.16.$$

Solution 12.3.4. (2). In this case, $\bar{x} = \frac{1}{4}(6+4+3+5) = 4.5$, $n = 4$, $\bar{x} \sim N(\mu_i, \frac{1}{n})$, $i = 1, 2$, $\frac{(k-\mu_1)}{\sigma/\sqrt{n}} = 2(4 - 5) = -2$ and $\frac{(k-\mu_2)}{\sigma/\sqrt{n}} = 2(4 - 3) = 2$. Since the observed \bar{x} is $4.5 > 4 = k$, we assign the sample to $P_1(X) : N(\mu_1, 1)$, the criterion being $A_1 : \bar{x} \geq k$ and $A_2 : \bar{x} < k$. Additionally,

$P(1|1, A)$ = Probability of a correct classification

$$\begin{aligned} &= Pr\{\bar{x} \geq k | P_1(X)\} = \int_k^\infty \frac{e^{-\frac{n}{2}(\bar{x}-\mu_1)^2}}{\sqrt{(2\pi)}} d\bar{x} = \int_{-2}^\infty \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du \\ &= 0.5 + \int_0^2 \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du \approx 0.98, \end{aligned}$$

and

$P(1|2, A)$ = Probability of misclassification

$$= Pr\{\bar{x} \geq k | P_2(X)\} = \int_k^\infty \frac{e^{-\frac{n}{2}(\bar{x}-\mu_2)^2}}{\sqrt{(2\pi)}} d\bar{x} = \int_2^\infty \frac{e^{-\frac{1}{2}u^2}}{\sqrt{(2\pi)}} du \approx 0.023.$$

Example 12.3.5. Let π_1 and π_2 be two p -variate real nonsingular normal populations sharing the same covariance matrix, $\pi_1 : N_p(\mu^{(1)}, \Sigma)$, $\Sigma > O$, and $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$, whose mean values are such that $\mu^{(1)} \neq \mu^{(2)}$. Let the prior probabilities be $q_1 = q_2$ and the cost functions be $C(1|2) = C(2|1)$. Consider a single p -vector X to be classified into π_1 or π_2 . Determine the regions of misclassification and the corresponding probabilities.

Solution 12.3.5. The p -variate real normal densities are the following:

$$P_i(X) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu^{(i)})' \Sigma^{-1}(X-\mu^{(i)})} \quad (i)$$

for $i = 1, 2$, $\Sigma > O$, $\mu^{(1)} \neq \mu^{(2)}$. Consider the inequality

$$\begin{aligned} \frac{P_1(X)}{P_2(X)} &\geq \frac{C(1|2) q_2}{C(2|1) q_1} = 1 \Rightarrow \\ e^{-\frac{1}{2}[(X-\mu^{(1)})' \Sigma^{-1}(X-\mu^{(1)}) - (X-\mu^{(2)})' \Sigma^{-1}(X-\mu^{(2)})]} &\geq 1. \end{aligned}$$

Taking logarithms, we have

$$\begin{aligned} -\frac{1}{2}[(X-\mu^{(1)})' \Sigma^{-1}(X-\mu^{(1)}) - (X-\mu^{(2)})' \Sigma^{-1}(X-\mu^{(2)})] &\geq 0 \Rightarrow \\ (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1}(\mu^{(1)} + \mu^{(2)}) &\geq 0. \end{aligned}$$

Let

$$u = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1}(\mu^{(1)} + \mu^{(2)}). \quad (12.3.7)$$

Then, u has a univariate normal distribution since it is a linear function of the components of X , which is a p -variate normal. Thus,

$$\begin{aligned}\text{Var}(u) &= \text{Var}[(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X] \\ &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} \text{Cov}(X) \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \\ &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) = \Delta^2\end{aligned}\quad (12.3.8)$$

where Δ^2 is Mahalanobis' distance. The mean values of u under π_1 and π_2 are respectively,

$$\begin{aligned}E(u)|\pi_1 &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} E(X)|\pi_1 - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) \\ &= \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) = \frac{1}{2}\Delta^2,\end{aligned}\quad (12.3.9)$$

$$\begin{aligned}E(u)|\pi_2 &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} \mu^{(2)} - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) \\ &= \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(2)} - \mu^{(1)}) = -\frac{1}{2}\Delta^2,\end{aligned}\quad (12.3.10)$$

so that

$$\begin{aligned}u &\sim N_1(\frac{1}{2}\Delta^2, \Delta^2) \text{ under } \pi_1, \\ u &\sim N_1(-\frac{1}{2}\Delta^2, \Delta^2) \text{ under } \pi_2.\end{aligned}\quad (12.3.11)$$

Accordingly, the regions of misclassification are

$$A_2 : u < 0 | \pi_1 : u \sim N_1(\frac{1}{2}\Delta^2, \Delta^2) \text{ and } A_1 : u \geq 0 | \pi_2 : u \sim N_1(-\frac{1}{2}\Delta^2, \Delta^2), \quad (12.3.12)$$

and the probabilities of misclassification are as follows:

$$\begin{aligned}P(2|1) &= \int_{-\infty}^0 \frac{1}{\Delta\sqrt{2\pi}} e^{-\frac{1}{2\Delta^2}(u-\frac{1}{2}\Delta^2)^2} du \\ &= \int_{-\infty}^{\frac{0-\frac{1}{2}\Delta^2}{\Delta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(-\frac{1}{2}\Delta) \quad (ii) \\ P(1|2) &= \int_0^{\infty} \frac{1}{\Delta\sqrt{2\pi}} e^{-\frac{1}{2\Delta^2}(u+\frac{1}{2}\Delta^2)^2} du \\ &= \int_{\frac{0+\frac{1}{2}\Delta^2}{\Delta}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1 - \Phi(\frac{1}{2}\Delta) \quad (iii)\end{aligned}$$

where $\Phi(\cdot)$ denotes the distribution function of a univariate standard normal variable.

Note 12.3.1. If no conditions are imposed on the prior probabilities, q_1 and q_2 , or on the costs of misclassification, $C(2|1)$ and $C(1|2)$, then the regions are determined as $A_1 : u \geq k$, $k = \ln \frac{C(1|2)q_2}{C(2|1)q_1}$, and $A_2 : u < k$. In this case, the probabilities of misclassification will be $\Phi\left(\frac{k-\frac{1}{2}\Delta^2}{\Delta}\right)$ and $1 - \Phi\left(\frac{k+\frac{1}{2}\Delta^2}{\Delta}\right)$, respectively.

Note 12.3.2. If the prior probabilities q_1 and q_2 are not known, we may assume that the two populations π_1 and π_2 are equally likely to be chosen or equivalently that $q_1 = q_2 = \frac{1}{2}$, in which instance $k = \ln \frac{C(1|2)}{C(2|1)}$. Then, the correct decisions are to assign the vector X at hand to π_1 in the region A_1 and to π_2 in the region A_2 , where $A_1 : u \geq k$ and $A_2 : u < k$, $k = \ln \frac{q_2 C(1|2)}{q_1 C(2|1)}$ with $q_1, q_2, C(2|1)$ and $C(1|2)$ assumed to be known and

$$u = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1}(\mu^{(1)} + \mu^{(2)})$$

whose first term, namely $(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X$, is known as the linear discriminant function, which is utilized to discriminate or to separate two p -variate populations, not necessarily normally distributed, having mean value vectors $\mu^{(1)}$ and $\mu^{(2)}$ and sharing the same covariance matrix $\Sigma > O$.

Example 12.3.6. Assume that no prior probabilities or costs are involved. Applicants to a certain training program are given tests to evaluate their aptitude for languages and aptitude for science. Let the test scores be denoted by x_1 and x_2 , respectively. Let X be the bivariate vector $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. After completing the training program, their aptitudes are tested again. Let $X^{(1)'} = [x_1^{(1)}, x_2^{(1)}]$ be the score vector in the group of successful trainees and let $X^{(2)'} = [x_1^{(2)}, x_2^{(2)}]$ be the score vector in the group of unsuccessful trainees. From previous experience of conducting such tests over the years, it is known that $X^{(1)} \sim N_2(\mu^{(1)}, \Sigma)$, $\Sigma > O$, and $X^{(2)} \sim N_2(\mu^{(2)}, \Sigma)$, $\Sigma > O$, where

$$\mu^{(1)} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \mu^{(2)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

Then (1): one applicant taken at random before the training program started obtained the test scores $X_0 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$; (2): three applicants chosen at random before the training program started had the following scores:

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

In (1), classify X_0 to π_1 or π_2 and in (2), classify the entire sample of three vectors into π_1 or π_2 .

Solution 12.3.6. Let us compute certain quantities which are needed to answer the questions:

$$\begin{aligned}\frac{1}{2}(\mu^{(1)} + \mu^{(2)}) &= \frac{1}{2} \left(\begin{bmatrix} 4 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}; \\ \mu^{(1)} - \mu^{(2)} &= \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X = [2, -2] X = 2x_1 - 2x_2; \\ \Delta^2 &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) = [2, 0] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 4; \\ (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) &= [2, 0] \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = 8.\end{aligned}$$

Hence,

$$\begin{aligned}u &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2} (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) \\ &= 2x_1 - 2x_2 - 4;\end{aligned}$$

$$u|\pi_1 \sim N_1\left(\frac{1}{2}\Delta^2, \Delta^2\right), \quad u|\pi_2 \sim N_1\left(-\frac{1}{2}\Delta^2, \Delta^2\right);$$

$$A_1 : u \geq 0, \quad A_2 : u < 0.$$

Since, in (1), the observed $X_0 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, the observed u is $u = 2x_1 - 2x_2 - 4 = 8 - 2 - 4 = 2 > 0$ and we classify the observed X_0 into $\pi_1 : N_1\left(\frac{1}{2}\Delta^2, \Delta^2\right)$, the criterion being $A_1 : u \geq 0$ and $A_2 : u < 0$. Thus,

$P(1|1, A)$ = Probability of making a correct classification decision

$$\begin{aligned}&= Pr\{u \geq 0|\pi_1\} = \int_0^\infty \frac{e^{-\frac{1}{2\Delta^2}(u-\frac{1}{2}\Delta^2)^2}}{\Delta\sqrt{(2\pi)}} du = \int_{-\frac{\Delta}{2}}^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv \\ &= \int_{-1}^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv = 0.5 + \int_0^1 \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv \approx 0.841;\end{aligned}$$

$P(1|2, A)$ = Probability of misclassification

$$= \int_0^\infty \frac{e^{-\frac{1}{2\Delta^2}(u+\frac{1}{2}\Delta^2)^2}}{\Delta\sqrt{(2\pi)}} du = \int_1^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv \approx 0.159.$$

When solving (2), the entire sample is to be classified. Proceeding as in the derivation of the criterion u in case (1), it is seen that for the problem at hand, X_0 will be replaced by \bar{X} ,

the average of the sample vectors or the sample mean value vector, and then u will become $u_1 = 2\bar{x}_1 - 2\bar{x}_2 - 4$ where $\bar{X}' = [\bar{x}_1, \bar{x}_2]$. Thus, we require the sample average:

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 12 \\ 4 \end{bmatrix} \Rightarrow \text{observed sample mean} = \frac{1}{3} \begin{bmatrix} 12 \\ 4 \end{bmatrix}.$$

This means that $\bar{x}_1 = \frac{12}{3} = 4$, $\bar{x}_2 = \frac{4}{3}$, and the observed $u_1 = 2\bar{x}_1 - 2\bar{x}_2 - 4 = 8 - \frac{8}{3} - 4 > 0$. Hence, we classify the whole sample to π_1 as the criterion is $A_1 : u_1 \geq 0$ and $A_2 : u_1 < 0$. Since \bar{X} is normally distributed with $E[\bar{X}] = \mu^{(i)}$ and $\text{Cov}(\bar{X}) = \frac{1}{n}\Sigma$, $i = 1, 2$, where n is the sample size, the densities of u_1 under π_1 and π_2 are the following:

$$\begin{aligned} u_1|\pi_1 &\sim N_1\left(\frac{1}{2}\Delta^2, \frac{1}{3}\Delta^2\right), \quad n = 3, \\ u_1|\pi_2 &\sim N_1\left(-\frac{1}{2}\Delta^2, \frac{1}{3}\Delta^2\right). \end{aligned}$$

Moreover,

$P(1|1, A)$ = Probability of making a correct classification decision

$$\begin{aligned} &= Pr\{u_1 \geq 0|\pi_1\} = \int_0^\infty \frac{\sqrt{3}}{\Delta\sqrt{(2\pi)}} e^{-\frac{3}{2\Delta^2}(u_1 - \frac{1}{2}\Delta^2)^2} du_1 \\ &= \int_{-\frac{\sqrt{3}\Delta}{2}}^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv = \int_{-\sqrt{3}}^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv = 0.5 + \int_0^{\sqrt{3}} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv \approx 0.958 \end{aligned}$$

and

$P(1|2, A)$ = Probability of misclassification

$$\begin{aligned} &= Pr\{u_1 \geq 0|\pi_2\} = \int_0^\infty \frac{\sqrt{3}}{\Delta\sqrt{(2\pi)}} e^{-\frac{3}{2\Delta^2}(u_1 + \frac{1}{2}\Delta^2)^2} du_1 \\ &= \int_{\sqrt{3}}^\infty \frac{e^{-\frac{1}{2}v^2}}{\sqrt{(2\pi)}} dv \approx 0.042. \end{aligned}$$

12.4. Linear Discriminant Function

Let X be a $p \times 1$ vector and B a $p \times 1$ arbitrary constant vector, $B' = (b_1, \dots, b_p)$. Consider the arbitrary linear function $w = B'X$. Then, the mean value and variance of w are the following: $E(w) = B'E(X)$ and $\text{Var}(w) = \text{Var}(B'X) = B'\text{Cov}(X)B = B'\Sigma B$ where $\Sigma > O$ is the covariance matrix of X . Suppose that the X could be from a p -variate real population π_1 with mean value vector $\mu^{(1)}$ or from the p -variate real population π_2 with mean value vector $\mu^{(2)}$. Suppose that both the populations π_1 and π_2 have the same

covariance matrix $\Sigma > O$. Then, a measure of discrimination or separation between π_1 and π_2 is $|B'\mu^{(1)} - B'\mu^{(2)}|$ as measured in terms of the standard deviation $\sqrt{\text{Var}(w)}$ for determining the best choice of B . Taking the squared distance, let

$$\delta = \frac{[B'\mu^{(1)} - B'\mu^{(2)}]^2}{B'\Sigma B} = \frac{[B'(\mu^{(1)} - \mu^{(2)})]^2}{B'\Sigma B} = \frac{B'(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'B}{B'\Sigma B} \quad (12.4.1)$$

since the square of a scalar quantity is the scalar quantity times its transpose, $B'(\mu^{(1)} - \mu^{(2)})$ being a scalar quantity. Accordingly, we will maximize δ as specified in (12.4.1). This will be achieved by selecting a particular B in such a way that δ attains a maximum which corresponds to the maximum distance between π_1 and π_2 . Without any loss of generality, we may assume that $B'\Sigma B = 1$, so that only the numerator in (12.4.1) need be maximized, subject to the condition $B'\Sigma B = 1$. Let λ denote a Lagrangian multiplier and

$$\eta = B'(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'B - \lambda(B'\Sigma B - 1).$$

Let us take the partial derivative of η with respect to the vector B and equate the result to a null vector (the reader may refer to Chap. 1 for the derivative of a scalar variable with respect to a vector variable):

$$\begin{aligned} \frac{\partial \eta}{\partial B} = O &\Rightarrow 2(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'B - 2\lambda\Sigma B = O \\ &\Rightarrow \Sigma^{-1}(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'B = \lambda B. \end{aligned} \quad (i)$$

Note that $(\mu^{(1)} - \mu^{(2)})'B \equiv \alpha$ is a scalar quantity and B is a specific vector coming from (i) and hence we may write (i) as

$$B = \frac{\alpha}{\lambda} \Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) \equiv c \Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) \quad (ii)$$

where c is a real scalar quantity. Observe that δ as given in (12.4.1) will remain the same if B is multiplied by any scalar quantity. Thus, we may take $c = 1$ in (ii) without any loss of generality. The linear discriminant function then becomes

$$B'X = (\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}X, \quad (12.4.2)$$

and when $B'X$ is as given in (12.4.2), δ as defined in (12.4.1), can be expressed as follows:

$$\begin{aligned} \delta &= \frac{(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)})(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)})}{(\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)})} \\ &= (\mu^{(1)} - \mu^{(2)})'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) = \Delta^2 \equiv \text{Mahalanobis' distance} \\ &= \text{Var}(w) = \text{Variance of the discriminant function.} \end{aligned} \quad (12.4.3)$$

This δ is also the generalized squared distance between the vectors $\mu^{(1)}$ and $\mu^{(2)}$ or the squared distance between the vectors $\Sigma^{-\frac{1}{2}}\mu^{(1)}$ and $\Sigma^{-\frac{1}{2}}\mu^{(2)}$ in the mathematical sense (Euclidean distance). Hence Mahalanobis' distance between two p -variate populations with different mean value vectors and the same covariance matrix is a measure of discrimination or separation between the populations, and the linear discriminant function is given in (12.4.2). Hence for an observed value X , if $u = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X > 0$ when $\mu^{(1)}$, $\mu^{(2)}$ and Σ are known, then we choose population π_1 with mean value $\mu^{(1)}$, and if $u < 0$, then we select population π_2 with mean value $\mu^{(2)}$. When $u = 0$, both π_1 and π_2 are equally favored.

Example 12.4.1. In a small township, there is only one grocery store. The town is laid out on the East and West sides of the sole main road. We will refer to the villagers as East-enders and West-enders. These townspeople shop only once a week for groceries. The grocery store owner found that the East-enders and West-enders have somewhat different buying habits. Consider the following items: x_1 = grain items in kilograms, x_2 = vegetable items in kilograms, x_3 = dairy products in kilograms, and let $[x_1, x_2, x_3] = X'$ where X is the vector of weekly purchases. Then, the expected quantities bought by the East-enders and West-enders are $E(X) = \mu^{(1)}$ and $E(X) = \mu^{(2)}$, respectively, with the common covariance matrix $\Sigma > O$. From past history, the grocery store owner determined that

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \mu^{(1)} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \mu^{(2)} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Consider the following situations: (1) A customer walked in and bought $x_1 = 1$ kg of grain items, $x_2 = 2$ kg of vegetable items, and $x_3 = 1$ kg of dairy products. Is she likely to be an East-enders or West-enders? (2): Another customer bought the three types of items in the quantities (10, 1, 1), respectively. Is she more likely to be an East-enders than a West-enders?

Solution 12.4.1. The inverse of the covariance matrix, $\mu^{(1)} - \mu^{(2)}$, as well as other relevant quantities are the following:

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \mu^{(1)} - \mu^{(2)} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

$$(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} = [1, 0, -1] \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} = [\frac{1}{3}, -1, -2].$$

In (1), $X' = (1, 2, 1)$ and since

$$(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X = [\frac{1}{3}, -1, -2] \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} < 0,$$

we classify this customer as a West-ender from her buying pattern. In (2),

$$(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X = [\frac{1}{3}, -1, -2] \begin{bmatrix} 10 \\ 1 \\ 1 \end{bmatrix} > 0,$$

so that, given her purchases, this customer is classified as an East-ender.

12.5. Classification When the Population Parameters are Unknown

We now consider the classification problem involving two populations π_1 and π_2 for which the parameters of the corresponding densities are unknown. Since the structure of the parameters in these general densities $P_1(X)$ and $P_2(X)$ is not known, we will present a specific example: Consider the two p -variate normal populations of Example 12.3.3. Let $\pi_1 : N_p(\mu^{(1)}, \Sigma)$ and $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, which share the same positive definite covariance matrix Σ . Suppose that we have a single observation vector X to be classified into π_1 or π_2 . When the parameters $\mu^{(1)}$, $\mu^{(2)}$ and Σ are unknown, we will have to estimate them from some training samples. But, for a problem such as classifying skeletal remains, one does not have samples from the respective ancestral groups. Nevertheless, one can obtain training samples from living racial groups, and so, secure estimates of the parameters involved. Assume that we have simple random samples of sizes n_1 and n_2 from $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$, respectively. Denote the sample values by $X_1^{(1)}, \dots, X_{n_1}^{(1)}$, and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$, and let $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ be the sample averages. That is,

$$X_j^{(1)} = \begin{bmatrix} x_{1j}^{(1)} \\ \vdots \\ x_{pj}^{(1)} \end{bmatrix}, \quad j = 1, \dots, n_1; \quad X_j^{(2)} = \begin{bmatrix} x_{1j}^{(2)} \\ \vdots \\ x_{pj}^{(2)} \end{bmatrix}, \quad j = 1, \dots, n_2;$$

$$\bar{X}^{(1)} = \begin{bmatrix} \bar{x}_1^{(1)} \\ \vdots \\ \bar{x}_p^{(1)} \end{bmatrix}, \quad \bar{x}_k^{(1)} = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{kj}^{(1)}; \quad \bar{X}^{(2)} = \begin{bmatrix} \bar{x}_1^{(2)} \\ \vdots \\ \bar{x}_p^{(2)} \end{bmatrix}, \quad \bar{x}_k^{(2)} = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{kj}^{(2)}. \quad (12.5.1)$$

Let the sample matrices be denoted by bold-faced letters where the $p \times n_1$ matrix $\mathbf{X}^{(1)}$ and the $p \times n_2$ matrix $\mathbf{X}^{(2)}$ are the sample matrices and let $\bar{\mathbf{X}}^{(1)}$ and $\bar{\mathbf{X}}^{(2)}$ be the matrices of sample means. Thus, we have

$$\begin{aligned}
 \mathbf{X}^{(1)} &= [X_1^{(1)}, \dots, X_{n_1}^{(1)}] = \begin{bmatrix} x_{11}^{(1)} & \dots & x_{1n_1}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{p1}^{(1)} & \dots & x_{pn_1}^{(1)} \end{bmatrix}, \\
 \bar{\mathbf{X}}^{(1)} &= [\bar{X}_1^{(1)}, \dots, \bar{X}^{(1)}] = \begin{bmatrix} \bar{x}_1^{(1)} & \dots & \bar{x}_1^{(1)} \\ \vdots & \ddots & \vdots \\ \bar{x}_p^{(1)} & \dots & \bar{x}_p^{(1)} \end{bmatrix}, \\
 \mathbf{X}^{(2)} &= [X_1^{(2)}, \dots, X_{n_2}^{(2)}] = \begin{bmatrix} x_{11}^{(2)} & \dots & x_{1n_2}^{(2)} \\ \vdots & \ddots & \vdots \\ x_{p1}^{(2)} & \dots & x_{pn_2}^{(2)} \end{bmatrix}, \\
 \bar{\mathbf{X}}^{(2)} &= [\bar{X}_1^{(2)}, \dots, \bar{X}^{(2)}] = \begin{bmatrix} \bar{x}_1^{(2)} & \dots & \bar{x}_1^{(2)} \\ \vdots & \ddots & \vdots \\ \bar{x}_p^{(2)} & \dots & \bar{x}_p^{(2)} \end{bmatrix}. \tag{12.5.2}
 \end{aligned}$$

Then, the sample sum of products matrices are

$$\begin{aligned}
 S_i &= (\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)})(\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)})', \quad i = 1, 2; \\
 S_m &= (s_{ij}^{(m)}), \quad s_{ij}^{(m)} = \sum_{k=1}^{n_m} (x_{ik}^{(m)} - \bar{x}_i^{(m)})(x_{jk}^{(m)} - \bar{x}_j^{(m)}), \quad m = 1, 2, \quad S = S_1 + S_2. \tag{12.5.3}
 \end{aligned}$$

The unbiased estimators of $\mu^{(1)}, \mu^{(2)}$ and Σ are respectively $\bar{X}^{(1)}, \bar{X}^{(2)}$ and $\frac{S}{n_{(2)}} = \frac{S_1+S_2}{n_{(2)}}$, $n_{(2)} = n_1 + n_2 - 2$. The criteria for classification, the regions, the statistic, and so on, are available from Example 12.3.3. That is,

$$A_1 : u \geq k, \quad A_2 : u < k, \quad k = \ln \frac{C(1|2)q_2}{C(2|1)q_1},$$

where

$$u = X' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) - \frac{1}{2} (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}).$$

Note that q_1 and q_2 are the prior probabilities of selecting the populations π_1 and π_2 and $C(1|2)$ and $C(2|1)$ are the costs or loss associated with misclassification. We will assume that $q_1, q_2, C(1|2)$ and $C(2|1)$ are all known but the parameters $\mu^{(1)}, \mu^{(2)}$ and Σ are

estimated by their unbiased estimators. Denoting the estimator of u as v , we obtain the following criterion, assuming that we have one p -vector X to be classified into π_1 or π_2 :

$$\begin{aligned} A_1 : v \geq k, \quad A_2 : v < k, \quad k &= \ln \frac{q_2 C(1|2)}{q_1 C(2|1)}, \\ v &= n_{(2)} X' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)}) - n_{(2)} \frac{1}{2} (\bar{X}^{(1)} - \bar{X}^{(2)})' S^{-1} (\bar{X}^{(1)} + \bar{X}^{(2)}) \\ &= n_{(2)} [X - \frac{1}{2} (\bar{X}^{(1)} + \bar{X}^{(2)})]' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)}). \end{aligned} \quad (12.5.4)$$

As it turns out, it already proves quite challenging to obtain the exact distribution of v as given in (12.5.4) where X is a single p -vector either from π_1 or from π_2 .

12.5.1. Some asymptotic results

Before considering asymptotic properties of u and v as defined in Sect. 12.4, let us recall certain results obtained in earlier chapters. Let the $p \times 1$ vectors Y_j , $j = 1, \dots, n$, be iid vectors from some population for which $E[Y_j] = \mu$ and $\text{Cov}(Y_j) = \Sigma > O$, $j = 1, \dots, n$. Let the sample matrix, the matrix of sample means wherein the sample mean $\bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$ and the sample sum of products matrix S be the as follows:

$$\begin{aligned} \mathbf{Y} &= [Y_1, \dots, Y_n], \quad \bar{\mathbf{Y}} = [\bar{Y}, \dots, \bar{Y}], \quad S = (s_{ij}), \quad s_{ij} = \sum_{k=1}^n (y_{ik} - \bar{y}_i)(y_{jk} - \bar{y}_j), \\ S &= [\mathbf{Y} - \bar{\mathbf{Y}}][\mathbf{Y} - \bar{\mathbf{Y}}]' = \mathbf{Y} [I_n - J J' / n] \mathbf{Y}, \quad Y'_j = [y_{1j}, y_{2j}, \dots, y_{pj}], \end{aligned} \quad (i)$$

where J is a $n \times 1$ vector of unities. Since a matrix of the form $\mathbf{Y} - \bar{\mathbf{Y}}$ is present, we may let $\mu = O$ without any loss of generality in the following computations since $Y_j - \bar{Y} = (Y_j - \mu) - (\bar{Y} - \mu)$. Note that $B = B' = I_n - \frac{1}{n} J J' = B^2$ and hence, B is idempotent and of rank $n - 1$. Since $B = B'$, there exists an orthonormal matrix Q such that $Q' B Q = \text{diag}(1, \dots, 1, 0) = D$, $Q Q' = I$, $Q' Q = I$, the diagonal elements being 1's and 0 since $B = B^2$ and of rank $n - 1$. Then,

$$\begin{aligned} S &= \mathbf{Y} Q \text{diag}(1, \dots, 1, 0) Q' \mathbf{Y}' = \mathbf{Y} Q D D' Q' \mathbf{Y}', \\ D &= \text{diag}(1, \dots, 1, 0). \end{aligned} \quad (ii)$$

Consider $\Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}}$. Let $U_j = \Sigma^{-\frac{1}{2}} Y_j$, $j = 1, \dots, n$, where Y_j is the j -th column of \mathbf{Y} and it is assumed that $\mu = O$. Observe that $E[U_j] = O$, $\text{Cov}(U_j) = I_p$, $j = 1, \dots, n$, and the U_j 's are uncorrelated. Letting $\mathbf{U} = [U_1, \dots, U_n]$, (ii) implies that

$$\Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}} = \mathbf{U} Q D D' Q' \mathbf{U}'. \quad (iii)$$

Denoting by $U_{(j)}$ the j -th row of \mathbf{U} , it follows that the elements of $U_{(j)}$ are iid uncorrelated real scalar variables with mean value zero and variance 1. Consider the transformation $V_{(j)} = U_{(j)}\mathbf{Q}$; then $E[V_{(j)}] = \mathbf{O}$ and $\text{Cov}[V_{(j)}] = I_n$, $j = 1, \dots, p$, the $V_{(j)}$'s being the uncorrelated. Let \mathbf{V} be the $p \times n$ matrix whose rows are $V_{(j)}$, $j = 1, \dots, p$. Let the columns of \mathbf{V} be V_j , $j = 1, \dots, n$, that is, $\mathbf{V} = [V_1, \dots, V_n]$. Then, (iii) implies the following:

$$\begin{aligned} \Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}} &= \mathbf{V}\mathbf{D}\mathbf{D}'\mathbf{V}' = \{[V_1, \dots, V_n]\mathbf{D}\}\{[V_1, \dots, V_n]\mathbf{D}\}' \\ &= [V_1, \dots, V_{n-1}, \mathbf{O}][V_1, \dots, V_{n-1}, \mathbf{O}]' = V_1V_1' + \dots + V_{n-1}V_{n-1}' \Rightarrow \\ E[\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}] &= E[V_1V_1'] + \dots + E[V_{n-1}V_{n-1}'] = I_p + \dots + I_p = (n-1)I_p \Rightarrow \\ E[S] &= (n-1)\Sigma \text{ or } E\left[\frac{S}{n-1}\right] = \Sigma. \end{aligned} \tag{iv}$$

Additionally,

$$\begin{aligned} \text{Cov}(\bar{Y}) &= \frac{1}{n^2}\text{Cov}[Y_1 + \dots + Y_n] = \frac{1}{n^2}[\text{Cov}(Y_1) + \dots + \text{Cov}(Y_n)] \\ &= \frac{1}{n^2}[\Sigma + \dots + \Sigma] = \frac{n}{n^2}\Sigma = \frac{\Sigma}{n} \rightarrow \mathbf{O} \text{ as } n \rightarrow \infty, \end{aligned} \tag{v}$$

when Σ is finite with respect to any norm of Σ , namely $\|\Sigma\| < \infty$. Appealing to the extended Chebyshev inequality, this shows that the unbiased estimator of μ , namely \bar{Y} , converges to μ in probability, that is,

$$Pr(\bar{Y} \rightarrow \mu) \rightarrow 1 \text{ when } n \rightarrow \infty \text{ or } \lim_{n \rightarrow \infty} Pr(\bar{Y} \rightarrow \mu) = 1. \tag{vi}$$

An unbiased estimator of Σ is $\hat{\Sigma} = \frac{S}{n-1}$ with $E[\hat{\Sigma}] = \Sigma$. Will $\hat{\Sigma}$ also converge to Σ in probability when $n \rightarrow \infty$? In order to establish this, we require the covariance structure of the elements in S . For arbitrary populations, it is somewhat difficult to verify this result; however, it is rather straightforward for normal populations. We will examine this aspect next.

12.5.2. Another method

Let the $p \times 1$ vectors X_j , $j = 1, \dots, n$, be a simple random sample of size n from a population having a real $N_p(\mu, \Sigma)$, $\Sigma > \mathbf{O}$, distribution. Letting S denote the sample sum of products matrix, S will be distributed as a Wishart matrix with $m = n - 1$ degrees of freedom and $\Sigma > \mathbf{O}$ as its parameter matrix, whose density is

$$f(S) = \frac{1}{2^{\frac{mp}{2}} |\Sigma|^{\frac{m}{2}} \Gamma_p(\frac{m}{2})} |S|^{\frac{m}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}, \quad S > \mathbf{O}, \quad m \geq p; \tag{i}$$

the reader may also refer to real matrix-variate gamma density discussed in Chap. 5. This is usually written as $S \sim W_p(m, \Sigma)$, $\Sigma > O$. Letting $S_{(*)} = \Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}}$, $S_{(*)} \sim W_p(m, I)$. Consider the transformation $S_{(*)} = TT'$ where $T = (t_{ij})$ is a lower triangular matrix whose diagonal elements are positive, that is, $t_{ij} = 0$, $i < j$, and $t_{jj} > 0$, $j = 1, \dots, p$. It was explained in Chaps. 1 and 3 that the t_{ij} 's are mutually independently distributed with the t_{ij} 's such that $i > j$ distributed as standard normal variables and t_{jj}^2 , as a chisquare variable having $m - (j - 1)$ degrees of freedom. The j -th diagonal element of TT' is of the form $t_{j1}^2 + \dots + t_{jj-1}^2 + t_{jj}^2$, where $t_{jk}^2 \sim \chi_1^2$, for $k = 1, \dots, j - 1$, that is, the square of a real standard normal variable. Thus, the j -th diagonal element is distributed as $\chi_1^2 + \dots + \chi_1^2 + \chi_{m-(j-1)}^2 \sim \chi_m^2$ since all the individual chisquare variables are independently distributed, in which case the resulting number of degrees of freedom is the sum of the degrees of freedom of the chisquares. Now, noting that for a χ_ν^2 ,

$$E[\chi_\nu^2] = \nu \text{ and } \text{Var}(\chi_\nu^2) = 2\nu, \quad (ii)$$

the expected value of each of the diagonal elements in TT' , which are the diagonal elements in $S_{(*)}$, will be $m = n - 1$. The non-diagonal elements in TT' result from a sum of terms of the form $t_{ik}t_{ii}$, $k < i$, whose expected value is $E[t_{ik}t_{ii}] = E[t_{ik}]E[t_{jj}]$; but since $E[t_{ik}] = 0$, $i > k$, all the non-diagonal elements will have zero as their expected values. Accordingly,

$$E[S_{(*)}] = \text{diag}(m, \dots, m) \Rightarrow E\left[\frac{S_{(*)}}{m}\right] = I \Rightarrow E\left[\frac{S}{m}\right] = \Sigma, \quad m = n - 1, \quad (iii)$$

and the estimator $\hat{\Sigma} = \frac{S}{m}$ is unbiased for Σ , m being equal to $n - 1$. Now, let us examine the covariance structure of $S_{(*)}$. Let W denote a single vector comprising all the distinct elements of $S_{(*)} = TT'$ and consider its covariance structure. In this vector of order $\frac{p(p+1)}{2} \times 1$, convert all the original t_{ij} 's and t_{jj} 's in terms of standard normal and chisquare variables. Let $z_1, \dots, z_{\frac{p(p-1)}{2}}$ be the standard normal variables and y_1, \dots, y_p denote the chisquare variables. Then, each element of $\text{Cov}(W) = [W - E(W)][W - E(W)]'$ will be a sum of terms of the type

$$[\text{Var}(y_k)][\text{Var}(z_j)] = \text{Var}(y_k) = [\text{twice the number of degrees of freedom of } y_k], \quad (iv)$$

which happens to be a linear function of m . Our estimator being $\hat{\Sigma} = \frac{S}{m} = \Sigma^{\frac{1}{2}} \frac{S_{(*)}}{m} \Sigma^{\frac{1}{2}}$, the covariance structure of $\frac{S_{(*)}}{m}$ which is $\frac{1}{m^2} \text{Cov}(W)$ tends to O when $m \rightarrow \infty$, since each element of $\text{Cov}(W)$ is of the form $am + b$ where a and b are real scalars, so that $\frac{am+b}{m^2} \rightarrow 0$ as $m \rightarrow \infty$, or equivalently, as $n \rightarrow \infty$ since $m = n - 1$. Thus, it follows from an extended version of Chebyshev's inequality that

$$Pr\left(\frac{S}{m} \rightarrow \Sigma\right) \rightarrow 1 \text{ as } m \rightarrow \infty \text{ or as } n \rightarrow \infty \text{ since } m = n - 1. \quad (v)$$

These last two results are stated next as a theorem.

Theorem 12.5.1. *Let the $p \times 1$ vectors X_j , $j = 1, \dots, n$, be iid with $E[X_j] = \mu$ and $\text{Cov}(X_j) = \Sigma$, $j = 1, \dots, n$. Assume that Σ is finite in the sense that $\|\Sigma\| < \infty$. Then, letting $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ denote the sample mean,*

$$Pr(\bar{X} \rightarrow \mu) \rightarrow 1 \text{ as } n \rightarrow \infty. \quad (12.5.5)$$

Further, letting $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$,

$$Pr\left(\hat{\Sigma} = \frac{S}{m} \rightarrow \Sigma\right) \rightarrow 1 \text{ as } m \rightarrow \infty \text{ or as } n \rightarrow \infty \text{ since } m = n - 1. \quad (12.5.6)$$

Let us now examine the criterion in (12.5.4). In this case, we can obtain an asymptotic distribution of the criterion v for large $n_{(2)}$ or when $n_{(2)} \rightarrow \infty$ in the sense that $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$. When $n_{(2)} \rightarrow \infty$, we have $\bar{X}^{(1)} \rightarrow \mu^{(1)}$, $\bar{X}^{(2)} \rightarrow \mu^{(2)}$ and $\frac{S}{n_{(2)}} \rightarrow \Sigma$, so that the criterion v in (12.5.4) becomes

$$\begin{aligned} u &= (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2} (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) \\ &= [X - \frac{1}{2} (\mu^{(1)} + \mu^{(2)})]' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}), \end{aligned} \quad (12.5.7)$$

which is nothing but u as specified in (12.3.7) with the densities $N_1(\frac{1}{2}\Delta^2, \Delta^2)$ in π_1 and $N_1(-\frac{1}{2}\Delta^2, \Delta^2)$ in π_2 . Hence, the following result:

Theorem 12.5.2. *When $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, the criterion v provided in (12.5.4) becomes u as specified in (12.5.7) with the univariate normal densities $N_1(\frac{1}{2}\Delta^2, \Delta^2)$ in π_1 and $N_1(-\frac{1}{2}\Delta^2, \Delta^2)$ in π_2 , where Δ^2 is Mahalanobis' distance given in (12.3.8). We classify X , the observation vector at hand, to π_1 when $X \in A_1$ and, to π_2 when $X \in A_2$ where $A_1 : u \geq k$ and $A_2 : u < k$ with $k = \ln \frac{C(1|2)q_2}{C(2|1)q_1}$, q_1 and q_2 being the prior probabilities of selecting the populations π_1 and π_2 , respectively, and $C(2|1)$ and $C(1|2)$ denoting the costs or loss associated with misclassification.*

In a practical situation, when n_1 and n_2 are large, we may replace Δ^2 in Theorem 12.5.2 by the corresponding sample value $n_{(2)}(\bar{X}^{(1)} - \bar{X}^{(2)})' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)})$ where $S = S_1 + S_2$ and $n_{(2)} = n_1 + n_2 - 2$ and utilize the criterion u as specified in (12.5.7) to classify the given vector X into π_1 and π_2 . It is assumed that q_1 , q_2 , $C(2|1)$ and $C(1|2)$ are available.

12.5.3. A new sample from π_1 or π_2

As in Examples 12.3.1 and 12.3.2, suppose that a simple random sample of size n_3 is available either from $\pi_1 : N_p(\mu^{(1)}, \Sigma)$ or from $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$. Letting the new sample be $X_1^{(3)}, \dots, X_{n_3}^{(3)}$, the $p \times n_3$ sample matrix, the sample mean $\bar{X}^{(3)} = \frac{1}{n_3} \sum_{j=1}^{n_3} X_j^{(3)}$, the $p \times n_3$ matrix of sample means and the sample sum of products matrix are the following:

$$\begin{aligned} \mathbf{X}^{(3)} &= [X_1^{(3)}, \dots, X_{n_3}^{(3)}], \quad \bar{\mathbf{X}}^{(3)} = [\bar{X}^{(3)}, \bar{X}^{(3)}, \dots, \bar{X}^{(3)}], \\ S_3 &= [\mathbf{X}^{(3)} - \bar{\mathbf{X}}^{(3)}][\mathbf{X}^{(3)} - \bar{\mathbf{X}}^{(3)}]' = (s_{ij}^{(3)}), \\ s_{ij}^{(3)} &= \sum_{k=1}^{n_3} (x_{ik}^{(3)} - \bar{x}_i^{(3)})(x_{jk}^{(3)} - \bar{x}_j^{(3)}). \end{aligned} \quad (12.5.8)$$

An unbiased estimate from this third sample is $\hat{\Sigma} = \frac{S_3}{n_3-1}$, as $E[\hat{\Sigma}] = \Sigma$. A pooled estimate of Σ obtained from the three samples is

$$\frac{S_1 + S_2 + S_3}{n_1 + n_2 + n_3 - 3} \equiv \frac{S}{n_{(3)}}, \quad S = S_1 + S_2 + S_3, \quad n_{(3)} = n_1 + n_2 + n_3 - 3. \quad (12.5.9)$$

Then, the criterion corresponding to (12.3.4) changes to:

$$A_1: w \geq k \text{ and } A_2: w < k, \quad k = \ln \frac{C(1|2) q_2}{C(2|1) q_1}, \quad (12.5.10)$$

where

$$w = n_{(3)}[\bar{X}^{(3)} - \frac{1}{2}(\bar{X}^{(1)} + \bar{X}^{(2)})]' S^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)}) \quad (12.5.11)$$

with $S = S_1 + S_2 + S_3$, $n_{(3)} = n_1 + n_2 + n_3 - 3$ and $\bar{X}^{(3)}$ being the sample average from the third sample, which either comes from $\pi_1 : N_p(\mu^{(1)}, \Sigma)$ or $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$. Thus, the classification rule is the following:

$$A_1: w \geq k \text{ and } A_2: w < k, \quad k = \ln \frac{C(1|2) q_2}{C(2|1) q_1}, \quad (12.5.12)$$

w being as defined in (12.5.11). That is, classify the new sample into π_1 if $w \geq k$ and, into π_2 if $w < k$.

As was explained in Sect. 12.5.2, as $n_j \rightarrow \infty$, $j = 1, 2$, $\bar{X}^{(i)} \rightarrow \mu^{(i)}$, $i = 1, 2$, and although n_3 usually remains finite, as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, we have $n_{(3)} \rightarrow \infty$ and

$\frac{S}{n^{(3)}} \rightarrow \Sigma$. Accordingly, the criterion w as given in (12.5.11) converges to w_1 for large values of n_1 and n_2 , where

$$w_1 = [\bar{X}^{(3)} - \frac{1}{2}(\mu^{(1)} + \mu^{(2)})]' \Sigma^{-1}(\mu^{(1)} - \mu^{(2)}). \quad (12.5.13)$$

Compared to u as specified in (12.3.7), the only difference is that X associated with u is replaced by $\bar{X}^{(3)}$ in w_1 . Hence, the variance in u will be multiplied by $\frac{1}{n_3}$, and the asymptotic distributions will be as follows:

$$w_1|\pi_1 \sim N_1\left(\frac{1}{2}\Delta^2, \frac{1}{n_3}\Delta^2\right) \text{ and } w_1|\pi_2 \sim N_1\left(-\frac{1}{2}\Delta^2, \frac{1}{n_3}\Delta^2\right), \quad (12.5.14)$$

as $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$.

Theorem 12.5.3. Consider two populations $\pi_1 : N_p(\mu^{(1)}, \Sigma)$ and $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$, and simple random samples of respective sizes n_1 and n_2 from these two populations. Suppose that a simple random sample of size n_3 is available, either from π_1 or π_2 . For classifying the third sample into π_1 or π_2 , the criterion to be utilized is w as given in (12.5.11). Then, the asymptotic distribution of w , when $n_i \rightarrow \infty$, $i = 1, 2$, is that of w_1 specified in (12.5.13) and the regions of classification are as given in (12.5.12).

In a practical situation, when the sample sizes n_1 and n_2 are large, one may replace Δ^2 by its sample analogue, and then use (12.5.14) to reach a decision. As it turns out, it proves quite difficult to derive the exact density of w .

Example 12.5.1. A certain milk collection and distribution center collects and sells the milk supplied by local farmers to the community, the balance, if any, being dispatched to a nearby city. In that locality, there are two types of cows. Some farmers only keep Jersey cows and others, only Holstein cows. Samples of the same quantities of milk are taken and the following characteristics are evaluated: x_1 , the fat content, x_2 , the glucose content, and x_3 , the protein content. It is known that $X' = (x_1, x_2, x_3)$ is normally distributed as $X \sim N_3(\mu^{(1)}, \Sigma)$, $\Sigma > O$, for Jersey cows, and $X \sim N_3(\mu^{(2)}, \Sigma)$, $\Sigma > O$, for Holstein cows, with $\mu^{(1)} \neq \mu^{(2)}$, the covariance matrices Σ being assumed identical. These parameters which are not known, are estimated on the basis of 100 milk samples from Jersey cows and 102 samples from Holstein cows, all the samples being of equal volume. The following are the summarized data with our standard notations, where S_1 and S_2 are the sample sums of products matrices:

$$\bar{X}^{(1)} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \quad \bar{X}^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 50 & -50 & 50 \\ -50 & 100 & 0 \\ 50 & 0 & 150 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 150 & -150 & 150 \\ -150 & 300 & 0 \\ 150 & 0 & 450 \end{bmatrix}.$$

Three farmers just brought in their supply of milk and (1): a sample denoted by X_1 is collected from the first farmer's supply and evaluated; (2) a sample, X_2 , is taken from a second farmer's supply and evaluated; (3) a set of 5 random samples are collected from a third farmer's supply, the sample average being \bar{X} . The data is

$$X_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \bar{X} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad n = 5.$$

Classify, X_1 , X_2 and the sample of size 5 to either coming from Jersey or Holstein cows.

Solution 12.5.1. The following preliminary calculations are needed:

$$\frac{S}{n_1 + n_2 - 2} = \frac{S_1 + S_2}{n_1 + n_2 - 2} = \frac{S_1 + S_2}{200} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix},$$

$$\left(\frac{S}{200}\right)^{-1} = \begin{bmatrix} 6 & 3 & -2 \\ 3 & 2 & -1 \\ -2 & -1 & 1 \end{bmatrix}, \quad \bar{X}^{(1)} - \bar{X}^{(2)} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \bar{X}^{(1)} + \bar{X}^{(2)} = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}.$$

Then,

$$(\bar{X}^{(1)} - \bar{X}^{(2)})' \left(\frac{S}{200}\right)^{-1} = [1, -1, 0] \begin{bmatrix} 6 & 3 & -2 \\ 3 & 2 & -1 \\ -2 & -1 & 1 \end{bmatrix} = [3, 1, -1],$$

$$\frac{1}{2}(\bar{X}^{(1)} - \bar{X}^{(2)})' \left(\frac{S}{200}\right)^{-1} (\bar{X}^{(1)} + \bar{X}^{(2)}) = \frac{1}{2}[3, 1, -1] \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} = 4,$$

$$(\bar{X}^{(1)} - \bar{X}^{(2)})' \left(\frac{S}{200}\right)^{-1} X = [3, 1, -1]X = 3x_1 + x_2 - x_3 \Rightarrow w = 3x_1 + x_2 - x_3 - 4$$

where the w is given in (12.5.11). For answering (1), we substitute X_1 to X in w . That is, w at X_1 is $3(2) + (1) - (1) - 4 = 2 > 0$. Hence, we assign X_1 to Jersey cows. For answering (2), we replace X in w by X_2 , that is, $3(1) + (1) - (2) - 4 = -2 < 0$. Thus, we assign X_2 to Holstein cows. For answering (3), we replace X in w by \bar{X} . That is, $3(2) + (2) - (1) - 4 = 3 > 0$. Accordingly, we classify this sample as coming from Jersey cows.

12.6. Maximum Likelihood Method of Classification

As before, let π_1 be the p -variate real normal population $N_p(\mu^{(1)}, \Sigma)$, $\Sigma > O$, with the simple random sample $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ of size n_1 drawn from that population,

and $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$, with the simple random sample $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ of size n_2 so distributed. A p -vector X at hand is to be classified into π_1 or π_2 . Let the sample means and the sample sums of products matrices be $\bar{X}^{(1)}$, $\bar{X}^{(2)}$, S_1 and S_2 . Then, the problem of classification of X into π_1 or π_2 can be stated in terms of testing a hypothesis of the following type: X and $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ are from $N_p(\mu^{(1)}, \Sigma)$ and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ are from π_2 constitutes the null hypothesis, versus, the alternative X and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ are from $N_p(\mu^{(2)}, \Sigma)$ and $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ are from $N_p(\mu^{(1)}, \Sigma)$. Let the likelihood functions under the null and alternative hypotheses be denoted as L_0 and L_1 , respectively, where

$$L_0 = \left\{ \prod_{j=1}^{n_1} \frac{e^{-\frac{1}{2}(X_j - \mu^{(1)})' \Sigma^{-1} (X_j - \mu^{(1)})}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \right\} \frac{e^{-\frac{1}{2}(X - \mu^{(1)})' \Sigma^{-1} (X - \mu^{(1)})}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \\ \times \left\{ \prod_{j=1}^{n_2} \frac{e^{-\frac{1}{2}(X_j - \mu^{(2)})' \Sigma^{-1} (X_j - \mu^{(2)})}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \right\},$$

$$L_0 = \frac{e^{-\frac{1}{2}\rho_1}}{(2\pi)^{\frac{(n_1+n_2+1)p}{2}} |\Sigma|^{\frac{n_1+n_2+1}{2}}}, \quad \rho_1 = v + (X - \mu^{(1)})' \Sigma^{-1} (X - \mu^{(1)}), \quad (i)$$

$$L_1 = \frac{e^{-\frac{1}{2}\rho_2}}{(2\pi)^{\frac{(n_1+n_2+1)p}{2}} |\Sigma|^{\frac{n_1+n_2+1}{2}}}, \quad \rho_2 = v + (X - \mu^{(2)})' \Sigma^{-1} (X - \mu^{(2)}), \quad (ii)$$

where

$$v = \text{tr}(\Sigma^{-1} S_1) + \frac{n_1}{2} (\bar{X}^{(1)} - \mu^{(1)})' \Sigma^{-1} (\bar{X}^{(1)} - \mu^{(1)}) \\ + \text{tr}(\Sigma^{-1} S_2) + \frac{n_2}{2} (\bar{X}^{(2)} - \mu^{(2)})' \Sigma^{-1} (\bar{X}^{(2)} - \mu^{(2)}) \quad (iii)$$

and S_1 and S_2 are the sample sums of products matrices from the samples $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$, respectively. Referring to Chaps. 1 and 3 for vector/matrix derivatives and the maximum likelihood estimators (MLE's) of the parameters of normal populations, the MLE's obtained from (i) are the following, denoting the estimators/estimates with a hat: The MLE's under L_0 are the following:

$$\hat{\mu}^{(1)} = \frac{n_1 \bar{X}^{(1)} + X}{n_1 + 1}, \quad \hat{\mu}^{(2)} = \bar{X}^{(2)}, \quad \hat{\Sigma} = \frac{S_1 + S_2 + S_3^{(1)}}{n_1 + n_2 + 1} \equiv \hat{\Sigma}_1, \\ S_3^{(1)} = (X - \hat{\mu}^{(1)})(X - \hat{\mu}^{(1)})' = \left(X - \frac{n_1 \bar{X}^{(1)} + X}{n_1 + 1} \right) \left(X - \frac{n_1 \bar{X}^{(1)} + X}{n_1 + 1} \right)' \\ = \left(\frac{n_1}{n_1 + 1} \right)^2 (X - \bar{X}^{(1)})(X - \bar{X}^{(1)})', \quad (12.6.1)$$

observing that the scalar quantity

$$(X - \hat{\mu}^{(1)})' \Sigma^{-1} (X - \hat{\mu}^{(1)}) = \text{tr}(X - \hat{\mu}^{(1)})' \Sigma^{-1} (X - \hat{\mu}^{(1)}) = \text{tr}(\Sigma^{-1} S_3^{(1)}).$$

By substituting the MLE's in L_0 , we obtain the maximum of L_0 :

$$\begin{aligned} \max L_0 &= \frac{e^{-\frac{(n_1+n_2+1)p}{2}}}{(2\pi)^{\frac{(n_1+n_2+1)p}{2}} |\hat{\Sigma}_1|^{\frac{(n_1+n_2+1)}{2}}}, \\ \hat{\Sigma}_1 &= \frac{S_1 + S_2 + \left(\frac{n_1}{n_1+1}\right)^2 (X - \bar{X}^{(1)})(X - \bar{X}^{(1)})'}{n_1 + n_2 + 1}. \end{aligned} \quad (12.6.2)$$

Under L_1 , the MLE's are

$$\begin{aligned} \hat{\mu}^{(2)} &= \frac{n_2 \bar{X}^{(2)} + X}{n_2 + 1}, \quad \hat{\mu}^{(1)} = \bar{X}^{(1)}, \quad \hat{\Sigma} = \frac{S_1 + S_2 + S_3^{(2)}}{n_1 + n_2 + 1} \equiv \hat{\Sigma}_2, \\ S_3^{(2)} &= \left(\frac{n_2}{n_2+1}\right)^2 (X - \bar{X}^{(2)})(X - \bar{X}^{(2)})'. \end{aligned} \quad (12.6.3)$$

Thus,

$$\begin{aligned} \max L_1 &= \frac{e^{-\frac{(n_1+n_2+1)p}{2}}}{(2\pi)^{\frac{(n_1+n_2+1)p}{2}} |\hat{\Sigma}_2|^{\frac{n_1+n_2+1}{2}}}, \\ \hat{\Sigma}_2 &= \frac{1}{n_1 + n_2 + 1} \left[S_1 + S_2 + \left(\frac{n_2}{n_2+1}\right)^2 (X - \bar{X}^{(2)})(X - \bar{X}^{(2)})' \right]. \end{aligned}$$

Hence,

$$\begin{aligned} \lambda_1 &= \frac{\max L_0}{\max L_1} = \left(\frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} \right)^{\frac{n_1+n_2+1}{2}} \Rightarrow \lambda_1^{\frac{2}{n_1+n_2+1}} = z_1 = \frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|}, \quad \text{so that} \\ z_1 &= \frac{|S_1 + S_2 + \left(\frac{n_2}{n_2+1}\right)^2 (X - \bar{X}^{(2)})(X - \bar{X}^{(2)})'|}{|S_1 + S_2 + \left(\frac{n_1}{n_1+1}\right)^2 (X - \bar{X}^{(1)})(X - \bar{X}^{(1)})'|}. \end{aligned} \quad (12.6.4)$$

If $z_1 \geq 1$, then $\max L_0 \geq \max L_1$, which means that the likelihood of X coming from π_1 is greater than or equal to the likelihood of X originating from π_2 . Hence, we may classify X into π_1 if $z_1 \geq 1$ and classify X into π_2 if $z_1 < 1$. In other words,

$$A_1 : z_1 \geq 1 \quad \text{and} \quad A_2 : z_1 < 1. \quad (iv)$$

If we let $S = S_1 + S_2$, then $z_1 \geq 1 \Rightarrow$

$$\begin{aligned} & \left| S + \left(\frac{n_2}{n_2 + 1} \right)^2 (X - \bar{X}^{(2)})(X - \bar{X}^{(2)})' \right| \\ & \geq \left| S + \left(\frac{n_1}{n_1 + 1} \right)^2 (X - \bar{X}^{(1)})(X - \bar{X}^{(1)})' \right|. \end{aligned} \quad (v)$$

We can re-express this last inequality in a more convenient form. Expanding the following partitioned determinant in two different ways, we have the following, where S is $p \times p$ and Y is $p \times 1$:

$$\begin{aligned} \left| \begin{array}{cc} S & -Y \\ Y' & 1 \end{array} \right| &= |S + YY'| = |S| |1 + Y'S^{-1}Y| \\ &= |S|[1 + Y'S^{-1}Y], \end{aligned} \quad (vi)$$

observing that $1 + Y'S^{-1}Y$ is a scalar quantity. Accordingly, $z_1 \geq 1$ means that

$$1 + \left(\frac{n_2}{n_2 + 1} \right)^2 (X - \bar{X}^{(2)})' S^{-1} (X - \bar{X}^{(2)}) \geq 1 + \left(\frac{n_1}{n_1 + 1} \right)^2 (X - \bar{X}^{(1)})' S^{-1} (X - \bar{X}^{(1)}).$$

That is,

$$\begin{aligned} z_2 &= \left(\frac{n_2}{n_2 + 1} \right)^2 (X - \bar{X}^{(2)})' S^{-1} (X - \bar{X}^{(2)}) \\ &\quad - \left(\frac{n_1}{n_1 + 1} \right)^2 (X - \bar{X}^{(1)})' S^{-1} (X - \bar{X}^{(1)}) \geq 0 \Rightarrow \\ z_3 &= \left(\frac{n_2}{n_2 + 1} \right)^2 (X - \bar{X}^{(2)})' \left(\frac{S}{n_1 + n_2 - 2} \right)^{-1} (X - \bar{X}^{(2)}) \\ &\quad - \left(\frac{n_1}{n_1 + 1} \right)^2 (X - \bar{X}^{(1)})' \left(\frac{S}{n_1 + n_2 - 2} \right)^{-1} (X - \bar{X}^{(1)}) \geq 0. \end{aligned} \quad (12.6.5)$$

Hence, the regions of classification are the following:

$$A_1 : z_3 \geq 0 \text{ and } A_2 : z_3 < 0. \quad (vii)$$

Thus, classify X into π_1 when $z_3 \geq 0$ and, X into π_2 when $z_3 < 0$. For large n_1 and n_2 , some interesting results ensue. When $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, we have $\frac{n_i}{n_i + 1} \rightarrow 1$, $i = 1, 2$, $\bar{X}^{(i)} \rightarrow \mu^{(i)}$, $i = 1, 2$, and $\frac{S}{n_1 + n_2 - 2} \rightarrow \Sigma$. Then, z_3 converges to z_4 where

$$\begin{aligned} z_4 &= \frac{1}{2} (X - \mu^{(2)})' \Sigma^{-1} (X - \mu^{(2)}) - (X - \mu^{(1)})' \Sigma^{-1} (X - \mu^{(1)}) \geq 0 \quad (viii) \\ &\Rightarrow [X - \frac{1}{2}(\mu^{(1)} + \mu^{(2)})]' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \geq 0 \Rightarrow u \geq 0 \end{aligned}$$

where u is the same criterion u as that specified in (12.5.7). Hence, we have the following result:

Theorem 12.6.1. Let $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ be a simple random sample of size n_1 from $\pi_1 : N_p(\mu^{(1)}, \Sigma)$, $\Sigma > O$ and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ be a simple random sample of size n_2 from the population $\pi_2 : N_p(\mu^{(2)}, \Sigma)$, $\Sigma > O$. Letting X be a vector at hand to be classified into π_1 or π_2 , when $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, the likelihood ratio criterion for classification is the following: Classify X into π_1 if $u \geq 0$ and, X into π_2 if $u < 0$ or equivalently, $A_1 : u \geq 0$ and $A_2 : u < 0$ where $u = [X - \frac{1}{2}(\mu^{(1)} + \mu^{(2)})]' \Sigma^{-1}(\mu^{(1)} - \mu^{(2)})$ whose density is $u \sim N_1(\frac{1}{2}\Delta^2, \Delta^2)$ when X is assigned to π_1 and $u \sim N_1(-\frac{1}{2}\Delta^2, \Delta^2)$ when X is assigned to π_2 , with $\Delta^2 = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1}(\mu^{(1)} - \mu^{(2)})$ denoting Mahalanobis' distance.

The likelihood ratio criterion for classification specified in (12.6.5) can also be given the following interpretation: For large values of n_1 and n_2 , the criterion reduces to the following: $(X - \mu^{(2)})' \Sigma^{-1}(X - \mu^{(2)}) - (X - \mu^{(1)})' \Sigma^{-1}(X - \mu^{(1)}) \geq 0$ where $(X - \mu^{(2)})' \Sigma^{-1}(X - \mu^{(2)})$ is the generalized distance between X and $\mu^{(2)}$, and $(X - \mu^{(1)})' \Sigma^{-1}(X - \mu^{(1)})$ is the generalized distance between X and $\mu^{(1)}$, which means that the generalized distance between X and $\mu^{(2)}$ is larger than the generalized distance between X and $\mu^{(1)}$ when $u > 0$. That is, X is closer to $\mu^{(1)}$ than $\mu^{(2)}$ and accordingly, we classify X into π_1 , which is the case $u > 0$. Similarly, if X is closer to $\mu^{(2)}$ when compared to the distance to $\mu^{(1)}$, we assign X to π_2 , which is the case $u < 0$. The case $u = 0$ is also included in the first inequality, but only for convenience. However, when $Pr\{u = 0|\pi_i, i = 1, 2\} = 0$, replacing $u > 0$ by $u \geq 0$ is fully justified.

Note 12.6.1. The reader may refer to Example 12.3.3 for an illustration of the computations involved in connection with the probabilities of misclassification. For large values of n_1 and n_2 , one has the z_4 of (viii) as an approximation to the u appearing in the same equation as well as the u of (12.5.7) or that of Example 12.3.3. In order to apply Theorem 12.6.1, one needs to know the parameters $\mu^{(1)}$, $\mu^{(2)}$ and Σ . When they are not available, one may substitute to them the corresponding estimates $\bar{X}^{(1)}$, $\bar{X}^{(2)}$ and $\hat{\Sigma} = \frac{S_1 + S_2}{n_1 + n_2 - 2}$ when n_1 and n_2 are large. Then, the approximate probabilities of misclassification can be determined.

Example 12.6.1. Redo the problem considered in Example 12.5.1 by making use of the maximum likelihood procedure.

Solution 12.6.1. In order to answer the questions, we need to compute

$$z_4 = \left(\frac{n_2}{n_2 + 1}\right)^2 (X - \bar{X}^{(2)})' \left(\frac{S}{n_1 + n_2 - 1}\right)^{-1} (X - \bar{X}^{(2)}) \\ - \left(\frac{n_1}{n_1 + 1}\right)^2 (X - \bar{X}^{(1)})' \left(\frac{S}{n_1 + n_2 - 2}\right)^{-1} (X - \bar{X}^{(1)}).$$

In this case, $\frac{n_1}{n_1+1} = \frac{100}{101} \approx 1$ and $\frac{n_2}{n_2+1} = \frac{102}{103} \approx 1$ and hence, the criterion z_4 is the same as w of (12.5.4) and the decisions arrived at an Example 12.5.1 will remain unchanged in this example. Since n_1 and n_2 are large, we have reasonably accurate approximations of the parameters as

$$\bar{X}^{(1)} \rightarrow \mu^{(1)}, \bar{X}^{(2)} \rightarrow \mu^{(2)} \text{ and } \frac{S}{n_1 + n_2 - 2} \rightarrow \Sigma,$$

so that the probabilities of misclassification can be evaluated by using their estimates. The approximate distributions are then given by

$$w|\pi_1 \sim N_1(\frac{1}{2}\hat{\Delta}^2, \hat{\Delta}^2) \text{ and } w|\pi_2 \sim N_1(-\frac{1}{2}\hat{\Delta}^2, \hat{\Delta}^2)$$

where $\hat{\Delta}^2 = (\bar{X}^{(1)} - \bar{X}^{(2)})'(\frac{S}{n_1+n_2-2})^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$. From the computations done in Example 12.5.1, we have

$$(\bar{X}^{(1)} - \bar{X}^{(2)})' = [1, -1, 0], (\bar{X}^{(1)} - \bar{X}^{(2)})' \left(\frac{S}{n_1 + n_2 - 2} \right)^{-1} = [3, 1, -1],$$

$$\hat{\Delta}^2 = [3, 1, -1] \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 2$$

$$\Rightarrow w|\pi_1 \sim N_1(1, 2) \text{ and } w|\pi_2 \sim N_1(-1, 2), \text{ approximately.}$$

As well, $A_1 : w \geq 0$ and $A_2 : w < 0$. For the data pertaining to (1) of Example 12.5.1, we have $w > 0$ and X_1 is assigned to π_1 . Observing that $w \rightarrow u$ of (12.5.7),

$$\begin{aligned} P(1|1, A) &= \text{Probability of arriving at a correct decision} \\ &= Pr\{u > 0|\pi_1\} = \int_0^\infty \frac{1}{\sqrt{2}\sqrt{(2\pi)}} e^{-\frac{1}{4}(u-1)^2} du \\ &= \int_{\frac{0-1}{\sqrt{2}}}^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}v^2} dv \approx 0.76; \end{aligned}$$

$$\begin{aligned} P(1|2, A) &= \text{Probability of misclassification} \\ &= Pr\{u > 0|\pi_2\} = \int_0^\infty \frac{1}{\sqrt{2}\sqrt{(2\pi)}} e^{-\frac{1}{4}(u+1)^2} du \\ &= \int_{\frac{1}{\sqrt{2}}}^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}v^2} dv \approx 0.24. \end{aligned}$$

In Example 12.5.1, the observed vector provided for (2) is classified into π_2 since $w < 0$. Thus, the probability of making the right decision is $P(2|2, A) = Pr\{u < 0|\pi_2\} \approx 0.76$

and the probability of misclassification is $P(2|1, A) = Pr\{u < 0|\pi_1\} \approx 0.24$. Given the data related to (3) of Example 12.5.1, the only difference is that the distributions in π_1 and π_2 will be slightly different, the mean values remaining the same but the variance $\hat{\Delta}^2$ being replaced by $\hat{\Delta}^2/n$ where $n = 5$. The computations are similar to those provided for (1), the sample mean being assigned to π_1 in this case.

12.7. Classification Involving k Populations

Consider the p -variate populations π_1, \dots, π_k and let X be a p -vector at hand to be classified into one of these k populations. Let q_1, \dots, q_k be the prior probabilities of selecting these populations, $q_j > 0$, $j = 1, \dots, k$, with $q_1 + \dots + q_k = 1$. Let the cost of misclassification of a p -vector belonging to π_i being improperly classified into π_j be $C(j|i)$ for $i \neq j$ so that $C(i|i) = 0$, $i = 1, \dots, k$. A decision rule $A = (A_1, \dots, A_k)$ determines subspaces $A_j \subset R_p$, $j = 1, \dots, k$, with $A_i \cap A_j = \phi$ (the empty set) for all $i \neq j$. Let the probability/density functions associated with the k populations be $P_j(X)$, $j = 1, \dots, k$, respectively. Let $P(j|i, A) = Pr\{X \in A_j|\pi_i : P_i(X), A\}$ = probability of an observation coming from or belonging to the population π_i or originating from the probability/density function $P_i(X)$, being improperly assigned to π_j or misclassified as coming from $P_j(X)$, and the cost associated with this misclassification be denoted by $C(j|i)$. Under the rule $A = (A_1, \dots, A_k)$, the probabilities of correctly classifying and misclassifying an observed vector are the following, assuming that the $P_j(X)$'s, $j = 1, \dots, k$, are densities:

$$P(i|i, A) = \int_{A_i} P_i(X)dX \quad \text{and} \quad P(j|i, A) = \int_{A_j} P_i(X)dX, \quad i, j = 1, \dots, k, \quad (i)$$

where $P(i|i, A)$ is a probability of achieving a correct classification, that is, of assigning an observation X to π_i when the population is actually π_i , and $P(j|i, A)$ is the probability of an observation X coming from π_i being misclassified as originating from π_j . Consider a p -vector X at hand. What is then the probability that this X came from $P_i(X)$, given that X is an observation vector from one of the populations π_1, \dots, π_k ? This is in fact a conditional statement involving

$$\frac{q_i P_i(X)}{q_1 P_1(X) + q_2 P_2(X) + \dots + q_k P_k(X)}.$$

Suppose that for specific i and j , the conditional probability

$$\frac{q_i P_i(X)}{q_1 P_1(X) + \dots + q_k P_k(X)} \geq \frac{q_j P_j(X)}{q_1 P_1(X) + \dots + q_k P_k(X)}. \quad (ii)$$

This is tantamount to presuming that the likeliness of X originating from $P_i(X)$ is greater than or equal to that of X coming from $P_j(X)$. In this case, we would like to assign X to

π_i rather than π_j . If (ii) holds for all $j = 1, \dots, k, j \neq i$, then we classify X into π_i . Equation (ii) for $j = 1, \dots, k, j \neq i$, implies that

$$q_i P_i(X) \geq q_j P_j(X) \Rightarrow \frac{P_i(X)}{P_j(X)} \geq \frac{q_j}{q_i}, \quad j = 1, \dots, k, \quad j \neq i. \quad (12.7.1)$$

Accordingly, we adopt (12.7.1) as a decision rule $A = (A_1, \dots, A_k)$. This decision rule corresponds to the following: When $X \in A_1 \subset R_p$ or X falls in A_1 , then X is classified into π_1 , when $X \in A_2$, then X is assigned to π_2 , and so on. What is the expected cost of an X belonging to π_i being misclassified into π_j under some decision rule $B = (B_1, \dots, B_k), B_j \subset R_p, j = 1, \dots, k, B_i \cap B_j = O, i \neq j$, for all i and j ? This is $q_i P_i(X) C(j|i) \equiv E_i(B)$. The expected cost of an X belonging to π_j being misclassified into π_i under the same decision rule B is $E_j(B) = q_j P_j(X) C(i|j)$. If $E_i(B) < E_j(B)$, then we favor $P_i(X)$ over $P_j(X)$ as it is always desirable to minimize the expected cost in any procedure or decision. If $E_i(B) < E_j(B)$ for all $j = 1, \dots, k, j \neq i$, then $P_i(X)$ or π_i is preferred over all other populations to which X could be assigned. Note that

$$E_i(B) < E_j(B) \Rightarrow q_i P_i(X) C(j|i) < q_j P_j(X) C(i|j) \Rightarrow \frac{P_i(X)}{P_j(X)} < \frac{q_j C(i|j)}{q_i C(j|i)}, \quad (iii)$$

for $j = 1, \dots, k, j \neq i$, so that (iii) is the situation resulting from the following misclassification rule: if

$$\frac{P_i(X)}{P_j(X)} \geq \frac{q_j C(i|j)}{q_i C(j|i)}, \quad j = 1, \dots, k, \quad j \neq i, \quad (12.7.2)$$

we classify X into π_i or equivalently, $X \in A_i$, which is the decision rule $A = (A_1, \dots, A_k)$. Thus, the decision rule B in (iii) is identical to A . Observing that when $C(i|j) = C(j|i)$, (12.7.2) reduces to (12.7.1); the decision rule $A = (A_1, \dots, A_k)$ in (12.7.1) is seen to yield the maximum probability of assigning an observation X at hand to π_i compared to the probability of assigning X to any other $\pi_j, j = 1, \dots, k, j \neq i$, when the costs of misclassification are equal. As well, it follows from (12.7.2) that the decision rule $A = (A_1, \dots, A_k)$ gives the minimum expected cost associated with assigning the observation X at hand to π_i compared to assigning X to any other population $\pi_j, j = 1, \dots, k, j \neq i$.

12.7.1. Classification when the populations are real Gaussian

Let the populations be p -variate real normal, that is, $\pi_j \sim N_p(\mu^{(j)}, \Sigma), \Sigma > O, j = 1, \dots, k$, with different mean value vectors but the same covariance matrix $\Sigma > O$. Let the density of π_j be denoted by $P_j(X) \simeq N_p(\mu^{(j)}, \Sigma), \Sigma > O$. A vector X at hand is to be assigned to one of the π_i 's, $i = 1, \dots, k$. In Sect. 12.3 or Example 12.3.3, the decision

rule involves two populations. Letting the two populations be $\pi_i : P_i(X)$ and $\pi_j : P_j(X)$ for specific i and j , it was determined that the decision rule consists of classifying X into π_i if $\ln \frac{P_i(X)}{P_j(X)} \geq \ln \rho$, $\rho = \frac{q_j C(i|j)}{q_i C(j|i)}$, with $\rho = 1$ so that $\ln \rho = 0$ whenever $C(i|j) = C(j|i)$ and $q_i = q_j$. When $\ln \rho = 0$, we have seen that the decision rule is to classify the p -vector X into π_i or $P_i(X)$ if $u_{ij}(X) \geq 0$ and to assign X to $P_j(X)$ or π_j if $u_{ij}(X) < 0$, where

$$\begin{aligned} u_{ij}(X) &= (\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} X - \frac{1}{2} (\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} (\mu^{(i)} + \mu^{(j)}) \\ &= [X - \frac{1}{2} (\mu^{(i)} + \mu^{(j)})]' \Sigma^{-1} (\mu^{(i)} - \mu^{(j)}). \end{aligned} \quad (iv)$$

Now, on applying the result obtained in (iv) to (12.7.1) and (12.7.2), one arrives at the following decision rule:

$$A_i : u_{ij}(X) \geq 0 \text{ or } A_i : u_{ij}(X) \geq \ln k, \quad k = \frac{q_j C(i|j)}{q_i C(j|i)}, \quad j = 1, \dots, k, \quad j \neq i, \quad (12.7.3)$$

with $\ln \rho = 0$ occurring when $q_i = q_j$ and $C(i|j) = C(j|i)$.

Note 12.7.1. What will interchanging i and j in $u_{ij}(X)$ entail? Note that, as defined, $u_{ij}(X)$ involves the terms $(\mu^{(i)} - \mu^{(j)}) = -(\mu^{(j)} - \mu^{(i)})$ and $(\mu^{(i)} + \mu^{(j)})$, the latter being unaffected by the interchange of $\mu^{(i)}$ and $\mu^{(j)}$. Hence, for all i and j ,

$$u_{ij}(X) = -u_{ji}(X), \quad i \neq j. \quad (12.7.4)$$

When the underlying population is $X \sim N_p(\mu^{(i)}, \Sigma)$, $E[u_{ij}(X)|\pi_i] = \frac{1}{2} \Delta_{ij}^2$, which implies that $E[u_{ji}|\pi_i] = -\frac{1}{2} \Delta_{ij}^2 = -E[u_{ij}(X)|\pi_i]$ where $\Delta_{ij}^2 = (\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} (\mu^{(i)} - \mu^{(j)})$.

Note 12.7.2. For computing the probabilities of correctly classifying and misclassifying an observed vector, certain assumptions regarding the distributions associated with the populations π_j , $j = 1, \dots, k$, are needed, the normality assumption being the most convenient one.

Example 12.7.1. A certain milk collection and distribution center collects and sells the milk supplied by local farmers to the community, the balance, if any, being dispatched to a nearby city. In that locality, there are three dairy cattle breeds, namely, Jersey, Holstein and Guernsey, and each farmer only keeps one type of cows. Samples are taken and the following characteristics are evaluated in grams per liter: x_1 , the fat content, x_2 , the glucose content, and x_3 , the protein content. It has been determined that $X' = (x_1, x_2, x_3)$

is normally distributed as $X \sim N_3(\mu^{(1)}, \Sigma)$ for Jersey cows, $X \sim N_3(\mu^{(2)}, \Sigma)$ for Holstein cows and $X \sim N_3(\mu^{(3)}, \Sigma)$ for Guernsey cows, with a common covariance matrix $\Sigma > O$, where

$$\mu^{(1)} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \mu^{(2)} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \mu^{(3)} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

(1): A farmer brought in his supply of milk from which one liter was collected. The three variables were evaluated, the result being $X'_0 = (2, 3, 4)$. (2): Another one liter sample was taken from a second farmer's supply and it was determined that the vector of the resulting measurements was $X'_1 = (2, 2, 2)$. No prior probabilities or costs are involved. Which breed of dairy cattle is each of these farmers likely to own?

Solution 12.7.1. Our criterion is based on $u_{ij}(X)$ where

$$u_{ij}(X) = (\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} X - \frac{1}{2}(\mu^{(i)} - \mu^{(j)})' \Sigma^{-1}(\mu^{(i)} + \mu^{(j)}).$$

Let us evaluate the various quantities of interest:

$$\begin{aligned} \mu^{(1)} - \mu^{(2)} &= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mu^{(1)} - \mu^{(3)} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}, \mu^{(2)} - \mu^{(3)} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix}, \\ \mu^{(1)} + \mu^{(2)} &= \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix}, \mu^{(1)} + \mu^{(3)} = \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix}, \mu^{(2)} + \mu^{(3)} = \begin{bmatrix} 3 \\ 6 \\ 5 \end{bmatrix}; \end{aligned}$$

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix};$$

$$\begin{aligned} A_1 &: \{u_{12}(X) \geq 0, u_{13}(X) \geq 0\}, A_2 : \{u_{21}(X) \geq 0, u_{23}(X) \geq 0\}, \\ A_3 &: \{u_{31}(X) \geq 0, u_{32}(X) \geq 0\}; \end{aligned}$$

$$\begin{aligned} (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X &= (1, 0, -1) \Sigma^{-1} X = (\frac{1}{3}, -1, -2) X = \frac{1}{3}x_1 - x_2 - 2x_3 \\ (\mu^{(1)} - \mu^{(3)})' \Sigma^{-1} X &= (0, 0, -2) \Sigma^{-1} X = (0, -2, -4) X = -2x_2 - 4x_3 \\ (\mu^{(2)} - \mu^{(3)})' \Sigma^{-1} X &= (-1, 0, -1) \Sigma^{-1} X = (-\frac{1}{3}, -1, -2) X = -\frac{1}{3}x_1 - x_2 - 2x_3; \end{aligned}$$

$$\begin{aligned} \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(2)}) &= \frac{1}{2}[\frac{1}{3}, -1, -2] \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix} = -\frac{11}{2} \\ \frac{1}{2}(\mu^{(1)} - \mu^{(3)})' \Sigma^{-1} (\mu^{(1)} + \mu^{(3)}) &= \frac{1}{2}[0, -2, -4] \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix} = -14 \\ \frac{1}{2}(\mu^{(2)} - \mu^{(3)})' \Sigma^{-1} (\mu^{(2)} + \mu^{(3)}) &= \frac{1}{2}[-\frac{1}{3}, -1, -2] \begin{bmatrix} 3 \\ 6 \\ 5 \end{bmatrix} = -\frac{17}{2}. \end{aligned}$$

Hence,

$$\begin{aligned} u_{12}(X) &= \frac{1}{3}x_1 - x_2 - 2x_3 + \frac{11}{2}; & u_{13}(X) &= -2x_2 - 4x_3 + 14; \\ u_{21}(X) &= -\frac{1}{3}x_1 + x_2 + 2x_3 - \frac{11}{2}; & u_{23}(X) &= -\frac{1}{3}x_1 - x_2 - 2x_3 + \frac{17}{2}; \\ u_{31}(X) &= 2x_2 + 4x_3 - 14; & u_{32}(X) &= \frac{1}{3}x_1 + x_2 + 2x_3 - \frac{17}{2}. \end{aligned}$$

In order to answer (1), we substitute X_0 to X and first, evaluate $u_{12}(X_0)$ and $u_{13}(X_0)$ to determine whether they are ≥ 0 . Since $u_{12}(X_0) = \frac{1}{3}(2) - (3) - 2(4) + \frac{11}{2} < 0$, the condition is violated and hence we need not check for $u_{13}(X_0) \geq 0$. Thus, X_0 is not in A_1 . Now, consider $u_{21}(X_0) = -\frac{1}{3}(2) + 3 + 2(4) - \frac{11}{2} > 0$ and $u_{23}(X_0) = -\frac{1}{3}(2) - (3) - 2(4) - \frac{17}{2} < 0$; again the condition is violated and we deduce that X_0 is not in A_2 . Finally, we verify A_3 : $u_{31}(X_0) = 2(3) + 2(4) - 14 = 0$ and $u_{32}(X_0) = \frac{1}{3}(2) + (3) + 2(4) - \frac{17}{2} > 0$. Thus, $X_0 \in A_3$, that is, we conclude that the sample milk came from Guernsey cows.

For answering (2), we substitute X_1 to X in $u_{ij}(X)$. Noting that $u_{12}(X_1) = \frac{1}{3}(2) - (2) - 2(2) + \frac{11}{2} > 0$ and $u_{13}(X_1) = -2(2) - 4(2) + 14 > 0$, we can surmise that $X_1 \in A_1$, that is, the sample milk came from Jersey cows. Let us verify A_2 and A_3 to ascertain that no mistake has been made in the calculations. Since $u_{21}(X_1) < 0$, X_1 is not in A_2 , and since $u_{31}(X_0) < 0$, X_1 is not in A_3 . This completes the computations.

12.7.2. Some distributional aspects

For computing the probabilities of correctly classifying and misclassifying an observation, we require the distributions of our criterion $u_{ij}(X)$. Let the populations be normally distributed, that is, $\pi_j \sim N_p(\mu^{(j)}, \Sigma)$, $\Sigma > O$, with the same covariance matrix Σ for all k populations, $j = 1, \dots, k$. Then, the probability of achieving a correct classification when X is assigned to π_i is the following under the decision rule $A = (A_1, \dots, A_k)$:

$$P(i|i, A) = \int_{A_i} P_i(X) dX \quad (12.7.5)$$

where $dX = dx_1 \wedge \dots \wedge dx_p$ and the integral is actually a multiple integral. But A_i is defined by the inequalities $u_{i1}(X) \geq 0, u_{i2}(X) \geq 0, \dots, u_{ik}(X) \geq 0$, where $u_{ii}(X)$ is excluded. This is the case when no prior probabilities and costs are involved or when the prior probabilities are equal and the cost functions are identical. Otherwise, the region is $\{A_i : u_{ij}(X) \geq \ln k_{ij}, k_{ij} = \frac{q_j C(i|j)}{q_i C(j|i)}, j = 1, \dots, k, j \neq i\}$. Integrating (12.7.5) is challenging as the region is determined by $k - 1$ inequalities.

When the parameters $\mu^{(j)}, j = 1, \dots, k$, and Σ are known, we can evaluate the joint distributions of $u_{ij}(X), j = 1, \dots, k, j \neq i$, under the normality assumption for $\pi_j, j = 1, \dots, k$. Let us examine the distributions of $u_{ij}(X)$ for normally distributed $\pi_i : P_i(X), i = 1, \dots, k$. In this instance, $E[X]|\pi_i = \mu^{(i)}$, and under π_i ,

$$\begin{aligned} E[u_{ij}(X)]|\pi_i &= (\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} \mu^{(i)} - \frac{1}{2}(\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} (\mu^{(i)} + \mu^{(j)}) \\ &= \frac{1}{2}(\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} (\mu^{(i)} - \mu^{(j)}) = \frac{1}{2} \Delta_{ij}^2; \\ \text{Var}(u_{ij}(X))|\pi_i &= \text{Var}[(\mu^{(i)} - \mu^{(j)})' \Sigma^{-1} X] = \Delta_{ij}^2. \end{aligned}$$

Since $u_{ij}(X)$ is a linear function of the vector normal variable X , it is normal and the distribution of $u_{ij}(X)|\pi_i$ is

$$u_{ij}(X) \sim N_1(\frac{1}{2} \Delta_{ij}^2, \Delta_{ij}^2), j = 1, \dots, k, j \neq i. \tag{12.7.6}$$

This normality holds for each $j, j = 1, \dots, k, j \neq i$, and for a fixed i . Then, we can evaluate the joint density of $u_{i1}(X), u_{i2}(X), \dots, u_{ik}(X)$, excluding $u_{ii}(X)$, and we can evaluate $P(i|i, A)$ from this joint density. Observe that for $j = 1, \dots, k, j \neq i$, the $u_{ij}(X)$'s are linear functions of the same vector normal variable X and hence, they have a joint normal distribution. In that case, the mean value vector is a $(k - 1)$ -vector, denoted by $\mu^{(ii)}$, whose elements are $\frac{1}{2} \Delta_{ij}^2, j = 1, \dots, k, j \neq i$, for a fixed i , or equivalently,

$$\mu^{(ii)} = [\frac{1}{2} \Delta_{i1}^2, \dots, \frac{1}{2} \Delta_{ik}^2] = E[U'_{ii}] \text{ with } U'_{ii} = [u_{i1}(X), \dots, u_{ik}(X)],$$

excluding the elements $u_{ii}(X)$ and $\Delta_{ii}^2 = 0$. The subscript ii in U_{ii} indicates the region A_i and the original population $P_i(X)$. The covariance matrix of U_{ii} , denoted by Σ_{ii} , will be a $(k - 1) \times (k - 1)$ matrix of the form $\Sigma_{ii} = [\text{Cov}(u_{ir}, u_{it})] = (c_{rt}), c_{rt} = \text{Cov}(u_{ir}(X), u_{it}(X))$. The subscript ii in Σ_{ii} indicates the region A_i and the original population $P_i(X)$. Observe that for two linear functions $t_1 = C'X = c_1x_1 + \dots + c_px_p$ and $t_2 = B'X = b_1x_1 + \dots + b_px_p$, having a common covariance matrix $\text{Cov}(X) = \Sigma$, we have $\text{Var}(t_1) = C' \Sigma C, \text{Var}(t_2) = B' \Sigma B$ and $\text{Cov}(t_1, t_2) = C' \Sigma B = B' \Sigma C$. Therefore,

$$c_{rt} = (\mu^{(i)} - \mu^{(r)})' \Sigma^{-1} (\mu^{(i)} - \mu^{(t)}), i \neq r, t; \Sigma_{ii} = (c_{rt}).$$

Let the vector U_{ii} be such that $U'_{ii} = (u_{i1}(X), \dots, u_{ik}(X))$, excluding $u_{ii}(X)$. Thus, for a specific i ,

$$U_{ii} \sim N_{k-1}(\mu_{(ii)}, \Sigma_{ii}), \quad \Sigma_{ii} > O,$$

and its density function, denoted by $g_{ii}(U_{ii})$, is

$$g_{ii}(U_{ii}) = \frac{1}{(2\pi)^{\frac{k-1}{2}} |\Sigma_{ii}|^{\frac{1}{2}}} e^{-\frac{1}{2}(U_{ii}-\mu_{(ii)})' \Sigma_{ii}^{-1} (U_{ii}-\mu_{(ii)})}.$$

Then,

$$\begin{aligned} P(i|i, A) &= \int_{u_{ij}(X) \geq 0, j=1, \dots, k, j \neq i} g_{ii}(U_{ii}) dU_{ii} \\ &= \int_{u_{i1}(X)=0}^{\infty} \cdots \int_{u_{ik}(X)=0}^{\infty} g_{ii}(U_{ii}) du_{i1}(X) \wedge \dots \wedge du_{ik}(X), \end{aligned} \quad (12.7.7)$$

the differential du_{ii} being absent from dU_{ii} , which is also the case for $u_{ii}(X) \geq 0$ in the integral. If prior probabilities and cost functions are involved, then replace $u_{ij}(X) \geq 0$ in the integral (12.7.7) by $u_{ij}(X) \geq \ln k_{ij}$, $k_{ij} = \frac{q_j C(i|j)}{q_i C(j|i)}$. Thus, the problem reduces to determining the joint density $g_{ii}(U_{ii})$ and then evaluating the multiple integrals appearing in (12.7.7). In order to compute the probability specified in (12.7.7), we standardize the normal density by letting $V_{ii} = \Sigma_{ii}^{-\frac{1}{2}} U_{ii}$ where $V_{ii} \sim N_{k-1}(O, I)$, and with the help of this standard normal, we may compute this probability through V_{ii} . Note that (12.7.7) holds for each i , $i = 1, \dots, k$, and thus, the probabilities of achieving a correct classification, $P(i|i, A)$ for $i = 1, \dots, k$, are available from (12.7.7).

For computing probabilities of misclassification of the type $P(i|j, A)$, we can proceed as follows: In this context, the basic population is $\pi_j : P_j(X) \sim N_p(-\frac{1}{2}\Delta_{ij}^2, \Delta_{ij}^2)$, the region of integration being $A_i : \{u_{i1}(X) \geq 0, \dots, u_{ik}(X) \geq 0\}$, excluding the element $u_{ii}(X) \geq 0$. Consider the vector U_{ij} corresponding to the vector U_{ii} . In U_{ij} , i stands for the region A_i and j , for the original population $P_j(X)$. The elements of U_{ij} are the same as those of U_{ii} , that is, $U'_{ij} = (u_{i1}(X), \dots, u_{ik}(X))$, excluding $u_{ii}(X)$. We then proceed as before and compute the covariance matrix Σ_{ij} of U_{ij} in the original population $P_j(X)$. The variances of $u_{im}(X)$, $m = 1, \dots, k$, $m \neq i$, will remain the same but the covariances will be different since they depend on the mean values. Thus, $U_{ij} \sim N_{k-1}(\mu_{(ij)}, \Sigma_{ij})$, and on standardizing, one has $V_{ij} \sim N_{k-1}(O, I)$, so that the required probability $P(i|j, A)$ can be computed from the elements of V_{ij} . Note that when the prior probabilities and costs are equal,

$$\begin{aligned}
 P(i|j, A) &= \int_{u_{i1}(X) \geq 0, \dots, u_{ik}(X) \geq 0} g_{ij}(U_{ij}) \, du_{i1}(X) \wedge \dots \wedge du_{ik}(X) \\
 &= \int_{u_{i1}(X)=0}^{\infty} \dots \int_{u_{ik}(X)=0}^{\infty} g_{ij}(U_{ij}) \, dU_{ij}, \tag{12.7.8}
 \end{aligned}$$

excluding $u_{ii}(X)$ in the integral as well as the differential $du_{ii}(X)$. Thus, $dU_{ij} = du_{i1}(X) \wedge \dots \wedge du_{ik}(X)$, excluding $du_{ii}(X)$.

Example 12.7.2. Given the data provided in Example 12.7.1, what is the probability of correctly assigning X to π_1 ? That is, compute the probability $P(1|1, A)$.

Solution 12.7.2. Observe that the joint density of $u_{12}(X)$ and $u_{13}(X)$ is that of a bivariate normal distribution since $u_{12}(X)$ and $u_{13}(X)$ are linear functions of the same vector X where X has a multivariate normal distribution. In order to compute the joint bivariate normal density, we need $E[u_{1j}(X)]$, $\text{Var}(u_{1j}(X))$, $j = 2, 3$ and $\text{Cov}(u_{12}(X), u_{13}(X))$. The following quantities are evaluated from the data given in Example 12.7.1:

$$\begin{aligned}
 \text{Var}(u_{12}(X)) &= \text{Var}[(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X - \frac{1}{2}(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1}(\mu^{(1)} + \mu^{(2)})] \\
 &= \text{Var}[(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} X] = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \\
 &= [\frac{1}{3}, -1, -2] \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \frac{7}{3} \Rightarrow E[u_{12}(X)] = \frac{7}{6};
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(u_{13}(X)) &= (\mu^{(1)} - \mu^{(3)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(3)}) \\
 &= [0, -2, -4] \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} = 8 \Rightarrow E[u_{13}(X)] = 4;
 \end{aligned}$$

$$\text{Cov}(u_{12}(X), u_{13}(X)) = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(3)}) = [\frac{1}{3}, -1, -2] \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} = 4.$$

Hence, the covariance matrix of $U_{11} = \begin{bmatrix} u_{12}(X) \\ u_{13}(X) \end{bmatrix}$, denoted by Σ_{11} , is the following:

$$\begin{aligned}
 \Sigma_{11} &= \begin{bmatrix} \frac{7}{3} & 4 \\ 4 & 8 \end{bmatrix} \Rightarrow |\Sigma_{11}| = \frac{8}{3} \\
 \Sigma_{11}^{-1} &= \frac{3}{8} \begin{bmatrix} 8 & -4 \\ -4 & \frac{7}{3} \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 24 & -12 \\ -12 & 7 \end{bmatrix},
 \end{aligned}$$

where

$$\frac{1}{8} \begin{bmatrix} 24 & -12 \\ -12 & 7 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 2\sqrt{6} & 0 \\ -\sqrt{6} & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{6} & -\sqrt{6} \\ 0 & 1 \end{bmatrix} = B'B \text{ with}$$

$$B = \frac{1}{\sqrt{8}} \begin{bmatrix} 2\sqrt{6} & -\sqrt{6} \\ 0 & 1 \end{bmatrix} \Rightarrow |B| = |B'| = |\Sigma_{11}|^{-\frac{1}{2}} = \frac{2\sqrt{6}}{8}.$$

The bivariate normal density of U_{11} is the following:

$$U_{11} = \begin{bmatrix} u_{12}(X) \\ u_{13}(X) \end{bmatrix} \sim N_2(\mu_{(1)}, \Sigma_{11}), \quad \mu_{(1)} = \begin{bmatrix} 7/6 \\ 4 \end{bmatrix}, \quad (12.7.9)$$

with Σ_{11} and $\Sigma_{11}^{-1} = B'B$ as previously specified. Letting $Y = B(U_{11} - E[U_{11}])$, $Y \sim N_2(O, I)$. Note that

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad U_{11} - E[U_{11}] = \begin{bmatrix} u_{12}(X) - E[u_{12}(X)] \\ u_{13}(X) - E[u_{13}(X)] \end{bmatrix} = \begin{bmatrix} u_{12}(X) - 7/6 \\ u_{13}(X) - 4 \end{bmatrix},$$

$$y_1 = \frac{2\sqrt{6}}{\sqrt{8}}(u_{12}(X) - 7/6) - \frac{\sqrt{6}}{\sqrt{8}}(u_{13}(X) - 4),$$

$$y_2 = \frac{\sqrt{6}}{\sqrt{8}}(u_{13}(X) - 4).$$

Then,

$$B^{-1} = \frac{\sqrt{8}}{2\sqrt{6}} \begin{bmatrix} 1 & \sqrt{6} \\ 0 & 2\sqrt{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix},$$

and we have

$$\begin{bmatrix} u_{12}(X) - 7/6 \\ u_{13}(X) - 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

which yields $u_{12}(X) = \frac{7}{6} + \frac{1}{\sqrt{3}}y_1 + \sqrt{2}y_2$ and $u_{13}(X) = 4 + 2\sqrt{2}y_2$. The intersection of the two lines corresponding to $u_{12}(X) = 0$ and $u_{13}(X) = 0$ is the point $(y_1, y_2) = (\sqrt{3}(\frac{5}{6}), -\sqrt{2})$. Thus, $u_{12}(X) \geq 0$ and $u_{13}(X) \geq 0$ give $y_2 \geq -\frac{4}{2\sqrt{2}} = -\sqrt{2}$ and $\frac{7}{6} + \frac{1}{\sqrt{3}}y_1 + \sqrt{2}y_2 \geq 0$. We can express the resulting probability as $\rho_1 - \rho_2$ where

$$\rho_1 = \int_{y_2=-\sqrt{2}}^{\infty} \int_{y_1=-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(y_1^2+y_2^2)} dy_1 \wedge dy_2 = 1 - \Phi(-\sqrt{2}), \quad (12.7.10)$$

which is explicitly available, where $\Phi(\cdot)$ denotes the distribution function of a standard normal variable, and

$$\begin{aligned} \rho_2 &= \int_{y_1=-\infty}^{\sqrt{3}(\frac{5}{6})} \int_{y_2=-\sqrt{2}}^{\frac{1}{\sqrt{2}}(\frac{7}{6} + \frac{1}{\sqrt{3}}y_1)} \frac{1}{2\pi} e^{-\frac{1}{2}(y_1^2+y_2^2)} dy_1 \wedge dy_2 \\ &= \int_{y_1=-\infty}^{\sqrt{3}(\frac{5}{6})} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y_1^2} [\Phi(-\frac{1}{\sqrt{2}}(\frac{7}{6} + \frac{1}{\sqrt{3}}y_1)) - \Phi(-\sqrt{2})] dy_1 \\ &= \int_{y_1=-\infty}^{\sqrt{3}(\frac{5}{6})} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y_1^2} \Phi(-\frac{1}{\sqrt{2}}(\frac{7}{6} + \frac{1}{\sqrt{3}}y_1)) dy_1 - \Phi(\sqrt{3}(\frac{5}{6}))\Phi(-\sqrt{2}). \end{aligned} \quad (12.7.11)$$

Therefore, the required probability is

$$\begin{aligned} \rho_1 - \rho_2 &= 1 - \Phi(-\sqrt{2}) + \Phi(-\sqrt{2})\Phi(\sqrt{3}(\frac{5}{6})) \\ &\quad - \int_{y_1=-\infty}^{\sqrt{3}(\frac{5}{6})} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y_1^2} \Phi(-\frac{1}{\sqrt{2}}(\frac{7}{6} + \frac{1}{\sqrt{3}}y_1)) dy_1. \end{aligned} \quad (12.7.12)$$

Note that all quantities, except the integral, are explicitly available from standard normal tables. The integral part can be read from a bivariate normal table. If a bivariate normal table is used, then one can approximate the required probability from (12.7.9). Alternatively, once evaluated numerically, the integral is found to be equal to 0.2182 which subtracted from 0.9941, yields a probability of 0.7759 for $P(1|1, A)$.

12.7.3. Classification when the population parameters are unknown

When training samples are available from the populations π_i , $i = 1, \dots, k$, we can estimate the parameters and proceed with the classification. Let $X_j^{(i)}$, $j = 1, \dots, n_i$, be a simple random sample of size n_i from the i -th population π_i . Then, the sample average is $\bar{X}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^{(i)}$, and with our usual notations, the sample matrix, the matrix of sample means and sample sum of products matrix are the following:

$$\begin{aligned} \mathbf{X}^{(i)} &= [X_1^{(i)}, \dots, X_{n_i}^{(i)}], \quad \bar{\mathbf{X}}^{(i)} = [\bar{X}^{(i)}, \dots, \bar{X}^{(i)}], \\ S_i &= [\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)}][\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)}]', \quad i = 1, \dots, k, \end{aligned}$$

where

$$\mathbf{X}^{(i)} = \begin{bmatrix} x_{11}^{(i)} & \dots & x_{1n_i}^{(i)} \\ \vdots & \ddots & \vdots \\ x_{p1}^{(i)} & \dots & x_{pn_i}^{(i)} \end{bmatrix}, \quad S_i = (s_{rt}^{(i)}), \quad s_{rt}^{(i)} = \sum_{m=1}^{n_i} (x_{rm}^{(i)} - \bar{x}_r^{(i)})(x_{tm}^{(i)} - \bar{x}_t^{(i)}), \quad i = 1, \dots, k.$$

Note that $\mathbf{X}^{(i)}$ and $\bar{\mathbf{X}}^{(i)}$ are $p \times n_i$ matrices and $X_j^{(i)}$ is a $p \times 1$ vector for each $j = 1, \dots, n_i$, and $i = 1, \dots, k$. Let the population mean value vectors and the common covariance matrix be $\mu^{(1)}, \dots, \mu^{(k)}$, and $\Sigma > O$, respectively. Then, the unbiased estimators for these parameters are the following, identifying the estimators/estimates by a hat: $\hat{\mu}_j^{(i)} = \bar{X}^{(i)}$, $i = 1, \dots, k$, and $\hat{\Sigma} = \frac{S}{n_1 + \dots + n_k - k}$, $S = S_1 + \dots + S_k$. On replacing the population parameters by their unbiased estimators, the classification criteria $u_{ij}(X)$, $j = 1, \dots, k$, $j \neq i$, become the following: Classify an observation vector X into π_i if $\hat{u}_{ij}(X) \geq \ln k_{ij}$, $k_{ij} = \frac{q_j C(i|j)}{q_i C(j|i)}$, $j = 1, \dots, k$, $j \neq i$, or $\hat{u}_{ij} \geq 0$, $j = 1, \dots, k$, $j \neq i$, if $q_1 = \dots = q_k$, and the $C(i|j)$'s are equal $j = 1, \dots, k$, $j \neq i$, where

$$\hat{u}_{ij}(X) = (\bar{X}^{(i)} - \bar{X}^{(j)})' \hat{\Sigma}^{-1} X - \frac{1}{2} (\bar{X}^{(i)} - \bar{X}^{(j)})' \hat{\Sigma}^{-1} (\bar{X}^{(i)} + \bar{X}^{(j)}) \quad (12.7.13)$$

for $j = 1, \dots, k$, $j \neq i$. Unfortunately, the exact distribution of $\hat{u}_{ij}(X)$ is difficult to obtain even when the populations π_i 's have p -variate normal distributions. However, when $n_j \rightarrow \infty$, $\bar{X}^{(j)} \rightarrow \mu^{(j)}$, $j = 1, \dots, k$, and when $n_j \rightarrow \infty$, $j = 1, \dots, k$, $\hat{\Sigma} \rightarrow \Sigma$. Then, asymptotically, that is, when $n_j \rightarrow \infty$, $j = 1, \dots, k$, $\hat{u}_{ij}(X) \rightarrow u_{ij}(X)$, so that the theory discussed in the previous sections is applicable. As well, the classification probabilities can then be evaluated as illustrated in Example 12.7.2.

12.8. The Maximum Likelihood Method when the Population Covariances Are Equal

Consider k real normal populations $\pi_i : P_i(X) \simeq N_p(\mu^{(i)}, \Sigma)$, $\Sigma > O$, $i = 1, \dots, k$, having the same covariance matrix but different mean value vectors $\mu^{(i)}$, $i = 1, \dots, k$. A p -vector X at hand is to be classified into one of these populations π_j , $j = 1, \dots, k$. Consider a simple random sample $X_1^{(i)}, X_2^{(i)}, \dots, X_{n_i}^{(i)}$ of sizes n_i from π_i for $i = 1, \dots, k$. Employing our usual notations, the sample means, sample matrices, matrices of sample means and the sample sum of products matrices are as follows:

$$\begin{aligned} \bar{X}^{(i)} &= \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^{(i)}, \quad \mathbf{X}^{(i)} = [X_1^{(i)}, \dots, X_{n_i}^{(i)}] = \begin{bmatrix} x_{11}^{(i)} & \dots & x_{1n_i}^{(i)} \\ \vdots & \ddots & \vdots \\ x_{p1}^{(i)} & \dots & x_{pn_i}^{(i)} \end{bmatrix}, \\ \bar{\mathbf{X}}^{(i)} &= [\bar{X}^{(i)}, \dots, \bar{X}^{(i)}], \quad S^{(i)} = [\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)}][\mathbf{X}^{(i)} - \bar{\mathbf{X}}^{(i)}]', \\ S^{(i)} &= (s_{rt}^{(i)}), \quad s_{rt}^{(i)} = \sum_{m=1}^{n_i} (x_{rm}^{(i)} - \bar{x}_r^{(i)})(x_{tm}^{(i)} - \bar{x}_t^{(i)}), \quad S = S^{(1)} + S^{(2)} + \dots + S^{(k)}. \end{aligned} \quad (12.8.1)$$

Then, the unbiased estimators of the population parameters, denoted with a hat, are

$$\hat{\mu}^{(i)} = \bar{X}^{(i)}, \quad i = 1, \dots, k, \quad \text{and} \quad \hat{\Sigma} = \frac{S}{n_1 + n_2 + \dots + n_k - k}. \quad (12.8.2)$$

The null hypothesis can be taken as $X_1^{(i)}, \dots, X_{n_i}^{(i)}$ and X originating from π_i and $X_1^{(j)}, \dots, X_{n_j}^{(j)}$ coming from π_j , $j = 1, \dots, k$, $j \neq i$, the alternative hypothesis being: X and $X_1^{(j)}, \dots, X_{n_j}^{(j)}$ coming from π_j for $j = 1, \dots, k$, $j \neq i$, and $X_1^{(i)}, \dots, X_{n_i}^{(i)}$ originating from π_i . On proceeding as in Sect. 12.6, when the prior probabilities are equal and the cost functions are identical, the criterion for classification of the observed vector X to π_i for a specific i is

$$A_i : \left(\frac{n_j}{n_j + 1} \right)^2 (X - \bar{X}^{(j)})' \left(\frac{S}{n_{(k)}} \right)^{-1} (X - \bar{X}^{(j)}) - \left(\frac{n_i}{n_i + 1} \right)^2 (X - \bar{X}^{(i)})' \left(\frac{S}{n_{(k)}} \right)^{-1} (X - \bar{X}^{(i)}) \geq 0 \quad (12.8.3)$$

for $j = 1, \dots, k$, $j \neq i$, where the decision rule is $A = (A_1, \dots, A_k)$, $S = S^{(1)} + \dots + S^{(k)}$ and $n_{(k)} = n_1 + n_2 + \dots + n_k - k$. Note that (12.8.3) holds for each i , $i = 1, \dots, k$, and hence, A_1, \dots, A_k are available from (12.8.3). Thus, the vector X at hand is classified into A_i , that is, assigned to the population π_i , if the inequalities in (12.8.3) are satisfied. This statement holds for each i , $i = 1, \dots, k$. The exact distribution of the criterion in (12.8.3) is difficult to establish but the probabilities of classification can be computed from the asymptotic theory discussed in Sect. 12.7 by observing the following:

When $n_i \rightarrow \infty$, $\bar{X}^{(i)} \rightarrow \mu^{(i)}$, $i = 1, \dots, k$, and when $n_1 \rightarrow \infty, \dots, n_k \rightarrow \infty$, $\hat{\Sigma} \rightarrow \Sigma$. Thus, asymptotically, when $n_i \rightarrow \infty$, $i = 1, \dots, k$, the criterion specified in (12.8.3) reduces to the criterion (12.7.3) of Sect. 12.7. Accordingly, when $n_i \rightarrow \infty$ or for very large n_i 's, $i = 1, \dots, k$, one may utilize (12.7.3) for computing the probabilities of classification, which was illustrated in Examples 12.7.1 and 12.7.2.

12.9. Maximum Likelihood Method and Unequal Covariance Matrices

The likelihood procedure can also provide a classification rule when the normal population covariance matrices are different. For example, let $\pi_1 : P_1(X) \simeq N_p(\mu^{(1)}, \Sigma_1)$, $\Sigma_1 > O$, and $\pi_2 : P_2(X) \simeq N_p(\mu^{(2)}, \Sigma_2)$, $\Sigma_2 > O$, where $\mu^{(1)} \neq \mu^{(2)}$ and $\Sigma_1 \neq \Sigma_2$. Let a simple random sample $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ of size n_1 from π_1 and a simple random sample $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ of size n_2 from π_2 be available. Let $\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ be the sample averages and S_1 and S_2 be the sample sum of products matrices, respectively. In classification problems, there is an additional vector X which comes from π_1 under the null hypothesis and

from π_2 under the alternative. Then, the maximum likelihood estimators, denoted by a hat, will be the following:

$$\hat{\mu}^{(1)} = \bar{X}^{(1)}, \hat{\mu}^{(2)} = \bar{X}^{(2)}, \hat{\Sigma}_1 = \frac{S_1}{n_1} \text{ and } \hat{\Sigma}_2 = \frac{S_2}{n_2}, \quad (i)$$

respectively, when no additional vector is involved. However, these estimators will change in the presence of the additional vector X , where X is the vector at hand to be assigned to π_1 or π_2 . When X originates from π_1 or π_2 , $\mu^{(1)}$ and $\mu^{(2)}$ are respectively estimated as follows:

$$\hat{\mu}_*^{(1)} = \frac{n_1 \bar{X}_1 + X}{n_1 + 1} \text{ and } \hat{\mu}_*^{(2)} = \frac{n_2 \bar{X}_2 + X}{n_2 + 1}, \quad (ii)$$

and when X comes from π_1 or π_2 , Σ_1 and Σ_2 are estimated by

$$\hat{\Sigma}_{1*} = \frac{S_1 + S_3^{(1)}}{n_1 + 1} \text{ and } \hat{\Sigma}_{2*} = \frac{S_2 + S_3^{(2)}}{n_2 + 1} \quad (iii)$$

where

$$\begin{aligned} S_3^{(1)} &= (X - \hat{\mu}_*^{(1)})(X - \hat{\mu}_*^{(1)})' = \left(\frac{n_1}{n_1 + 1}\right)^2 (X - \bar{X}_1)(X - \bar{X}_1)' \\ S_3^{(2)} &= (X - \hat{\mu}_*^{(2)})(X - \hat{\mu}_*^{(2)})' = \left(\frac{n_2}{n_2 + 1}\right)^2 (X - \bar{X}_2)(X - \bar{X}_2)', \end{aligned} \quad (iv)$$

referring to the derivations provided in Sect. 12.6 when discussing maximum likelihood procedures. Thus, the null hypothesis can be X and $X_1^{(1)}, \dots, X_{n_1}^{(1)}$ are from π_1 and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ are from π_2 , versus the alternative: X and $X_1^{(2)}, \dots, X_{n_2}^{(2)}$ being from π_2 and $X_1^{(1)}, \dots, X_{n_1}^{(1)}$, from π_1 . Let L_0 and L_1 denote the likelihood functions under the null and alternative hypotheses, respectively. Observe that under the null hypothesis, Σ_1 is estimated by $\hat{\Sigma}_{1*}$ of (iii) and Σ_2 is estimated by $\hat{\Sigma}$ of (i), respectively, so that the likelihood ratio criterion λ is given by

$$\lambda = \frac{\max L_0}{\max L_1} = \frac{|\hat{\Sigma}_{2*}|^{\frac{n_2+1}{2}} |\hat{\Sigma}_1|^{\frac{n_1}{2}}}{|\hat{\Sigma}_{1*}|^{\frac{n_1+1}{2}} |\hat{\Sigma}_2|^{\frac{n_2}{2}}}. \quad (12.9.1)$$

The determinants in (12.9.1) can be represented as follows, referring to the simplifications discussed in Sect. 12.6:

$$\lambda = \frac{(n_1 + 1)^{\frac{p(n_1+1)}{2}} |S_2|^{\frac{n_2+1}{2}} [1 + (\frac{n_2}{n_2+1})^2 (X - \bar{X}_2)' S_2^{-1} (X - \bar{X}_2)]^{\frac{n_2+1}{2}} |\hat{\Sigma}_1|^{\frac{n_1}{2}}}{(n_2 + 1)^{\frac{p(n_2+1)}{2}} |S_1|^{\frac{n_1+1}{2}} [1 + (\frac{n_1}{n_1+1})^2 (X - \bar{X}_1)' S_1^{-1} (X - \bar{X}_1)]^{\frac{n_1+1}{2}} |\hat{\Sigma}_2|^{\frac{n_2}{2}}}. \quad (12.9.2)$$

The classification rule then consists of assigning the observed vector X to π_1 if $\lambda \geq 1$ and, to π_2 if $\lambda < 1$. We could have expressed the criterion in terms of $\lambda_1 = \lambda^{\frac{2}{n}}$ if $n_1 = n_2 = n$, which would have simplified the expressions appearing in (12.9.2).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 13

Multivariate Analysis of Variation



13.1. Introduction

We will employ the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

In this chapter, we only consider analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) problems involving real populations. Even though all the steps involved in the following discussion focusing on the real variable case can readily be extended to the complex domain, it does not appear that a parallel development of analysis of variance methodologies in the complex domain has yet been considered. In order to elucidate the various steps in the procedures, we will first review the univariate case. For a detailed exposition of the analysis of variance technique in the scalar variable case, the reader may refer Mathai and Haubold (2017). We will consider the cases of one-way classification or completely randomized design as well as two-way classification with-

out and with interaction or randomized block design. With this groundwork in place, the derivations of the results in the multivariate setting ought to prove easier to follow.

In the early nineteenth century, Gauss and Laplace utilized methodologies that may be regarded as forerunners to ANOVA in their analyses of astronomical data. However, this technique came to full fruition in Ronald Fisher's classic book titled "*Statistical Methods for Research Workers*", which was initially published in 1925. The principle behind ANOVA consists of partitioning the total variation present in the data into variations attributable to different sources. It is actually the total variation that is split rather than the total variance, the latter being a fraction of the former. Accordingly, the procedure could be more appropriately referred to as "analysis of variation". As has already been mentioned, we will initially consider the one-way classification model, which will then be extended to the multivariate situation.

Let us first focus on an experimental design called a completely randomized experiment. In this setting, the subject matter was originally developed for agricultural experiments, which influenced its terminology. For example, the basic experimental unit is referred to as a "plot", which is a piece of land in an agricultural context. When an experiment is performed on human beings, a plot translates into an individual. If the experiment is carried out on some machinery, then a machine corresponds to a plot. In a completely randomized experiment, a set of $n_1 + n_2 + \dots + n_k$ plots, which are homogeneous with respect to all factors of variation, are selected. Then, k treatments are applied at random to these plots, the first treatment to n_1 plots, the second treatment to n_2 plots, up to the k -th treatment being applied to n_k plots. For instance, if the effects of k different fertilizers on the yield of a certain crop are to be studied, then the treatments consist of these k fertilizers, the first treatment meaning one of the fertilizers, the second treatment, another one and so on, with the k -th treatment corresponding to the last fertilizer. If the experiment involves studying the yield of corn among k different varieties of corn, then a treatment coincides with a particular variety. If an experiment consists of comparing k teaching methods, then a treatment refers to a method of teaching and a plot corresponds to a student. When an experiment compares the effect of k different medications in curing a certain ailment, then a treatment is a medication, and so on. If the treatments are denoted by t_1, \dots, t_k , then treatment t_j is applied at random to n_j homogeneous plots or n_j homogeneous plots are selected at random and treatment t_j is applied to them, for $j = 1, \dots, k$. Random assignment is done to avoid possible biases or the influence of confounding factors, if any. Then, observations measuring the effect of these treatments on the experimental units are made. For example, in the case of various methods of teaching, the observation x_{ij} could be the final grade obtained by the j -th student who was subjected to the i -th teaching method. In the case of comparing k different varieties of corn, the observation x_{ij} could consist of the yield of corn observed at harvest time in the j -th plot which received the i -th variety of corn. Thus, in this instance, i stands for the treatment number and j represents the serial

number of the plot receiving the i -th treatment, x_{ij} being the final observation. Then, the corresponding linear additive fixed effect model is the following:

$$x_{ij} = \mu + \alpha_i + e_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k, \quad (13.1.1)$$

where μ is a general effect, α_i is the deviation from the general effect due to treatment t_i and e_{ij} is the random component, which includes the sum total contributions originating from unknown or uncontrolled factors. When the experiment is designed, the plots are selected so that they be homogeneous with respect to all possible factors of variation. The general effect μ can be interpreted as the grand average or the expected value of x_{ij} when α_i is not present or treatment t_i is not applied or has no effect. The simplest assumption that we will make is that $E[e_{ij}] = 0$ for all i and j and $\text{Var}(e_{ij}) = \sigma^2 > 0$ for all i and j and for some positive quantity σ^2 , where $E[\cdot]$ denotes the expected value of $[\cdot]$. It is further assumed that $\mu, \alpha_1, \dots, \alpha_k$ are all unknown constants. When $\alpha_1, \dots, \alpha_k$ are assumed to be random variables, model (13.1.1) is referred to as a “random effect model”. In the following discussion, we will solely consider fixed effect models. The first step consists of estimating the unknown quantities from the data. Since no distribution is assumed on the e_{ij} ’s, and thereby on the x_{ij} ’s, we will employ the method of least squares for estimating the parameters. In that case, one has to minimize the error sum of squares which is

$$\sum_{ij} e_{ij}^2 = \sum_{ij} [x_{ij} - \mu - \alpha_i]^2.$$

Applying calculus principles, we equate the partial derivatives of $\sum_{ij} e_{ij}^2$ with respect to μ to zero and then, equate the partial derivatives of $\sum_{ij} e_{ij}^2$ with respect to $\alpha_1, \dots, \alpha_k$ to zero and solve these equations. A convenient notation in this area is to represent a summation by a dot. As an example, if the subscript j is summed up, it is replaced by a dot, so that $\sum_j x_{ij} \equiv x_{i.}$; similarly, $\sum_{ij} x_{ij} \equiv x_{..}$. Thus,

$$\frac{\partial}{\partial \mu} \left[\sum_{ij} e_{ij}^2 \right] = 0 \Rightarrow -2 \sum_{ij} [x_{ij} - \mu - \alpha_i] = 0 \Rightarrow \sum_i \left(\sum_j [x_{ij} - \mu - \alpha_i] \right) = 0$$

that is,

$$\sum_i [x_{i.} - n_i \mu - n_i \alpha_i] = 0 \Rightarrow x_{..} - n. \mu - \sum_{i=1}^k n_i \alpha_i = 0,$$

and since we have taken α_i as a deviation from the general effect due to treatment t_i , we can let $\sum_i n_i \alpha_i = 0$ without any loss of generality. Then, $x_{..}/n.$ is an estimate of μ , and denoting estimates/estimators by a hat, we write $\hat{\mu} = x_{..}/n.$ Now, note that for example

α_1 appears in the terms $(x_{11} - \mu - \alpha_1)^2 + \cdots + (x_{1n_1} - \mu - \alpha_1)^2 = \sum_j (x_{1j} - \mu - \alpha_1)^2$ but does not appear in the other terms in the error sum of squares. Accordingly, for a specific i ,

$$\frac{\partial}{\partial \alpha_i} \left[\sum_{ij} e_{ij}^2 \right] = 0 \Rightarrow \sum_j [x_{ij} - \mu - \alpha_i] = 0 \Rightarrow x_{i.} - n_i \hat{\mu} - n_i \hat{\alpha}_i = 0,$$

that is, $\hat{\alpha}_i = \frac{x_{i.}}{n_i} - \hat{\mu}$. Thus,

$$\hat{\mu} = \frac{1}{n} x_{..} \quad \text{and} \quad \hat{\alpha}_i = \frac{1}{n_i} x_{i.} - \hat{\mu}. \quad (13.1.2)$$

The least squares minimum is obtained by substituting the least squares estimates of μ and α_i , $i = 1, \dots, k$, in the error sum of squares. Denoting the least squares minimum by s^2 ,

$$\begin{aligned} s^2 &= \sum_{ij} (x_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \sum_{ij} \left[x_{ij} - \frac{x_{..}}{n} - \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right) \right]^2 \\ &= \sum_{ij} \left[x_{ij} - \frac{x_{i.}}{n_i} \right]^2 = \sum_{ij} \left[x_{ij} - \frac{x_{..}}{n} \right]^2 - \sum_{ij} \left[\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right]^2. \end{aligned} \quad (13.1.3)$$

When the square is expanded, the middle term will become $-2 \sum_{ij} \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right) x_{ij}$, thus yielding the expression given in (13.1.3). As well, we have the following identity:

$$\sum_{ij} \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right)^2 = \sum_{i=1}^k n_i \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right)^2 = \sum_i \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n}.$$

Now, let us consider the hypothesis $H_o : \alpha_1 = \alpha_2 = \cdots = \alpha_k$, which is equivalent to the hypothesis $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$ since, by assumption, $\sum_i n_i \alpha_i = 0$. Proceeding as before, the least squares minimum, under the null hypothesis H_o , denoted by s_0^2 , is the following:

$$s_0^2 = \sum_{ij} \left(x_{ij} - \frac{x_{..}}{n} \right)^2$$

and hence the sum of squares due to the hypothesis or due to the presence of the α_j 's, is given by $s_0^2 - s^2 = \sum_{ij} \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right)^2$. Thus, the total variation is partitioned as follows:

$$\begin{aligned} s_0^2 &= [s_0^2 - s^2] + [s^2] \\ \sum_{ij} \left(x_{ij} - \frac{x_{..}}{n} \right)^2 &= \left[\sum_i n_i \left(\frac{x_{i.}}{n_i} - \frac{x_{..}}{n} \right)^2 \right] + \left[\sum_{ij} \left(x_{ij} - \frac{x_{i.}}{n_i} \right)^2 \right], \quad \text{that is,} \end{aligned}$$

Total variation (s_0^2) = variation due to the α_j 's ($s_0^2 - s^2$) + the residual variation (s^2),

which is the analysis of variation principle. If $e_{ij} \stackrel{iid}{\sim} N_1(0, \sigma^2)$ for all i and j where $\sigma^2 > 0$ is a constant, it follows from the chisquarredness and independence of quadratic forms, as discussed in Chaps. 2 and 3, that $\frac{s_0^2}{\sigma^2} \sim \chi_{n-1}^2$, a real chisquare variable having $n - 1$ degrees of freedom, $\frac{[s_0^2 - s^2]}{\sigma^2} \sim \chi_{k-1}^2$ under the hypothesis H_0 and $\frac{s^2}{\sigma^2} \sim \chi_{n-k}^2$, where the sum of squares due to the α_j 's, namely $s_0^2 - s^2$, and the residual sum of squares, namely s^2 , are independently distributed under the hypothesis. Usually, these findings are put into a tabular form known as the analysis of variation table or ANOVA table. The usual format is as follows:

ANOVA Table for the One-Way Classification

Variation due to (1)	df (2)	SS (3)	MS (3)/(2)
treatments	$k - 1$	$\sum_i n_i (\frac{x_i}{n_i} - \frac{x_{..}}{n})^2$	$(s_0^2 - s^2)/(k - 1)$
residuals	$n_{.} - k$	$\sum_{ij} (x_{ij} - \frac{x_i}{n_i})^2$	$s^2/(n_{.} - k)$
total	$n_{.} - 1$	$\sum_{ij} (x_{ij} - \frac{x_{..}}{n})^2$	

where df denotes the number of degrees of freedom, SS means sum of squares and MS stands for mean squares or the average of the squares. There is usually a last column which contains the F-ratio, that is, the ratio of the treatments MS to the residuals MS, and enables one to determine whether to reject the null hypothesis, in which case the test statistic is said to be “significant”, or not to reject the null hypothesis, when the test statistic is “not significant”. Further details on the real scalar variable case are available from Mathai and Haubold (2017).

In light of this brief review of the scalar variable case of one-way classification data or univariate data secured from a completely randomized design, the concepts will now be extended to the multivariate setting.

13.2. Multivariate Case of One-Way Classification Data Analysis

Extension of the results to the multivariate case is parallel to the scalar variable case. Consider a model of the type

$$X_{ij} = M + A_i + E_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k, \quad (13.2.1)$$

with X_{ij} , M , A_i and E_{ij} all being $p \times 1$ real vectors where X_{ij} denotes the j -th observation vector in the i -th group or the observed vectors obtained from the n_i plots receiving

the i -th vector of treatments, M is a general effect vector, A_i is a vector of deviations from M due to the i -th treatment vector so that $\sum_i n_i A_i = O$ since we are taking deviations from the general effect M , and E_{ij} is a vector of random components assumed to be normally distributed as follows: $E_{ij} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, for all i and j where Σ is a positive definite covariance matrix, that is,

$$\text{Cov}(E_{ij}) = E[(E_{ij} - O)(E_{ij} - O)'] = E[E_{ij}E_{ij}'] = \Sigma > O \text{ for all } i \text{ and } j,$$

where $E[\cdot]$ denotes the expected value operator. This normality assumption will be needed for testing hypotheses and developing certain distributional aspects. However, the multivariate analysis of variation can be set up without having to resort to any distributional assumption. In the real scalar variable case, we minimized the sum of the squares of the errors since the variations only involved single scalar variables. In the vector case, if we take the sum of squares of the elements in E_{ij} , that is, $E_{ij}'E_{ij}$ and its sum over all i and j , then we are only considering the variations in the individual elements of E_{ij} 's; however, in the vector case, there is joint variation among the elements of the vector and that is also to be taken into account. Hence, we should be considering all squared terms and cross product terms or the whole matrix of squared and cross product terms. This is given by $E_{ij}E_{ij}'$ and so, we should consider this matrix and carry out some type of minimization. Consider

$$\sum_{ij} E_{ij}E_{ij}' = \sum_{ij} [X_{ij} - M - A_i][X_{ij} - M - A_i]'. \quad (13.2.2)$$

For obtaining estimates of M and A_i , $i = 1, \dots, k$, we will minimize the trace of $\sum_{ij} E_{ij}E_{ij}'$ as a criterion. There are terms of the type $[X_{ij} - M - A_i]'[X_{ij} - M - A_i]$ in this trace. Thus,

$$\begin{aligned} \frac{\partial}{\partial M} \left[\text{tr} \left(\sum_{ij} E_{ij}E_{ij}' \right) \right] &= \sum_{ij} \frac{\partial}{\partial M} [X_{ij} - M - A_i]'[X_{ij} - M - A_i] = O \\ &\Rightarrow \sum_{ij} X_{ij} - n.M - \sum_i n_i A_i = O \Rightarrow \hat{M} = \frac{1}{n} X_{..}, \end{aligned}$$

noting that we assumed that $\sum_i n_i A_i = O$. Now, on differentiating the trace of $E_{ij}E_{ij}'$ with respect to A_i for a specific i , we have

$$\begin{aligned} \frac{\partial}{\partial A_i} \text{tr} \left[\sum_{ij} E_{ij}E_{ij}' \right] &= \frac{\partial}{\partial A_i} \sum_{ij} [X_{ij} - M - A_i]'[X_{ij} - M - A_i] = O \\ &\Rightarrow \sum_j [X_{ij} - M - A_i] = O \Rightarrow \hat{A}_i = \frac{1}{n_i} X_{i.} - \hat{M} = \frac{1}{n_i} X_{i.} - \frac{1}{n} X_{..}. \end{aligned}$$

Observe that there is only one critical vector for \hat{M} and for \hat{A}_i , $i = 1, \dots, k$. Accordingly, the critical point will either correspond to a minimum or a maximum of the trace. But for arbitrary M and A_i , the maximum occurs at plus infinity and hence, the critical point $(\hat{M}, \hat{A}_i, i = 1, \dots, k)$ corresponds to a minimum. Once evaluated at these estimates, the sum of squares and cross products matrix, denoted by S , is the following:

$$\begin{aligned} S &= \sum_{ij} [X_{ij} - \hat{M} - \hat{A}_i][X_{ij} - \hat{M} - \hat{A}_i]' = \sum_{ij} \left[X_{ij} - \frac{1}{n_i} X_{i.} \right] \left[X_{ij} - \frac{1}{n_i} X_{i.} \right]' \\ &= \sum_{ij} \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right] \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right]' - \sum_i n_i \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right] \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right]' \end{aligned} \quad (13.2.3)$$

Note that as in the scalar case, the middle terms and the last term will combine into the second term above. Now, let us impose the hypothesis $H_o : A_1 = A_2 = \dots = A_k = O$. Note that equality of the A_j 's will automatically imply that each one is null because the weighted sum is null as per our initial assumption in the model (13.2.1). Under this hypothesis, the model will be $X_{ij} = M + E_{ij}$, and then proceeding as in the univariate case, we end up with the following sum of squares and cross products matrix, denoted by S_0 :

$$S_0 = \sum_{ij} \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right] \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right]', \quad (13.2.4)$$

so that the sum of squares and cross products matrix due to the A_i 's is the difference

$$S_0 - S = \sum_{ij} \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right] \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right]'. \quad (13.2.5)$$

Thus, the following partitioning of the total variation in the multivariate data:

$$\begin{aligned} S_0 &= [S_0 - S] + S \\ \text{Total variation} &= [\text{Variation due to the } A_i\text{'s}] + [\text{Residual variation}] \\ \sum_{ij} \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right] \left[X_{ij} - \frac{1}{n_{.}} X_{..} \right]' &= \sum_{ij} \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right] \left[\frac{1}{n_i} X_{i.} - \frac{1}{n_{.}} X_{..} \right]' \\ &\quad + \sum_{ij} \left[X_{ij} - \frac{1}{n_i} X_{i.} \right] \left[X_{ij} - \frac{1}{n_i} X_{i.} \right]'. \end{aligned}$$

Under the normality assumption for the random component $E_{ij} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, we have the following properties, which follow from results derived in Chap. 5, the

notation $W_p(\nu, \Sigma)$ standing for a Wishart distribution having ν degrees of freedom and parameter matrix Σ :

$$\text{Total variation} = S_0 = \sum_{ij} \left[X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right] \left[X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right]',$$

$$S_0 \sim W_p(n_{\cdot} - 1, \Sigma);$$

$$\text{Variation due to the } A_i \text{'s} = S_0 - S = \sum_{ij} \left[\frac{1}{n_i} X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right] \left[\frac{1}{n_i} X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right]',$$

$$S_0 - S \sim W_p(k - 1, \Sigma) \text{ under the hypothesis } A_1 = A_2 = \dots = A_k = O;$$

$$\text{Residual variation} = S = \sum_{ij} \left[X_{ij} - \frac{1}{n_i} X_{i \cdot} \right] \left[X_{ij} - \frac{1}{n_i} X_{i \cdot} \right]',$$

$$S \sim W_p(n_{\cdot} - k, \Sigma).$$

We can summarize these findings in a tabular form known as the multivariate analysis of variation table or MANOVA table, where df means degrees of freedom in the corresponding Wishart distribution, and SSP represents the sum of squares and cross products matrix.

Multivariate Analysis of Variation (MANOVA) Table

Variation due to	df	SSP
treatments	$k - 1$	$\sum_{ij} \left[\frac{1}{n_i} X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right] \left[\frac{1}{n_i} X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right]'$
residuals	$n_{\cdot} - k$	$\sum_{ij} \left[X_{ij} - \frac{1}{n_i} X_{i \cdot} \right] \left[X_{ij} - \frac{1}{n_i} X_{i \cdot} \right]'$
total	$n_{\cdot} - 1$	$\sum_{ij} \left[X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right] \left[X_{ij} - \frac{1}{n_{\cdot}} X_{\cdot j} \right]'$

13.2.1. Some properties

The sample values from the i -th sample or the i -th group or the plots receiving the i -th treatment are $X_{i1}, X_{i2}, \dots, X_{in_i}$. In this case, the average is $\sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} = \frac{X_{i \cdot}}{n_i}$ and the i -th sample sum of squares and products matrix is

$$S_i = \sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{i \cdot}}{n_i} \right) \left(X_{ij} - \frac{X_{i \cdot}}{n_i} \right)'$$

As well, it follows from Chap. 5 that $S_i \sim W_p(n_i - 1, \Sigma)$ when $E_{ij} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$. Then, the residual sum of squares and products matrix can be written as follows, denoting it by the matrix V :

$$\begin{aligned}
 V &= \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)' = \sum_{i=1}^k \left[\sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)' \right] \\
 &= \sum_{i=1}^k S_i = S_1 + S_2 + \dots + S_k
 \end{aligned} \tag{13.2.6}$$

where $S_i \sim W_p(n_i - 1, \Sigma)$, $i = 1, \dots, k$, and the S_i 's are independently distributed since the sample values from the k groups are independently distributed among themselves (within the group) and between groups. Hence, $S \sim W_p(v, \Sigma)$, $v = (n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) = n - k$. Note that $\bar{X}_i = \frac{X_{i.}}{n_i}$ has $\text{Cov}(\bar{X}_i) = \frac{1}{n_i} \Sigma$, so that $\sqrt{n_i}(\bar{X}_i - \bar{X})$ are iid $N_p(O, \Sigma)$ where $\bar{X} = X_{..}/n_{..}$. Then, the sum of squares and products matrix due to the treatments or due to the A_i 's is the following, denoting it by U :

$$U = \sum_{i=1}^k n_i \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{..}} \right] \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n_{..}} \right]' \sim W_p(k - 1, \Sigma) \tag{13.2.7}$$

under the null hypothesis; when the hypothesis is violated, it is a noncentral Wishart distribution. Further, the sum of squares and products matrix due to the treatments and the residual sum of squares and products matrix are independently distributed. Thus, by comparing U and V , we should be able to reach a decision regarding the hypothesis. One procedure that is followed is to take the determinants of U and V and compare them. This does not have much of a basis and determinants should not be called ‘‘generalized variance’’ as previously explained since the basic condition of a norm is violated by the determinant. The basis for comparing determinants will become clear from the point of view of testing hypotheses by applying the likelihood ratio criterion, which is discussed next.

13.3. The Likelihood Ratio Criterion

Let $E_{ij} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$, and suppose that we have simple random samples of sizes n_1, \dots, n_k from the k groups relating to the k treatments. Then, the likelihood function, denoted by L , is the following:

$$\begin{aligned}
 L &= \prod_{ij} \frac{e^{-\frac{1}{2}(X_{ij}-M-A_i)' \Sigma^{-1}(X_{ij}-M-A_i)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \\
 &= \frac{e^{-\frac{1}{2} \sum_{ij} (X_{ij}-M-A_i)' \Sigma^{-1}(X_{ij}-M-A_i)}}{(2\pi)^{\frac{pn_{..}}{2}} |\Sigma|^{\frac{n_{..}}{2}}}.
 \end{aligned} \tag{13.3.1}$$

The maximum likelihood estimators/estimates (MLE's) of M is $\hat{M} = \frac{X_{..}}{n} = \bar{X}$ and that of A_i is $\hat{A}_i = \frac{X_{i.}}{n_i} - \hat{M}$. With a view to obtaining the MLE of Σ , we first note that the exponent is a real scalar quantity which is thus equal to its trace, so that we can express the exponent as follows, after substituting the MLE's of M and A_i :

$$\begin{aligned} & -\frac{1}{2} \sum_{ij} [X_{ij} - \hat{M} - \hat{A}_i]' \Sigma^{-1} [X_{ij} - \hat{M} - \hat{A}_i] \\ &= -\frac{1}{2} \sum_{ij} \text{tr} \left(\left[X_{ij} - \frac{X_{i.}}{n_i} \right]' \Sigma^{-1} \left[X_{ij} - \frac{X_{i.}}{n_i} \right] \right) \\ &= -\frac{1}{2} \sum_{ij} \text{tr} \left(\Sigma^{-1} \left[X_{ij} - \frac{X_{i.}}{n_i} \right] \left[X_{ij} - \frac{X_{i.}}{n_i} \right]' \right). \end{aligned}$$

Now, following through the estimation procedure of the MLE included in Chap. 3, we obtain the MLE of Σ as

$$\hat{\Sigma} = \frac{1}{n} \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)'. \quad (13.3.2)$$

After substituting \hat{M} , \hat{A}_i and $\hat{\Sigma}$, the exponent in the likelihood ratio criterion λ becomes $-\frac{1}{2} n \cdot \text{tr}(I_p) = -\frac{1}{2} n \cdot p$. Hence, the maximum value of the likelihood function L under the general model becomes

$$\max L = \frac{e^{-\frac{1}{2}(n \cdot p)} n^{\frac{n \cdot p}{2}}}{(2\pi)^{\frac{n \cdot p}{2}} \left| \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)' \right|^{\frac{n}{2}}}. \quad (13.3.3)$$

Under the hypothesis $H_o : A_1 = A_2 = \dots = A_p = O$, the model is $X_{ij} = M + E_{ij}$ and the MLE of M under H_o is still $\frac{1}{n} X_{..}$ and $\hat{\Sigma}$ under H_o is $\frac{1}{n} \sum_{ij} \left(X_{ij} - \frac{1}{n} X_{..} \right) \left(X_{ij} - \frac{1}{n} X_{..} \right)'$, so that $\max L$ under H_o , denoted by $\max L_o$, is

$$\max L_o = \frac{e^{-\frac{1}{2} n \cdot p} n^{\frac{n \cdot p}{2}}}{(2\pi)^{\frac{n \cdot p}{2}} \left| \sum_{ij} \left(X_{ij} - \frac{1}{n} X_{..} \right) \left(X_{ij} - \frac{1}{n} X_{..} \right)' \right|^{\frac{n}{2}}}. \quad (13.3.4)$$

Therefore, the λ -criterion is the following:

$$\begin{aligned} \lambda &= \frac{\max L_o}{\max L} = \frac{\left| \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)' \right|^{\frac{n}{2}}}{\left| \sum_{ij} \left(X_{ij} - \frac{X_{..}}{n} \right) \left(X_{ij} - \frac{X_{..}}{n} \right)' \right|^{\frac{n}{2}}} \\ &= \frac{|V|^{\frac{n}{2}}}{|U + V|^{\frac{n}{2}}} \end{aligned} \quad (13.3.5)$$

where

$$U = \sum_{ij} \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n} \right) \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n} \right)', \quad V = \sum_{ij} \left(X_{ij} - \frac{X_{i.}}{n_i} \right) \left(X_{ij} - \frac{X_{i.}}{n_i} \right)'$$

and $U \sim W_p(k - 1, \Sigma)$ under H_o is the sum of squares and cross products matrix due to the A_i 's and $V \sim W_p(n. - k, \Sigma)$ is the residual sum of squares and cross products matrix. It has already been shown that U and V are independently distributed. Then $W_1 = (U + V)^{-\frac{1}{2}} V (U + V)^{-\frac{1}{2}}$, with the determinant $\frac{|V|}{|U+V|}$, is a real matrix-variate type-1 beta with parameters $(\frac{n.-k}{2}, \frac{k-1}{2})$, as defined in Chap. 5, and $W_2 = V^{-\frac{1}{2}} U V^{-\frac{1}{2}}$ is a real matrix-variate type-2 beta with the parameters $(\frac{k-1}{2}, \frac{n.-k}{2})$. Moreover, $Y_1 = I - W_1 = (U + V)^{-\frac{1}{2}} U (U + V)^{-\frac{1}{2}}$ with $\frac{|U|}{|U+V|}$ is a real matrix-variate type-1 beta random variables with parameters $(\frac{k-1}{2}, \frac{n.-k}{2})$. Given the properties of independent real matrix-variate gamma random variables, we have seen in Chap. 5 that W_1 and $Y_2 = U + V$ are independently distributed. Similarly, $Y_1 = I - W_1$ and Y_2 are independently distributed. Further, $W_2^{-1} = V^{\frac{1}{2}} U^{-1} V^{\frac{1}{2}}$ is a real matrix-variate type-2 beta random variable with the parameters $(\frac{n.-k}{2}, \frac{k-1}{2})$. Observe that

$$|W_1| = \frac{|V|}{|U + V|} = \frac{1}{|V^{-\frac{1}{2}} U V^{-\frac{1}{2}} + I|} = \frac{1}{|W_2 + I|}, \quad W_1 = (I + W_2)^{-1}.$$

A one-to-one function of λ is

$$w = \lambda^{\frac{2}{n.}} = \frac{|V|}{|U + V|} = |W_1|. \tag{13.3.6}$$

13.3.1. Arbitrary moments of the likelihood ratio criterion

For an arbitrary h , the h -th moment of w as well as that of λ can be obtained from the normalizing constant of a real matrix-variate type-1 beta density with the parameters $(\frac{n.-k}{2}, \frac{k-1}{2})$. That is,

$$\begin{aligned} E[w^h] &= \frac{\Gamma_p(\frac{n.-k}{2} + h)}{\Gamma_p(\frac{n.-k}{2})} \frac{\Gamma_p(\frac{n.-1}{2})}{\Gamma_p(\frac{n.-1}{2} + h)}, \quad (n. - k) + (k - 1) = n. - 1, \\ &= \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{n.-1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n.-k}{2} - \frac{j-1}{2})} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{n.-k}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{n.-1}{2} - \frac{j-1}{2} + h)} \right\}. \end{aligned} \tag{13.3.7}$$

As $E[\lambda^h] = E[w^{\frac{n}{2}}]^h = E[w^{(\frac{n}{2})h}]$, the h -th moment of λ is obtained by replacing h by $(\frac{n}{2})h$ in (13.3.7). That is,

$$E[\lambda^h] = C_{p,k} \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{n-k}{2} - \frac{j-1}{2} + (\frac{n}{2})h)}{\Gamma(\frac{n-1}{2} - \frac{j-1}{2} + (\frac{n}{2})h)} \right\} \quad (13.3.8)$$

where

$$C_{p,k} = \left\{ \prod_{j=1}^p \frac{\Gamma(\frac{n-1}{2} - \frac{j-1}{2})}{\Gamma(\frac{n-k}{2} - \frac{j-1}{2})} \right\}.$$

13.3.2. Structural representation of the likelihood ratio criterion

It can readily be seen from (13.3.7) that the h -th moment of w is of the form of the h -th moment of a product of independently distributed real scalar type-1 beta random variables. That is,

$$E[w^h] = E[w_1 w_2 \cdots w_p]^h, \quad w = w_1 w_2 \cdots w_p, \quad (13.3.9)$$

where w_1, \dots, w_p are independently distributed and w_j is a real scalar type-1 beta random variable with the parameters $(\frac{n-k}{2} - \frac{j-1}{2}, \frac{k-1}{2})$, $j = 1, \dots, p$, for $n - k > p - 1$ and $n > k + p - 1$. Hence the exact density of w is available by constructing the density of a product of independently distributed real scalar type-1 beta random variables. For special values of p and k , one can obtain the exact densities in the forms of elementary functions. However, for the general case, the exact density corresponding to $E[w^h]$ as specified in (13.3.7) can be expressed in terms of a G-function and, in the case of $E[\lambda^h]$ as given in (13.3.8), the exact density can be represented in terms of an H-function. These representations are as follows, denoting the densities of w and λ as $f_w(w)$ and $f_\lambda(\lambda)$, respectively:

$$f_w(w) = C_{p,k} G_{p,p}^{p,0} \left[w \left| \begin{array}{l} \frac{n-1}{2} - \frac{j-1}{2} - 1, j=1, \dots, p \\ \frac{n-k}{2} - \frac{j-1}{2} - 1, j=1, \dots, p \end{array} \right. \right], \quad 0 < w \leq 1, \quad (13.3.10)$$

$$f_\lambda(\lambda) = C_{p,k} H_{p,p}^{p,0} \left[\lambda \left| \begin{array}{l} (\frac{n-1}{2} - \frac{j-1}{2} - \frac{n}{2}, \frac{n}{2}), j=1, \dots, p \\ (\frac{n-k}{2} - \frac{j-1}{2} - \frac{n}{2}, \frac{n}{2}), j=1, \dots, p \end{array} \right. \right], \quad 0 < \lambda \leq 1, \quad (13.3.11)$$

for $n > p + k - 1$, $p \geq 1$ and $f_w(w) = 0$, $f_\lambda(\lambda) = 0$, elsewhere. The evaluation of G and H-functions can be carried out with the help of symbolic computing packages such as *Mathematica* and MAPLE. Theoretical considerations, applications and several special cases of the G and H-functions are, for instance, available from Mathai (1993) and Mathai,

Saxena and Haubold (2010). The special cases listed therein can also be utilized to work out the densities for particular cases of (13.3.10) and (13.3.11). Explicit structures of the densities for certain special cases are listed in the next section.

13.3.3. Some special cases

Several particular cases can be worked out by examining the moment expressions in (13.3.7) and (13.3.8). The h -th moment of the $w = \lambda^{\frac{2}{n}}$, where λ is the likelihood ratio criterion, is available from (13.3.7) as

$$E[w^h] = C_{p,k} \frac{\Gamma(\frac{n-k}{2} + h)\Gamma(\frac{n-k}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n-k}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{n-1}{2} + h)\Gamma(\frac{n-1}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n-1}{2} - \frac{p-1}{2} + h)}. \tag{i}$$

Case (1): $p = 1$

In this case, from (i),

$$E[w^h] = C_{1,k} \frac{\Gamma(\frac{n-k}{2} + h)}{\Gamma(\frac{n-1}{2} + h)},$$

which is the h -th moment of a real scalar type-1 beta random variable with the parameters $(\frac{n-k}{2}, \frac{k-1}{2})$ and, in this case, w is simply a real scalar type-1 beta random variable with the parameters $(\frac{n-k}{2}, \frac{k-1}{2})$. We reject the null hypothesis $H_o : A_1 = A_2 = \cdots = A_k = O$ for small values of the λ -criterion and, accordingly, we reject H_o for small values of w or the hypothesis is rejected when the observed value of $w \leq w_\alpha$ where w_α is such that $\int_0^{w_\alpha} f_w(w)dw = \alpha$ for the preassigned size α of the critical region, $f_w(w)$ denoting the density of w for $p = 1, n. > k$.

Case (2): $p = 2$

From (i), we have

$$E[w^h] = C_{2,k} \frac{\Gamma(\frac{n-k}{2} + h)\Gamma(\frac{n-k}{2} - \frac{1}{2} + h)}{\Gamma(\frac{n-1}{2} + h)\Gamma(\frac{n-1}{2} - \frac{1}{2} + h)}$$

and therefore

$$E[w^{\frac{1}{2}}]^h = C_{2,k} \frac{\Gamma(\frac{n-k}{2} + \frac{h}{2})\Gamma(\frac{n-k}{2} - \frac{1}{2} + \frac{h}{2})}{\Gamma(\frac{n-1}{2} + \frac{h}{2})\Gamma(\frac{n-1}{2} - \frac{1}{2} + \frac{h}{2})}. \tag{ii}$$

The gamma functions in (ii) can be combined by making use of a duplication formula for gamma functions, namely,

$$\Gamma(z)\Gamma(z + 1/2) = \pi^{\frac{1}{2}}2^{1-2z}\Gamma(2z). \tag{13.3.12}$$

Take $z = \frac{n-k}{2} - \frac{1}{2} + \frac{h}{2}$ and $z = \frac{n-1}{2} - \frac{1}{2} + \frac{h}{2}$ in the part containing h and in the constant part wherein $h = 0$, and then apply formula (13.3.12) to obtain

$$E[w^{\frac{1}{2}}]^h = \frac{\Gamma(n-2) \Gamma(n-k-1+h)}{\Gamma(n-k-1) \Gamma(n-2+h)},$$

which is, for an arbitrary h , the h -th moment of a real scalar type-1 beta random variable with parameters $(n-k-1, k-1)$ for $n-k-1 > 0$, $k > 1$. Thus, $y = w^{\frac{1}{2}}$ is a real scalar type-1 beta random variable with the parameters $(n-k-1, k-1)$. We would then reject H_o for small values of w , that is, for small values of y or when the observed value of $y \leq y_\alpha$ with y_α such that $\int_0^{y_\alpha} f_y(y)dy = \alpha$ for a preassigned probability of type-I error which is the error of rejecting H_o when H_o is true, where $f_y(y)$ is the density of y for $p = 2$ whenever $n > k + 1$.

Case (3): $k = 2$, $p \geq 1$, $n > p + 1$

In this case, the h -th moment of w as specified in (13.3.7) is the following:

$$\begin{aligned} E[w^h] &= C_{p,2} \frac{\Gamma(\frac{n-2}{2} + h) \Gamma(\frac{n-2}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n-2}{2} - \frac{p-1}{2} + h)}{\Gamma(\frac{n-1}{2} + h) \Gamma(\frac{n-1}{2} - \frac{1}{2} + h) \cdots \Gamma(\frac{n-1}{2} - \frac{p-1}{2} + h)} \\ &= C_{p,2} \frac{\Gamma(\frac{n-1}{2} - \frac{p}{2} + h)}{\Gamma(\frac{n-1}{2} + h)} \end{aligned}$$

since the numerator gamma functions, except the last one, cancel with the denominator gamma functions except the first one. This expression happens to be the h -th moment of a real scalar type-1 beta random variable with the parameters $(\frac{n-1-p}{2}, \frac{p}{2})$ and hence, for $k = 2$, $n-1-p > 0$ and $p \geq 1$, w is a real scalar type-1 beta random variable. Then, we reject the null hypothesis H_o for small values of w or when the observed value of $w \leq w_\alpha$, with w_α such that $\int_0^{w_\alpha} f_w(w)dw = \alpha$ for a preassigned significance level α , $f_w(w)$ being the density of w for this case. We will use the same notation $f_w(w)$ for the density of w in all the special cases.

Case (4): $k = 3$, $p \geq 1$

Proceeding as in Case (3), we see that all the gammas in the h -th moment of w cancel out except the last two in the numerator and the first two in the denominator. Thus,

$$\begin{aligned} E[w^h] &= C_{p,3} \frac{\Gamma(\frac{n-3}{2} + \frac{1}{2} - \frac{p-1}{2} + h) \Gamma(\frac{n-3}{2} - \frac{p-1}{2})}{\Gamma(\frac{n-1}{2} + h) \Gamma(\frac{n-1}{2} - \frac{1}{2} + h)} \\ &= C_{p,3} \frac{\Gamma(\frac{n-1}{2} - \frac{p}{2} + h) \Gamma(\frac{n-1}{2} - \frac{p}{2} - \frac{1}{2} + h)}{\Gamma(\frac{n-1}{2} + h) \Gamma(\frac{n-1}{2} - \frac{1}{2} + h)}. \end{aligned}$$

After combining the gammas in $y = w^{\frac{1}{2}}$ with the help of the duplication formula (13.3.12), we have the following:

$$E[y^h] = \frac{\Gamma(n. - 2)}{\Gamma(n. - 2 - p)} \frac{\Gamma(n. - p - 2 + h)}{\Gamma(n. - 2 + h)}.$$

Therefore, $y = w^{\frac{1}{2}}$ is a real scalar type-1 random variable with the parameters $(n. - p - 2, p)$. We reject the null hypothesis for small values of y or when the observed value of $y \leq y_\alpha$, with y_α such that $\int_0^{y_\alpha} f_y(y)dy = \alpha$ for a preassigned significance level α . We will use the same notation $f_y(y)$ for the density of y in all special cases.

We can also obtain some special cases for $t_1 = \frac{1-w}{w}$ and $t_2 = \frac{1-y}{y}$, with $y = \sqrt{w}$. With this transformation, t_1 and t_2 will be available in terms of type-2 beta variables in the real scalar case, which conveniently enables us to relate this distribution to real scalar F random variables so that an F table can be used for testing the null hypothesis and reaching a decision. We have noted that

$$\begin{aligned} w &= \frac{|V|}{|U + V|} = |(U + V)^{-\frac{1}{2}}V(U + V)^{-\frac{1}{2}}| = |W_1| \\ &= \frac{1}{|V^{-\frac{1}{2}}UV^{-\frac{1}{2}} + I|} = \frac{1}{|W_2 + I|} \end{aligned}$$

where W_1 is a real matrix-variate type-1 beta random variable with the parameters $(\frac{n.-k}{2}, \frac{k-1}{2})$ and W_2 is a real matrix-variate type-2 beta random variable with the parameters $(\frac{k-1}{2}, \frac{n.-k}{2})$. Then, when $p = 1$, W_1 and W_2 are real scalar variables, denoted by w_1 and w_2 , respectively. Then for $p = 1$, we have one gamma ratio with h in the general h -th moment (13.3.7) and then,

$$t_1 = \frac{1 - w}{w} = \frac{1}{w} - 1 = (w_2 + 1) - 1 = w_2$$

where w_2 is a real scalar type-2 beta random variable with the parameters $(\frac{p-1}{2}, \frac{n.-k}{2})$. As well, in general, for a real matrix-variate type-2 beta matrix W_2 with the parameters $(\frac{\nu_1}{2}, \frac{\nu_2}{2})$, we have $\frac{\nu_2}{\nu_1}W_2 = F_{\nu_1, \nu_2}$ where F_{ν_1, ν_2} is a real matrix-variate F matrix random variable with degrees of freedom ν_1 and ν_2 . When $p = 1$ or in the real scalar case $\frac{\nu_2}{\nu_1}w_2 = F_{\nu_1, \nu_2}$ where, in this case, F is a real scalar F random variable with ν_1 and ν_2 degrees of freedom. We have used F for the scalar and matrix-variate case in order to avoid too many symbols. For $p = 2$, we combine the gamma functions in the numerator and denominator by applying the duplication formula for gamma functions (13.3.12); then, for $t_2 = \frac{1-y}{y}$ the situation turns out to be the same as in the case of t_1 , the only difference being that in the

real scalar type-2 beta w_2 , the parameters are $(k - 1, n. - k - 1)$. Note that the original $\frac{k-1}{2}$ has become $k - 1$ and the original $\frac{n.-k}{2}$ has become $n. - k - 1$. Thus, we can state the following two special cases.

Case (5): $p = 1, t_1 = \frac{1-w}{w}$

As was explained, t_1 is a real type-2 beta random variable with the parameters $(\frac{k-1}{2}, \frac{n.-k}{2})$, so that

$$\frac{n. - k}{k - 1} t_1 \simeq F_{k-1, n.-k},$$

which is a real scalar F random variable with $k - 1$ and $n. - k$ degrees of freedom. Accordingly, we reject H_o for small values of w and y , which corresponds to large values of F . Thus, we reject the null hypothesis H_o whenever the observed value of $F_{k-1, n.-k} \geq F_{k-1, n.-k, \alpha}$ where $F_{k-1, n.-k, \alpha}$ is the upper 100 $\alpha\%$ percentage point of the F distribution or $\int_a^\infty g(F)dF = \alpha$ where $a = F_{k-1, n.-k, \alpha}$ and $g(F)$ is the density of F in this case.

Case (6): $p = 2, t_2 = \frac{1-y}{y}, y = \sqrt{w}$

As previously explained, t_2 is a real scalar type-2 beta random variable with the parameters $(k - 1, n. - k - 1)$ or

$$\frac{n. - k - 1}{k - 1} t_2 \simeq F_{2(k-1), 2(n.-k-1)},$$

which is a real scalar F random variable having $2(k - 1)$ and $2(n. - k - 1)$ degrees of freedom. We reject the null hypothesis for large values of t_2 or when the observed value of $[\frac{n.-k-1}{k-1}]t_2 \geq b$ with b such that $\int_b^\infty g(F)dF = \alpha$, $g(F)$ denoting in this case the density of a real scalar random variable F with degrees of freedoms $2(k - 1)$ and $2(n. - k - 1)$, and $b = F_{2(k-1), 2(n.-k-1), \alpha}$.

Case (7): $k = 2, p \geq 1, t_1 = \frac{1-w}{w}$

For the case $k = 2$, we have already seen that the gamma functions with h in their arguments cancel out, leaving only one gamma in the numerator and one gamma in the denominator, so that w is distributed as a real scalar type-1 beta random variable with the parameters $(\frac{n.-1-p}{2}, \frac{p}{2})$. Thus, $t_1 = \frac{1-w}{w}$ is a real scalar type-2 beta with the parameters $(\frac{p}{2}, \frac{n.-p-1}{2})$, and

$$\left[\frac{n. - 1 - p}{p} \right] t_1 \simeq F_{p, n.-1-p},$$

which is a real scalar F random variable having p and $n. - 1 - p$ degrees of freedom. We reject H_o for large values of t_1 or when the observed value of $[\frac{n.-1-p}{p}]t_1 \geq b$ where b

is such that $\int_b^\infty g(F)dF = \alpha$ with $g(F)$ being the density of an F random variable with degrees of freedoms p and $n. - 1 - p$ in this special case.

Case (8): $k = 3, p \geq 1, t_2 = \frac{1-y}{y}, y = \sqrt{w}$

On combining Cases (4) and (6), it is seen that t_2 is a real scalar type-2 beta random variable with the parameters $(p, n. - p - 1)$, so that

$$\frac{n. - 1 - p}{p} t_2 \simeq F_{2p, 2(n. - p - 1)},$$

which is a real scalar F random variable with the degrees of freedoms $(2p, 2(n. - p - 1))$. Thus, we reject the hypothesis for large values of this F random variable. For a test at significance level α or with α as the size of its critical region, the hypothesis $H_o : A_1 = A_2 = \dots = A_k = O$ is rejected when the observed value of this $F \geq F_{2p, 2(n. - 1 - p), \alpha}$ where $F_{2p, 2(n. - 1 - p), \alpha}$ is the upper 100 $\alpha\%$ percentage point of the F distribution.

Example 13.3.1. In a dieting experiment, three different diets D_1, D_2 and D_3 are tried for a period of one month. The variables monitored are weight in kilograms (kg), waist circumference in centimeters (cm) and right mid-thigh circumference in centimeters. The measurements are $x_1 =$ final weight minus initial weight, $x_2 =$ final waist circumference minus initial waist reading and $x_3 =$ final minus initial thigh circumference. Diet D_1 is administered to a group of 5 randomly selected individuals ($n_1 = 5$), D_2 , to 4 randomly selected persons ($n_2 = 4$), and 6 randomly selected individuals ($n_3 = 6$) are subjected to D_3 . Since three variables are monitored, $p = 3$. As well, there are three treatments or three diets, so that $k = 3$. In our notation,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, X_{ij} = \begin{bmatrix} x_{1ij} \\ x_{2ij} \\ x_{3ij} \end{bmatrix}, X_{1j} = \begin{bmatrix} x_{11j} \\ x_{21j} \\ x_{31j} \end{bmatrix}, X_{2j} = \begin{bmatrix} x_{12j} \\ x_{22j} \\ x_{32j} \end{bmatrix}, X_{3j} = \begin{bmatrix} x_{13j} \\ x_{23j} \\ x_{33j} \end{bmatrix},$$

where i corresponds to the diet number and j stands for the sample serial number. For example, the observation vector on individual #3 within the group subjected to diet D_2 is denoted by X_{23} . The following are the data on x_1, x_2, x_3 :

Diet D_1 : $X_{1j}, j = 1, 2, 3, 4, 5 :$

$$X_{11} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, X_{12} = \begin{bmatrix} 4 \\ -2 \\ -1 \end{bmatrix}, X_{13} = \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}, X_{14} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, X_{15} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Diet D_2 : X_{2j} , $j = 1, 2, 3, 4$:

$$X_{21} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad X_{22} = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}, \quad X_{23} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad X_{24} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Diet D_3 : X_{3j} , $j = 1, 2, 3, 4, 5, 6$:

$$X_{31} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \quad X_{32} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \quad X_{33} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \\ X_{34} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \quad X_{35} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad X_{36} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

- (1): Perform an ANOVA test on the first component consisting of weight measurements;
 (2): Carry out a MANOVA test on the first two components, weight and waist measurements;
 (3): Do a MANOVA test on all the three variables, weight, waist and thigh measurements.

Solution 13.3.1. We first compute the vectors $X_{1.}$, \bar{X}_1 , $X_{2.}$, \bar{X}_2 , $X_{3.}$, \bar{X}_3 , $X_{..}$ and \bar{X} :

$$X_{1.} = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \quad \bar{X}_1 = \frac{X_{1.}}{n_1} = \frac{1}{5} \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad X_{2.} = \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix}, \quad \bar{X}_2 = \frac{X_{2.}}{n_2} = \frac{1}{4} \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \\ X_{3.} = \begin{bmatrix} 6 \\ 12 \\ 6 \end{bmatrix}, \quad \bar{X}_3 = \frac{X_{3.}}{6} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad X_{..} = \begin{bmatrix} 15 \\ 16 \\ 6 \end{bmatrix}, \quad \bar{X} = \frac{X_{..}}{n} = \frac{1}{15} \begin{bmatrix} 15 \\ 16 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 16/15 \\ 6/15 \end{bmatrix}.$$

Problem (1): ANOVA on the first component x_1 . The first components of the observations are x_{1ij} . The first components under diet D_1 are

$$[x_{111}, x_{112}, x_{113}, x_{114}, x_{115}] = [2, 4, -1, -1, 1] \text{ with } x_{11.} = 5;$$

the first components of observations under diet D_2 are

$$[x_{121}, x_{122}, x_{123}, x_{124}] = [1, 3, -1, 1] \text{ with } x_{12.} = 4;$$

and the first components under diet D_3 are

$$[x_{131}, x_{132}, x_{133}, x_{134}, x_{135}, x_{136}] = [2, 1, -1, 2, 2, 0] \text{ with } x_{13.} = 6.$$

Hence, the total on the first component $x_{1..} = 15$, and $\bar{x}_1 = \frac{x_{1..}}{n} = \frac{15}{15} = 1$. The first component model is the following:

$$x_{1ij} = \mu + \alpha_i + e_{1ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k.$$

Note again that estimators and estimates will be denoted by a hat. As previously mentioned, the same symbols will be used for the variables and the observations on those variables in order to avoid using too many symbols; however, the notations will be clear from the context. If the discussion pertains to distributions, then variables are involved, and if we are referring to numbers, then we are dealing with observations.

The least squares estimates are $\hat{\mu} = \frac{x_{1..}}{n} = 1$, $\hat{\alpha}_1 = \frac{x_{11}}{5} = \frac{5}{5} = 1$, $\hat{\alpha}_2 = \frac{x_{12}}{4} = \frac{4}{4} = 1$, $\hat{\alpha}_3 = \frac{x_{13}}{6} = \frac{6}{6} = 1$. The first component hypothesis is $\alpha_1 = \alpha_2 = \alpha_3 = 0$. The total sum of squares is

$$\begin{aligned} \sum_{ij} (x_{1ij} - \bar{x}_1)^2 &= \sum_{ij} x_{1ij}^2 - \frac{x_{1..}^2}{n} \\ &= (2 - 1)^2 + (4 - 1)^2 + (-1 - 1)^2 + (-1 - 1)^2 + (1 - 1)^2 \\ &\quad + (1 - 1)^2 + (3 - 1)^2 + (-1 - 1)^2 + (1 - 1)^2 \\ &\quad + (2 - 1)^2 + (1 - 1)^2 + (-1 - 1)^2 + (2 - 1)^2 + (2 - 1)^2 + (0 - 1)^2 \\ &= 34. \end{aligned}$$

The sum of squares due to the α_i 's is available from

$$\begin{aligned} \sum_i n_i \left(\frac{x_{1i.}}{n_i} - \frac{x_{1..}}{n} \right)^2 &= \sum_i \frac{x_{1i.}^2}{n_i} - \frac{x_{1..}^2}{n} \\ &= 5 \left(\frac{5}{5} - \frac{15}{15} \right)^2 + 4 \left(\frac{4}{4} - \frac{15}{15} \right)^2 + 6 \left(\frac{6}{6} - \frac{15}{15} \right)^2 = 0. \end{aligned}$$

Hence the following table:

ANOVA Table

Variation due to	<i>df</i>	<i>SS</i>	<i>MS</i>	F-ratio
diets	2	0	0	0
residuals	12	34		
total	14	34		

Since the sum of squares due to the α_i 's is null, the hypothesis is not rejected at any level.

Problem (2): MANOVA on the first two components. We are still using the notation X_{ij} for the two and three-component cases since our general notation does not depend on the number of components in the vector concerned; as well, we can make use of the computations pertaining to the first component in Problem (1). The relevant quantities computed from the data on the first two components are the following:

$$\text{Diet } D_1: \begin{bmatrix} 2 & 4 & -1 & -1 & 1 \\ 3 & -2 & -2 & 1 & 0 \end{bmatrix}, X_{1.} = \begin{bmatrix} x_{11.} \\ x_{21.} \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \bar{X}_1 = \frac{1}{5} \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$

$$\text{Diet } D_2: \begin{bmatrix} 1 & 3 & -1 & 1 \\ 2 & -1 & 2 & 1 \end{bmatrix}, X_{2.} = \begin{bmatrix} x_{12.} \\ x_{22.} \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \bar{X}_2 = \frac{1}{4} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

$$\text{Diet } D_3: \begin{bmatrix} 2 & 1 & -1 & 2 & 2 & 0 \\ 2 & 3 & 2 & 4 & 0 & 1 \end{bmatrix}, X_{3.} = \begin{bmatrix} 6 \\ 12 \end{bmatrix}, \bar{X}_3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

In this case, the grand total, denoted by $X_{..}$, and the grand average, denoted by \bar{X} , are the following:

$$X_{..} = \begin{bmatrix} 15 \\ 16 \end{bmatrix}, \bar{X} = \frac{1}{15} \begin{bmatrix} 15 \\ 16 \end{bmatrix} = \begin{bmatrix} 1 \\ 16/15 \end{bmatrix}.$$

Note that the total sum of squares and cross products matrix can be written as follows:

$$\begin{aligned} \sum_{ij} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' &= \sum_{j=1}^5 (X_{1j} - \bar{X})(X_{1j} - \bar{X})' + \sum_{j=1}^4 (X_{2j} - \bar{X})(X_{2j} - \bar{X})' \\ &\quad + \sum_{j=1}^6 (X_{3j} - \bar{X})(X_{3j} - \bar{X})'. \end{aligned}$$

Then,

$$\begin{aligned} \sum_{j=1}^5 (X_{1j} - \bar{X})(X_{1j} - \bar{X})' &= \begin{bmatrix} 1 & 29/15 \\ 29/15 & 29^2/15^2 \end{bmatrix} + \begin{bmatrix} 9 & -138/15 \\ -138/15 & 46^2/15^2 \end{bmatrix} \\ &\quad + \begin{bmatrix} 4 & 92/15 \\ 92/15 & 46^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & 2/15 \\ 2/15 & 1/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 16^2/15^2 \end{bmatrix} \\ &= \begin{bmatrix} 18 & -1 \\ -1 & 5330/15^2 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^4 (X_{2j} - \bar{X})(X_{2j} - \bar{X})' &= \begin{bmatrix} 0 & 0 \\ 0 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -62/15 \\ -62/15 & 31^2/15^2 \end{bmatrix} \\ &\quad + \begin{bmatrix} 4 & -28/15 \\ -28/15 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1/15^2 \end{bmatrix} = \begin{bmatrix} 8 & -6 \\ -6 & 1354/15^2 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^6 (X_{3j} - \bar{X})(X_{3j} - \bar{X})' &= \begin{bmatrix} 1 & 14/15 \\ 14/15 & 14^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 29^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -28/15 \\ -28/15 & 14^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 44/15 \\ 44/15 & 44^2/15^2 \end{bmatrix} + \begin{bmatrix} 1 & -16/15 \\ -16/15 & 16^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 1/15 \\ 1/15 & 1/15^2 \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 1 & 3426/15^2 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{ij} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' &= \begin{bmatrix} 18 & -1 \\ -1 & 5330/15^2 \end{bmatrix} + \begin{bmatrix} 8 & -6 \\ -6 & 1354/15^2 \end{bmatrix} + \begin{bmatrix} 8 & 1 \\ 1 & 3426/15^2 \end{bmatrix} \\ &= \begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} \text{ and } \begin{vmatrix} 34 & -6 \\ -6 & 674/15 \end{vmatrix} = 1491.73. \end{aligned}$$

Now, consider the residual sum of squares and cross products matrix:

$$\begin{aligned} \sum_{ij} (X_{ij} - \frac{X_{i.}}{n_i})(X_{ij} - \frac{X_{i.}}{n_i})' &= \sum_{j=1}^5 (X_{1j} - \frac{X_{1.}}{n_1})(X_{1j} - \frac{X_{1.}}{n_1})' \\ &+ \sum_{j=1}^4 (X_{2j} - \frac{X_{2.}}{n_2})(X_{2j} - \frac{X_{2.}}{n_2})' + \sum_{j=1}^6 (X_{3j} - \frac{X_{3.}}{n_3})(X_{3j} - \frac{X_{3.}}{n_3})'. \end{aligned}$$

That is,

$$\begin{aligned} \sum_{j=1}^5 (X_{1j} - \frac{X_{1.}}{n_1})(X_{1j} - \frac{X_{1.}}{n_1})' &= \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} + \begin{bmatrix} 9 & -6 \\ -6 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \\ &+ \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 18 & -1 \\ -1 & 18 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^4 (X_{2j} - \frac{X_{2.}}{n_2})(X_{2j} - \frac{X_{2.}}{n_2})' &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 8 & -6 \\ -6 & 6 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^6 (X_{3j} - \frac{X_{3.}}{n_3})(X_{3j} - \frac{X_{3.}}{n_3})' &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 1 & 10 \end{bmatrix}. \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{ij} (X_{ij} - \frac{X_{i.}}{n_i})(X_{ij} - \frac{X_{i.}}{n_i})' &= \begin{bmatrix} 18 & -1 \\ -1 & 18 \end{bmatrix} + \begin{bmatrix} 8 & -6 \\ -6 & 6 \end{bmatrix} + \begin{bmatrix} 8 & 1 \\ 1 & 10 \end{bmatrix} \\ &= \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix} \text{ and } \begin{vmatrix} 34 & -6 \\ -6 & 34 \end{vmatrix} = 1120. \end{aligned}$$

Therefore, the observed w is given by

$$w = \frac{1120}{1491.73} = 0.7508, \quad \sqrt{w} = 0.8665.$$

This is the case $p = 2$, $k = 3$, that is, our special Case (8). Then, the observed value of

$$t_2 = \frac{1 - \sqrt{w}}{\sqrt{w}} = \frac{0.1335}{0.8665} = 0.1540,$$

and

$$\frac{n. - 1 - p}{p} t_2 = \frac{15 - 1 - 2}{2} (0.1540) = 0.9244.$$

Our F-statistic is $F_{2p, 2(n.-p-1)} = F_{4, 24}$. Let us test the hypothesis $A_1 = A_2 = A_3 = O$ at the 5% significance level or $\alpha = 0.05$. Since the observed value $0.9244 < 5.77 = F_{4, 24, 0.05}$ which is available from F-tables, we do not reject the hypothesis.

Verification of the calculations

Denoting the total sum of squares and cross products matrix by S_t , the residual sum of squares and cross products matrix by S_r and the sum of squares and cross products matrix due to the hypothesis or due to the effects A_i 's by S_h , we should have $S_t = S_r + S_h$ where

$$S_t = \begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} \text{ and } S_r = \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix}$$

as previously determined. Let us compute

$$S_h = \sum_{i=1}^k n_i \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n.} \right) \left(\frac{X_{i.}}{n_i} - \frac{X_{..}}{n.} \right)'$$

For the first two components, we already have the following:

$$\begin{aligned} \frac{X_{1.}}{n_1} - \frac{X_{..}}{n.} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ -16/15 \end{bmatrix}, \quad n_1 = 5 \\ \frac{X_{2.}}{n_2} - \frac{X_{..}}{n.} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/15 \end{bmatrix}, \quad n_2 = 4 \\ \frac{X_{3.}}{n_3} - \frac{X_{..}}{n.} &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 16/15 \end{bmatrix} = \begin{bmatrix} 0 \\ 14/15 \end{bmatrix}, \quad n_3 = 6. \end{aligned}$$

Hence,

$$\begin{aligned} S_h &= 5 \begin{bmatrix} 0 & 0 \\ 0 & 16^2/15^2 \end{bmatrix} + 4 \begin{bmatrix} 0 & 0 \\ 0 & 1/15^2 \end{bmatrix} + 6 \begin{bmatrix} 0 & 0 \\ 0 & 14^2/15^2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 2460/15^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 164/15 \end{bmatrix}. \end{aligned}$$

As $34 + \frac{164}{15} = \frac{674}{15}$, $S_t = S_r + S_h$, that is,

$$\begin{bmatrix} 34 & -6 \\ -6 & 674/15 \end{bmatrix} = \begin{bmatrix} 34 & -6 \\ -6 & 34 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 164/15 \end{bmatrix}.$$

Thus, the result is verified.

Problem (3): Data on all the three variables. In this case, we have $p = 3, k = 3$. We will first use $X_1, \bar{X}_1, X_2, \bar{X}_2, X_3, \bar{X}_3, X_{..}$ and \bar{X} which have already been evaluated, to compute the residual sum of squares and cross product matrix. Since all the matrices are symmetric, for convenience, we will only display the diagonal elements and those above the diagonal. As in the case of two components, we compute the following, making use of the calculations already done for the 2-component case (the notations remaining the same since our general notation does not involve p):

$$\begin{aligned}
\sum_{j=1}^5 (X_{1j} - \frac{X_{1.}}{n_1})(X_{1j} - \frac{X_{1.}}{n_1})' &= \begin{bmatrix} 1 & 3 & 1 \\ 9 & 3 & \\ & & 1 \end{bmatrix} + \begin{bmatrix} 9 & -6 & -3 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 & -2 \\ & 4 & -2 \\ & & 1 \end{bmatrix} \\
&+ \begin{bmatrix} 4 & -2 & 2 \\ & 1 & -1 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix} = \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix}; \\
\sum_{j=1}^4 (X_{2j} - \frac{X_{2.}}{n_2})(X_{2j} - \frac{X_{2.}}{n_2})' &= \begin{bmatrix} 0 & 0 & 0 \\ & 1 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 4 & -4 & -4 \\ & 4 & 4 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 4 & -2 & -2 \\ & 1 & 1 \\ & & 1 \end{bmatrix} \\
&+ \begin{bmatrix} 0 & 0 & 0 \\ & 0 & 0 \\ & & 1 \end{bmatrix} = \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix}; \\
\sum_{j=1}^6 (X_{3j} - \frac{X_{3.}}{n_3})(X_{3j} - \frac{X_{3.}}{n_3})' &= \begin{bmatrix} 1 & 0 & 2 \\ & 0 & 0 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 1 & 0 \\ & & 0 \end{bmatrix} + \begin{bmatrix} 4 & 0 & -2 \\ & 0 & 0 \\ & & 1 \end{bmatrix} \\
&+ \begin{bmatrix} 1 & 2 & 1 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 1 & -2 & -1 \\ & 4 & 2 \\ & & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & -1 \\ & 1 & -1 \\ & & 1 \end{bmatrix} \\
&= \begin{bmatrix} 8 & 1 & -5 \\ & 10 & 3 \\ & & 8 \end{bmatrix}.
\end{aligned}$$

Then,

$$\begin{aligned}
\sum_{ij} (X_{ij} - \frac{X_{i.}}{n_i})(X_{ij} - \frac{X_{i.}}{n_i})' &= \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix} + \begin{bmatrix} 8 & 1 & -5 \\ & 10 & 3 \\ & & 8 \end{bmatrix} \\
&= \begin{bmatrix} 34 & -6 & -13 \\ & 34 & 12 \\ & & 22 \end{bmatrix}
\end{aligned}$$

whose determinant is equal to

$$\begin{aligned}
&= 34 \begin{vmatrix} 34 & 12 \\ 12 & 22 \end{vmatrix} + 6 \begin{vmatrix} -6 & 12 \\ -13 & 22 \end{vmatrix} - 13 \begin{vmatrix} -6 & 34 \\ -13 & 12 \end{vmatrix} \\
&= 15870.
\end{aligned}$$

The total sum of squares and cross products matrix is the following:

$$\sum_{ij} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' = \sum_{j=1}^5 (X_{1j} - \bar{X})(X_{1j} - \bar{X})' + \sum_{j=1}^4 (X_{2j} - \bar{X})(X_{2j} - \bar{X})' + \sum_{j=1}^6 (X_{3j} - \bar{X})(X_{3j} - \bar{X})',$$

with

$$\begin{aligned} \sum_{j=1}^5 (X_{1j} - \bar{X})(X_{1j} - \bar{X})' &= \begin{bmatrix} 1 & 29/15 & 9/15 \\ & 29^2/15^2 & (29 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 9 & -(3 \times 46)/15 & -(3 \times 21)/15 \\ & 46^2/15^2 & (46 \times 21)/15^2 \\ & & 21^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & 92/15 & -18/15 \\ & 46^2/15^2 & -(46 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 4 & 2/15 & 42/15 \\ & 1/15^2 & 21/15^2 \\ & & 21^2/15^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ & 16^2/15^2 & 96/15^2 \\ & & 36/15^2 \end{bmatrix} \\ &= \begin{bmatrix} 18 & -1 & -2 \\ & 5330/15^2 & 930/15^2 \\ & & 1080/15^2 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^4 (X_{2j} - \bar{X})(X_{2j} - \bar{X})' &= \begin{bmatrix} 0 & 0 & 0 \\ & 14^2/15^2 & (14 \times 24)/15^2 \\ & & 24^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 4 & -62/15 & -72/15 \\ & 31^2/15^2 & (31 \times 36)/15^2 \\ & & 36^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -28/15 & -18/15 \\ & 14^2/15^2 & (14 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ & 1/15^2 & 21/15^2 \\ & & 21^2/15^2 \end{bmatrix} = \begin{bmatrix} 8 & -6 & -6 \\ & 1354/15^2 & 1599/15^2 \\ & & 2394/15^2 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^6 (X_{3j} - \bar{X})(X_{3j} - \bar{X})' &= \begin{bmatrix} 1 & 14/15 & -21/15 \\ & 14^2/15^2 & -(14 \times 21)/15^2 \\ & & 21^2/15^2 \end{bmatrix} \\
&+ \begin{bmatrix} 0 & 0 & 0 \\ & 29^2/15^2 & (29 \times 9)/15^2 \\ & & 9^2/15^2 \end{bmatrix} + \begin{bmatrix} 4 & -28/15 & -48/15 \\ & 14^2/15^2 & (14 \times 24)/15^2 \\ & & 24^2/15^2 \end{bmatrix} \\
&+ \begin{bmatrix} 1 & 44/15 & 24/15 \\ & 44^2/15^2 & (44 \times 24)/15^2 \\ & & 24^2/15^2 \end{bmatrix} + \begin{bmatrix} 1 & -16/15 & -6/15 \\ & 16^2/15^2 & 96/15^2 \\ & & 6^2/15^2 \end{bmatrix} \\
&+ \begin{bmatrix} 1 & 1/15 & -24/15 \\ & 1/15^2 & -24/15^2 \\ & & 24^2/15^2 \end{bmatrix} = \begin{bmatrix} 8 & 1 & -5 \\ & 3426/15^2 & 1431/15^2 \\ & & 2286/15^2 \end{bmatrix}.
\end{aligned}$$

Hence the total sum of squares and cross products matrix is

$$\begin{aligned}
\sum_{ij} (X_{ij} - \bar{X})(X_{ij} - \bar{X})' &= \begin{bmatrix} 18 & -1 & -2 \\ & 5330/15^2 & 930/15^2 \\ & & 1080/15^2 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 1354/15^2 & 1599/15^2 \\ & & 2394/15^2 \end{bmatrix} \\
&+ \begin{bmatrix} 8 & 1 & -5 \\ & 3426/15^2 & 1431/15^2 \\ & & 2286/15^2 \end{bmatrix} = \begin{bmatrix} 34 & -6 & -13 \\ & -6 & 674/15 & 264/15 \\ & -13 & 264/15 & 384/15 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\text{and its determinant} &= 34 \begin{vmatrix} 674/15 & 264/15 \\ 264/15 & 384/15 \end{vmatrix} + 6 \begin{vmatrix} -6 & 264/15 \\ -13 & 384/15 \end{vmatrix} \\
&- 13 \begin{vmatrix} -6 & 674/15 \\ -13 & 264/15 \end{vmatrix} = 342126/15.
\end{aligned}$$

Then, the observed value of

$$w = \frac{15870 \times 15}{342126} = 0.6958, \quad \sqrt{w} = 0.8341.$$

Since $p = 3$ and $k = 3$, an exact distribution is available from our special Case (8) for $t_2 = \frac{1-\sqrt{w}}{\sqrt{w}}$ and an observed value of $t_2 = 0.1989$. Then,

$$\frac{n. - 1 - p}{p} t_2 = \frac{15 - 1 - 3}{3} t_2 = \frac{11}{3} t_2 \sim F_{2p, 2(n. - 1 - p)} = F_{6, 22}.$$

The critical value obtained from an F-table at the 5% significance level is $F_{6,22,.05} \approx 3.85$. Since the observed value of $F_{6,22}$ is $\frac{11}{3}(0.1989) = 0.7293 < 3.85$, the hypothesis $A_1 = A_2 = A_3 = O$ is not rejected. It can also be verified that $S_t = S_r + S_h$.

13.3.4. Asymptotic distribution of the λ -criterion

We can obtain an asymptotic real chisquare distribution for $n. \rightarrow \infty$. To this end, consider the general h -th moments of λ or $E[\lambda^h]$ from (13.3.8), that is,

$$\begin{aligned} E[\lambda^h] &= C_{p,k} \prod_{j=1}^p \left[\Gamma\left(\frac{n. - k}{2} - \frac{j - 1}{2} + \frac{n. h}{2}\right) / \Gamma\left(\frac{n. - 1}{2} - \frac{j - 1}{2} + \frac{n. h}{2}\right) \right] \\ &= C_{p,k} \prod_{j=1}^p \left[\Gamma\left(\frac{n.}{2}(1 + h) - \frac{j - 1}{2} - \frac{k}{2}\right) / \Gamma\left(\frac{n.}{2}(1 + h) - \frac{j - 1}{2} - \frac{1}{2}\right) \right]. \end{aligned}$$

Let us expand all the gamma functions in $E[\lambda^h]$ by using the first term in the asymptotic expansion of a gamma function or by making use of Stirling’s approximation formula, namely,

$$\Gamma(z + \delta) \approx \sqrt{(2\pi)} z^{z+\delta-\frac{1}{2}} e^{-z} \tag{13.3.13}$$

for $|z| \rightarrow \infty$ when δ is a bounded quantity. Taking $\frac{n.}{2} \rightarrow \infty$ in the constant part and $\frac{n.}{2}(1 + h) \rightarrow \infty$ in the part containing h , we have

$$\begin{aligned} &\Gamma\left(\frac{n.}{2}(1 + h) - \frac{j - 1}{2} - \frac{k}{2}\right) / \Gamma\left(\frac{n.}{2}(1 + h) - \frac{j - 1}{2} - \frac{1}{2}\right) \\ &\approx \left\{ \sqrt{(2\pi)} \left[\frac{n.}{2}(1 + h)\right]^{\frac{n.}{2}(1+h) - \frac{j-1}{2} - \frac{k}{2} - \frac{1}{2}} e^{-\frac{n.}{2}(1+h)} \right. \\ &\quad \left. / \sqrt{(2\pi)} \left[\frac{n.}{2}(1 + h)\right]^{\frac{n.}{2}(1+h) - \frac{j-1}{2} - \frac{1}{2} - \frac{1}{2}} e^{-\frac{n.}{2}(1+h)} \right\} \\ &= \left(\frac{n.}{2}\right)^{-(\frac{k-1}{2})} (1 + h)^{-(\frac{k-1}{2})}. \end{aligned}$$

The factor $\left(\frac{n.}{2}\right)^{-(\frac{k-1}{2})}$ is canceled from the expression coming from the constant part. Then, taking the product over $j = 1, \dots, p$, we have

$$\lambda^h \rightarrow (1 + h)^{-p(k-1)/2} \text{ or } \lambda^{-2h} \rightarrow (1 - 2h)^{-p(k-1)/2} \text{ for } 1 - 2h > 0,$$

which is the moment generating function (mgf) of a real scalar chisquare with $p(k - 1)$ degrees of freedom. Hence, we have the following result:

Theorem 13.3.1. Letting λ be the likelihood ratio criterion for testing the hypothesis $H_o : A_1 = A_2 = \dots = A_k = O$, the asymptotic distribution of $-2 \ln \lambda$ is a real chisquare random variable having $p(k-1)$ degrees of freedom as $n. \rightarrow \infty$, that is,

$$-2 \ln \lambda \rightarrow \chi_{p(k-1)}^2 \text{ as } n. \rightarrow \infty. \quad (13.3.14)$$

Observe that we only require the sum of the sample sizes $n_1 + \dots + n_k = n.$ to go to infinity, and not that the individual n_j 's be large. This chisquare approximation can be utilized for testing the hypothesis for large values of $n.$, and we then reject H_o for small values of λ , which means for large values of $-2 \ln \lambda$ or large values of $\chi_{p(k-1)}^2$, that is, when the observed $-2 \ln \lambda \geq \chi_{p(k-1), \alpha}^2$ where $\chi_{p(k-1), \alpha}^2$ denotes the upper 100 $\alpha\%$ percentage point of the chisquare distribution.

13.3.5. MANOVA and testing the equality of population mean values

In a one-way classification model, we have the following for the p -variate case:

$$X_{ij} = M + A_i + E_{ij} \text{ or } X_{ij} = M_i + E_{ij}, \text{ with } M_i = M + A_i, \quad (13.3.15)$$

for $j = 1, \dots, n_i, i = 1, \dots, k$. When the error vector is assumed to have a null expected value, that is, $E[E_{ij}] = O$, for all i and j , we have $E[X_{ij}] = M_i$ for all i and j . Thus, this assumption, in conjunction with the hypothesis $A_1 = A_2 = \dots = A_k = O$, implies that $M_1 = M_2 = \dots = M_k$, that is, the hypothesis of equality of the population mean value vectors or the test is equivalent to testing the equality of population mean value vectors in k independent populations with common covariance matrix $\Sigma > O$. We have already tackled this problem in Chap. 6 under both assumptions that Σ is known and unknown, when the populations are Gaussian, that is, $X_{ij} \sim N_p(M_i, \Sigma), \Sigma > O$. Thus, the hypothesis made in a one-way classification MANOVA setting and the hypothesis of testing the equality of mean value vectors in MANOVA are one and the same. In the scalar case too, the ANOVA in a one-way classification data coincides with testing the equality of population mean values in k independent univariate populations. In the ANOVA case, we are comparing the sum of squares attributable to the hypothesis to the residual sum of squares. If the hypothesis really holds true, then the sum of squares due to the hypothesis or to the α_j 's (deviations from the general effect due to the j -th treatment) must be zero and hence for large values of the sum of squares due to the presence of the α_j 's, as compared to the residual sum of squares, we reject the hypothesis. In MANOVA, we are comparing two sums of squares and cross product matrices, namely,

$$U = \sum_{ij} \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n.} \right] \left[\frac{X_{i.}}{n_i} - \frac{X_{..}}{n.} \right]' \text{ and } V = \sum_{ij} \left[X_{ij} - \frac{X_{i.}}{n_i} \right] \left[X_{ij} - \frac{X_{i.}}{n_i} \right]'$$

We have the following distributional properties:

$$\begin{aligned}
 T_1 &= (U + V)^{-\frac{1}{2}}U(U + V)^{-\frac{1}{2}} \sim \text{real matrix-variate type-1 beta with parameters} \\
 &\quad \left(\frac{k-1}{2}, \frac{n-k}{2}\right); \\
 T_2 &= (U + V)^{-\frac{1}{2}}V(U + V)^{-\frac{1}{2}} \sim \text{real matrix-variate type-1 beta with parameters} \\
 &\quad \left(\frac{n-k}{2}, \frac{k-1}{2}\right); \\
 T_3 &= V^{-\frac{1}{2}}UV^{-\frac{1}{2}} \sim \text{real matrix-variate type-2 beta with parameters } \left(\frac{k-1}{2}, \frac{n-k}{2}\right); \\
 T_4 &= U^{-\frac{1}{2}}VU^{-\frac{1}{2}} \sim \text{real matrix-variate type-2 beta with parameters } \left(\frac{n-k}{2}, \frac{k-1}{2}\right).
 \end{aligned}
 \tag{13.3.16}$$

The likelihood ratio criterion is

$$\lambda = \frac{|V|}{|U + V|} = |T_2| = \frac{1}{|T_3 + I|} = \frac{1}{\prod_{j=1}^p (1 + \eta_j)} \tag{13.3.17}$$

where the η_j 's are the eigenvalues of T_3 . We reject H_o for small values of λ which means for large values of $\prod_{j=1}^p [1 + \eta_j]$. The basic objective in MANOVA consists of comparing U and V , the matrices due to the presence of treatment effects and due to the residuals, respectively. We can carry out this comparison by using the type-1 beta matrices T_1 and T_2 or the type-2 beta matrices T_3 and T_4 or by making use of the eigenvalues of these matrices. In the type-1 beta case, the eigenvalues will be between 0 and 1, whereas in the type-2 beta case, the eigenvalues will be real positive or simply positive. We may also note that the eigenvalues of T_1 and its nonsymmetric forms $U(U + V)^{-1}$ or $(U + V)^{-1}U$ are identical. Similarly, the eigenvalues of the symmetric form T_2 and $V(U + V)^{-1}$ or $(U + V)^{-1}V$ are one and the same. As well, the eigenvalues of the symmetric form T_3 and the nonsymmetric forms UV^{-1} or $V^{-1}U$ are the same. Again, the eigenvalues of the symmetric form T_4 and its nonsymmetric forms $U^{-1}V$ or VU^{-1} are the same. Several researchers have constructed tests based on the matrices T_1, T_2, T_3, T_4 or their nonsymmetric forms or their eigenvalues. Some of the well-known test statistics are the following:

$$\text{Lawley-Hotelling trace} = \text{tr}(T_3)$$

$$\text{Roy's largest root} = \text{the largest eigenvalue of } T_2$$

$$\text{Pillai's trace} = \text{tr}(T_1)$$

$$\text{Wilks' lambda} = |T_2| = \text{the likelihood ratio statistic.}$$

For example, when the hypothesis is true, we expect the eigenvalues of T_3 to be small and hence we may reject the hypothesis when its smallest eigenvalue is large or the trace of T_3 is large. If we are using T_4 , then when the hypothesis is true, we expect T_4 to be large in the sense that the eigenvalues will be large, and therefore we may reject the hypothesis for small values of its largest eigenvalue or its trace. If we are utilizing T_1 , we are actually comparing the contribution attributable to the treatments to the total variation. We expect this to be small under the hypothesis and hence, we may reject the hypothesis for large values of its smallest eigenvalue or its trace. If we are using T_2 , we are comparing the residual part to the total variation. If the hypothesis is true, then we can expect a substantial contribution from the residual part so that we may reject the hypothesis for small values of the largest eigenvalue or the trace in this case. These are the main ideas in connection with constructing statistics for testing the hypothesis on the basis of the eigenvalues of the matrices T_1, T_2, T_3 and T_4 .

13.3.6. When H_0 is rejected

When $H_0 : A_1 = \dots = A_k = O$ is rejected, it is plausible that some of the differences may be non-null, that is, $A_i - A_j \neq O$ for some i and $j, i \neq j$. We may then test individual hypotheses of the type $H_{01} : A_i = A_j$ for $i \neq j$. There are $k(k-1)/2$ such differences. This type of test is equivalent to testing the equality of the mean value vectors in two independent p -variate Gaussian populations with the same covariance matrix $\Sigma > O$. This has already been discussed in Chap. 6 for the cases Σ known and Σ unknown. In this instance, we can use the special Case (7) where for $k = 2$, and the statistic t_1 is real scalar type-2 beta distributed with the parameters $(\frac{p}{2}, \frac{n_i - 1 - p}{2})$, so that

$$\frac{n_i - 1 - p}{p} t_1 \sim F_{p, n_i - 1 - p} \quad (13.3.18)$$

where $n_i = n_i + n_j$ for some specific i and j . We can make use of (13.3.18) for testing individual hypotheses. By utilizing Special Case (8) for $k = 3$, we can also test a hypothesis of the type $A_i = A_j = A_m$ for different i, j, m . Instead of comparing the results of all the $k(k-1)/2$ individual hypotheses, we may examine the estimates of A_i , namely, $\hat{A}_i = \frac{X_i}{n_i} - \frac{X_{..}}{n}$, $i = 1, \dots, k$. Consider the norms $\|\frac{X_i}{n_i} - \frac{X_j}{n_j}\|$, $i \neq j$ (the Euclidean norm may be taken for convenience). Start with the individual test corresponding to the maximum value of these norms. If this test is not rejected, it is likely that tests on all other differences will not be rejected either. If it is rejected, we then take the next largest difference and continue testing.

Note 13.3.1. Usually, before initiating a MANOVA, the assumption that the covariance matrices associated with the k populations or treatments are equal is tested. It may happen that the error variable E_{1j} , $j = 1, \dots, n_1$, may have the common covariance matrix Σ_1 , E_{2j} , $j = 1, \dots, n_2$, may have the common covariance matrix Σ_2 , and so on, where not all the Σ_j 's equal. In this instance, we may first test the hypothesis $H_o : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$. This test is already described in Chap. 6. If this hypothesis is not rejected, we may carry out the MANOVA analysis of the data. If this hypothesis is rejected, then some of the Σ_j 's may not be equal. In this case, we test individual hypotheses of the type $\Sigma_i = \Sigma_j$ for some specific i and j , $i \neq j$. Include all treatments for which the individual hypotheses are not rejected by the tests and exclude the data on the treatments whose Σ_j 's may be different, but distinct from those already selected. Continue with the MANOVA analysis of the data on the treatments which are retained, that is, those for which the Σ_j 's are equal in the sense that the corresponding tests of equality of covariance matrices did not reject the hypotheses.

Example 13.3.2. For the sake of illustration, test the hypothesis $H_o : A_1 = A_2$ with the data provided in Example 13.3.1.

Solution 13.3.2. We can utilize some of the computations done in the solution to Example 13.3.1. Here, $n_1 = 5$, $n_2 = 4$ and $n_{\cdot} = n_1 + n_2 = 9$. We disregard the third sample. The residual sum of squares and cross products matrix in the present case is available from the Solution 13.3.1 by omitting the matrix corresponding to the third sample. Then,

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{i\cdot}}{n_i} \right) \left(X_{ij} - \frac{X_{i\cdot}}{n_i} \right)' &= \begin{bmatrix} 18 & -1 & -2 \\ & 18 & 2 \\ & & 4 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ & 6 & 7 \\ & & 10 \end{bmatrix} \\ &= \begin{bmatrix} 26 & -7 & -8 \\ -7 & 24 & 9 \\ -8 & 9 & 14 \end{bmatrix} \end{aligned}$$

whose determinant is

$$26 \begin{vmatrix} 24 & 9 \\ 9 & 14 \end{vmatrix} + 7 \begin{vmatrix} -7 & 9 \\ -8 & 14 \end{vmatrix} - 8 \begin{vmatrix} -7 & 24 \\ -8 & 9 \end{vmatrix} = 5416.$$

Let us compute $\sum_{i=1}^2 \sum_{j=1}^{n_i} (X_{ij} - \frac{X_{..}}{n_{.}})(X_{ij} - \frac{X_{..}}{n_{.}})'$:

$$\begin{aligned} \sum_{j=1}^5 \left(X_{1j} - \frac{X_{..}}{n_{.}}\right) \left(X_{1j} - \frac{X_{..}}{n_{.}}\right)' &= \begin{bmatrix} 1 & 23/9 & 5/9 \\ 23^2/9^2 & 115/9^2 & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 9 & -66/9 & -39/9 \\ 22^2/9^2 & 286/9^2 & 13^2/9^2 \end{bmatrix} \\ &+ \begin{bmatrix} 4 & 22/9 & -10/9 \\ 22^2/9^2 & -110/9^2 & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 4 & -10/9 & 26/9 \\ 5^2/9^2 & -65/9^2 & 13^2/9^2 \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ 4^2/9^2 & 16/9^2 & 4^2/9^2 \end{bmatrix} = \begin{bmatrix} 18 & -31/9 & -18/9 \\ 1538/9^2 & 242/9^2 & 404/9^2 \end{bmatrix}; \end{aligned}$$

$$\begin{aligned} \sum_{j=1}^4 \left(X_{2j} - \frac{X_{..}}{n_{.}}\right) \left(X_{2j} - \frac{X_{..}}{n_{.}}\right)' &= \begin{bmatrix} 0 & 0 & 0 \\ 14^2/9^2 & 14^2/9^2 & 14^2/9^2 \end{bmatrix} + \begin{bmatrix} 4 & -26/9 & -44/9 \\ 13^2/9^2 & (13 \times 22)/9^2 & 22^2/9^2 \end{bmatrix} \\ &+ \begin{bmatrix} 4 & -28/9 & -10/9 \\ 14^2/9^2 & 70/9^2 & 5^2/9^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 5^2/9^2 & -65/9^2 & 13^2/9^2 \end{bmatrix} \\ &= \begin{bmatrix} 8 & -6 & -6 \\ 586/9^2 & 487/9^2 & 874/9^2 \end{bmatrix}. \end{aligned}$$

Hence the sum

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^{n_i} \left(X_{ij} - \frac{X_{..}}{n_{.}}\right) \left(X_{ij} - \frac{X_{..}}{n_{.}}\right)' &= \begin{bmatrix} 18 & -31/9 & -18/9 \\ 1538/9^2 & 242/9^2 & 404/9^2 \end{bmatrix} + \begin{bmatrix} 8 & -6 & -6 \\ 586/9^2 & 487/9^2 & 874/9^2 \end{bmatrix} \\ &= \begin{bmatrix} 26 & -85/9 & -8 \\ -85/9 & 236/9 & 9 \\ -8 & 9 & 142/9 \end{bmatrix} = U + V \end{aligned}$$

whose determinant is

$$26 \begin{vmatrix} 236/9 & 9 \\ 9 & 142/9 \end{vmatrix} + \frac{85}{9} \begin{vmatrix} -85/9 & 9 \\ -8 & 142/9 \end{vmatrix} - 8 \begin{vmatrix} -85/9 & 236/9 \\ -8 & 9 \end{vmatrix} = 8380.0549.$$

So, the observed values are as follows:

$$w = \frac{|V|}{|U + V|} = \frac{5416}{8380.055} = 0.6463$$

$$t_1 = \frac{1 - w}{w} = \frac{0.3537}{0.6463} = 0.5413$$

$$\frac{n. - 1 - p}{p} t_1 = \frac{5}{3} t_1 = \frac{5}{3} (0.5413) = 0.9022,$$

and $F_{p, n. - 1 - p} = F_{3, 5}$. Let us test the hypothesis at the 5% significance level. The critical value obtained from F-tables is $F_{3, 5, 0.05} = 9.01$. But since the observed value of F is $0.9022 < 9.01$, the hypothesis is not rejected. We expected this result because the hypothesis $A_1 = A_2 = A_3$ was not rejected. This example was mainly presented to illustrate the steps.

13.4. MANOVA for Two-Way Classification Data

As was done previously for the one-way classification, we will revisit the real scalar variable case first. Thus, we consider the case of two sets of treatments, instead of the single set analyzed in Sect. 13.3. In an agricultural experiment, suppose that we are considering r fertilizers as the first set of treatments, say F_1, \dots, F_r , along with a set of s different varieties of corn, V_1, \dots, V_s , as the second set of treatments. A randomized block experiment belongs to this category. In this case, r blocks of land, which are homogeneous with respect to all factors that may affect the yield of corn, such as precipitation, fertility of the soil, exposure to sunlight, drainage, and so on, are selected. Fertilizers F_1, \dots, F_r are applied to these r blocks at random, the first block receiving any one of F_1, \dots, F_r , and so on. Each block is divided into s equivalent plots, all the plots being of the same size, shape, and so on. Then, the s varieties of corn are applied to each block at random, with one variety to each plot. Such an experiment is called a randomized block experiment. This experiment is then replicated t times. This replication is done so that possible interaction between fertilizers and varieties of corn could be tested. If the randomized block experiment is carried out only once, no interaction can be tested from such data because each plot will have only one observation. Interaction between the i -th fertilizer and j -th variety is a joint effect for the (F_i, V_j) combination, that is, the effect of F_i on the yield varies with the variety of corn. For instance, an interaction will be present if the effect of F_1 is different when combined with V_1 or V_2 . In other words, there are individual effects and joint effects, a joint effect being referred to as an interaction between the two sets of treatments. As an example, consider one set of treatments consisting of r different methods of teaching and a second set of treatments that could be s levels of previous exposure of the students to the subject matter.

13.4.1. The model in a two-way classification

The additive, fixed effect, two-way classification or two-way layout model with interaction is the following:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad i = 1, \dots, r, \quad j = 1, \dots, s, \quad k = 1, \dots, t, \quad (13.4.1)$$

where μ is a general effect, α_i is the deviation from the general effect due to the i -th treatment of the first set, β_j is the deviation from the general effect due to the j -th treatment of the second set, and γ_{ij} is the effect due to interaction term or the joint effect of first and second sets of treatments. In a randomized block experiment, the treatments belonging to the first set are called “blocks” or “rows” and the treatments belonging to the second set are called “treatments” or “columns”; thus, the two sets correspond to rows, say R_1, \dots, R_r , and columns, say C_1, \dots, C_s . Then, γ_{ij} is the deviation from the general effect due to the combination (R_i, C_j) . The random component e_{ijk} is the sum total contributions coming from all unknown factors and x_{ijk} is the observation resulting from the effect of the combination of treatments (R_i, C_j) at the k -th replication or k -th identical repetition of the experiment. In an agricultural setting, the observation may be the yield of corn whereas, in a teaching experiment, the observation may be the grade obtained by the “ (i, j, k) ”-th student. In a fixed effect model, all parameters $\mu, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s$ are assumed to be unknown constants. In a random effect model $\alpha_1, \dots, \alpha_r$ or β_1, \dots, β_s or both sets are assumed to be random variables. We assume that $E[e_{ijk}] = 0$ and $\text{Var}(e_{ijk}) = \sigma^2 > 0$ for all i, j, k , where $E(\cdot)$ denotes the expected value of (\cdot) . In the present discussion, we will only consider the fixed effect model. Under this model, the data are called two-way classification data or two-way layout data because they can be classified according to the two sets of treatments, “rows” and “columns”. Since we are not making any assumption about the distribution of e_{ijk} , and thereby that of x_{ijk} , we will apply the method of least squares to estimate the parameters.

13.4.2. Estimation of parameters in a two-way classification

The error sum of squares is

$$e_{ijk}^2 = \sum_{ijk} (x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2.$$

Our first objective consists of isolating the sum of squares due to interaction and test the hypothesis of no interaction, that is, $H_o : \gamma_{ij} = 0$ for all i, j and k . If $\gamma_{ij} \neq 0$, part of the effect of the i -th row R_i is mixed up with the interaction and similarly, part of the effect of the j -th column, C_j , is intermingled with γ_{ij} , so that no hypothesis can be tested on the α_i 's and β_j 's unless γ_{ij} is zero or negligibly small or the hypothesis $\gamma_{ij} = 0$ is not

rejected. As well, on noting that in $[\mu + \alpha_i + \beta_j + \gamma_{ij}]$, the subscripts either appear none at a time, one at a time and both at a time, we may write $\mu + \alpha_i + \beta_j + \gamma_{ij} = m_{ij}$. Thus,

$$\begin{aligned} \sum_{ijk} e_{ijk}^2 &= \sum_{ijk} (x_{ijk} - m_{ij})^2 \Rightarrow \frac{\partial}{\partial m_{ij}} [e_{ijk}^2] = 0 \\ &\Rightarrow \sum_k (x_{ijk} - m_{ij}) = 0 \Rightarrow x_{ij.} - t \hat{m}_{ij} = 0 \text{ or } \hat{m}_{ij} = \frac{x_{ij.}}{t}. \end{aligned}$$

We employ the standard notation in this area, namely that a summation over a subscript is denoted by a dot. Then, the least squares minimum under the general model or the residual sum of squares, denoted by s^2 , is given by

$$s^2 = \sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{t} \right)^2. \quad (13.4.2)$$

Now, consider the hypothesis $H_o : \gamma_{ij} = 0$ for all i and j . Under this H_o , the model becomes

$$x_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \text{ or } \sum_{ijk} e_{ijk}^2 = \sum_{ijk} (x_{ijk} - \mu - \alpha_i - \beta_j)^2.$$

We differentiate this partially with respect to μ and α_i for a specific i , and to β_j for a specific j , and then equate the results to zero and solve to obtain estimates for μ , α_i and β_j . Since we have taken α_i , β_j and γ_{ij} as deviations from the general effect μ , we may let $\alpha. = \alpha_1 + \dots + \alpha_r = 0$, $\beta. = \beta_1 + \dots + \beta_s = 0$ and $\gamma_{i.} = 0$, for each i and $\gamma_{.j} = 0$ for each j , without any loss of generality. Then,

$$\begin{aligned} \frac{\partial}{\partial \mu} [e_{ijk}^2] &= 0 \Rightarrow \left(\sum_{ijk} x_{ijk} \right) - rst\mu - st\alpha. - rt\beta. = 0 \Rightarrow \hat{\mu} = \frac{x_{...}}{rst} \\ \frac{\partial}{\partial \alpha_i} [e_{ijk}^2] &= 0 \Rightarrow \sum_{jk} [x_{ijk} - \mu - \alpha_i - \beta_j] = 0 \\ &\Rightarrow x_{i..} - st\mu - st\alpha_i - t\beta. = 0 \Rightarrow \hat{\alpha}_i = \frac{x_{i..}}{st} - \hat{\mu} \\ \frac{\partial}{\partial \beta_j} [e_{ijk}^2] &= 0 \Rightarrow \sum_{ik} [x_{ijk} - \mu - \alpha_i - \beta_j] = 0 \\ &\Rightarrow x_{.j.} - rt\mu - t\alpha. - rt\beta_j = 0 \Rightarrow \hat{\beta}_j = \frac{x_{.j.}}{rt} - \hat{\mu}. \end{aligned}$$

Hence, the least squares minimum under the hypothesis H_0 , denoted by s_0^2 , is

$$\begin{aligned} s_0^2 &= \sum_{ijk} \left[\left(x_{ijk} - \frac{x_{\dots}}{rst} \right) - \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right) - \left(\frac{x_{.j.}}{rt} - \frac{x_{\dots}}{rst} \right) \right]^2 \\ &= \sum_{ijk} \left(x_{ijk} - \frac{x_{\dots}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right)^2 - rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{\dots}}{rst} \right)^2, \end{aligned}$$

the simplifications resulting from properties of summations with respect to subscripts. Thus, the sum of squares due to the hypothesis $H_0 : \gamma_{ij} = 0$ for all i and j or the interaction sum of squares, denoted by s_γ^2 is the following:

$$\begin{aligned} s_\gamma^2 &= s_0^2 - s^2 = \sum_{ijk} \left(x_{ijk} - \frac{x_{\dots}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right)^2 \\ &\quad - rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{\dots}}{rst} \right)^2 - \sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{t} \right)^2, \end{aligned}$$

and since

$$\sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{t} \right)^2 = \sum_{ijk} \left(x_{ijk} - \frac{x_{\dots}}{rst} \right)^2 - t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{\dots}}{rst} \right)^2,$$

the sum of squares due to the hypothesis or attributable to the γ_{ij} 's, that is, due to interaction is

$$s_\gamma^2 = t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{\dots}}{rst} \right)^2 - st \sum_i \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right)^2 - rt \sum_j \left(\frac{x_{.j.}}{rt} - \frac{x_{\dots}}{rst} \right)^2. \quad (13.4.3)$$

If the hypothesis $\gamma_{ij} = 0$ is not rejected, the effects of the γ_{ij} 's are deemed insignificant and then, setting the hypothesis $\gamma_{ij} = 0$, $\alpha_i = 0$, $i = 1, \dots, r$, we obtain the sum of squares due to the α_i 's or sum of squares due to the rows denoted as s_r^2 , is

$$s_r^2 = \sum_{ijk} \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right)^2 = st \sum_{i=1}^r \left(\frac{x_{i..}}{st} - \frac{x_{\dots}}{rst} \right)^2. \quad (13.4.4)$$

Similarly, the sum of squares attributable to the β_j 's or due to the columns, denoted as s_c^2 , is

$$s_c^2 = rt \sum_{j=1}^s \left(\frac{x_{.j.}}{rt} - \frac{x_{\dots}}{rst} \right)^2. \quad (13.4.5)$$

Observe that the sum of squares due to rows plus the sum of squares due to columns, once added to the interaction sum of squares, is the subtotal sum of squares, denoted by $s_{rc}^2 = t \sum_{ij} \left(\frac{x_{ij.}}{t} - \frac{x_{...}}{rst} \right)^2$ or this subtotal sum of squares is partitioned into the sum of squares due to the rows, due to the columns and due to interaction. This is equivalent to an ANOVA on the subtotals $\sum_k x_{ijk}$ or an ANOVA on a two-way classification with a single observation per cell. As has been pointed out, in that case, we cannot test for interaction, and moreover, this subtotal sum of squares plus the residual sum of squares is the grand total sum of squares. If we assume a normal distribution for the error terms, that is, $e_{ijk} \stackrel{iid}{\sim} N_1(0, \sigma^2)$, $\sigma^2 > 0$, for all i, j, k , then under the hypothesis $H_o : \gamma_{ij} = 0$, it can be shown that

$$\frac{s_{rc}^2}{\sigma^2} \sim \chi_v^2, \quad v = (rs - 1) - (r - 1) - (s - 1) = (r - 1)(s - 1), \quad (13.4.6)$$

and the residual variation s^2 has the following distribution whether H_o holds or not:

$$\frac{s^2}{\sigma^2} \sim \chi_{v_1}^2, \quad v_1 = rst - 1 - (rs - 1) = rs(t - 1), \quad (13.4.7)$$

where s_{rc}^2 and s^2 are independently distributed. Then, under the hypothesis $\gamma_{ij} = 0$ for all i and j or when this hypothesis is not rejected, it can be established that

$$\frac{s_r^2}{\sigma^2} \sim \chi_{r-1}^2, \quad \frac{s_c^2}{\sigma^2} \sim \chi_{s-1}^2 \quad (13.4.8)$$

and s_r^2 and s^2 as well as s_c^2 and s^2 are independently distributed whenever $H_o : \gamma_{ij} = 0$ is not rejected. Hence, under the hypothesis,

$$\frac{s_{rc}^2 / (r - 1)(s - 1)}{s^2 / (rs(t - 1))} \sim F_{v, v_1}, \quad v = (r - 1)(s - 1), \quad v_1 = rs(t - 1). \quad (13.4.9)$$

The total sum of squares is $\sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{rst} \right)^2$. Thus, the first decomposition and the first part of ANOVA in this two-way classification scheme is the following:

Total variation = Variation due to the subtotals + Residual variation,

the second stage being

Variation due to the subtotals = Variation due to the rows
+ Variation due to the columns + Variation due to interaction,

and the resulting ANOVA table is the following:

ANOVA Table for the Two-Way Classification

Variation due to	df (1)	SS (2)	MS (3)=(2)/(1)
rows	$r - 1$	$s_r^2 = st \sum_{i=1}^r (\frac{x_{i.}}{st} - \frac{x_{...}}{rst})^2$	$s_r^2/(r - 1) = D_1$
columns	$s - 1$	$s_c^2 = rt \sum_{j=1}^s (\frac{x_{.j}}{rt} - \frac{x_{...}}{rst})^2$	$s_c^2/(s - 1) = D_2$
interaction	$(r - 1)(s - 1)$	s_γ^2	$s_\gamma^2/(r - 1)(s - 1) = D_3$
subtotal	$rs - 1$	$t \sum_{ij} (\frac{x_{ij.}}{t} - \frac{x_{...}}{rst})^2$	
residuals	$rs(t - 1)$	s^2	$s^2/[rs(t - 1)] = D$
total	$rst - 1$	$\sum_{ijk} (x_{ijk} - \frac{x_{...}}{rst})^2$	

where df designates the number of degrees of freedom, SS means sum of squares, MS stands for mean squares, the expressions for the residual sum of squares is given in (13.4.2), that for the interaction in (13.4.3), that for the rows in (13.4.4) and that for columns in (13.4.5), respectively. Note that we test the hypothesis on the α_i 's and β_j 's or row effects and column effects, only if the hypothesis $\gamma_{ij} = 0$ is not rejected; otherwise there is no point in testing hypotheses on the α_i 's and β_j 's because they are confounded with the γ_{ij} 's.

13.5. Multivariate Extension of the Two-Way Layout

Instead of a single real scalar variable being studied, we consider a $p \times 1$ vector of real scalar variables. The multivariate two-way classification, the fixed effect model is the following:

$$X_{ijk} = M + A_i + B_j + \Gamma_{ij} + E_{ijk}, \quad (13.5.1)$$

for $i = 1, \dots, r$, $j = 1, \dots, s$, $k = 1, \dots, t$, where M , A_i , B_j , Γ_{ij} and E_{ijk} are all $p \times 1$ vectors. In this case, M is a general effect, A_i is the deviation from the general effect due to the i -th row, B_j is the deviation from the general effect due to the j -th column, Γ_{ij} is the deviation from the general effect due to interaction between the rows and the columns and E_{ijk} is the vector of the random or error component. For convenience, the two sets of treatments are referred to as rows and columns, the first set as rows and the second, as columns. In a two-way layout, two sets of treatments are tested. As in the scalar case of Sect. 13.4, we can assume, without any loss of generality, that $\sum_i A_i = A_1 + \dots + A_r = A$, $A = O$, $B = O$, $\sum_{i=1}^r \Gamma_{ij} = \Gamma_{.j} = O$ and $\sum_{j=1}^s \Gamma_{ij} = \Gamma_{i.} = O$. At this juncture, the procedures are parallel to those developed in Sect. 13.4 for the real scalar variable case. Instead of sums of squares, we now have sums of squares and cross products matrices. As

before, we may write $M_{ij} = M + A_i + B_j + \Gamma_{ij}$. Then, the trace of the sum of squares and cross products error matrix $E_{ijk}E'_{ijk}$ is minimized. Using the vector derivative operator, we have

$$\begin{aligned} \frac{\partial}{\partial M_{ij}} \text{tr} \left[\sum_{ijk} E_{ijk} E'_{ijk} \right] = O &\Rightarrow \sum_k (X_{ijk} - M_{ij}) = O \\ &\Rightarrow \hat{M}_{ij} = \frac{1}{t} X_{ij.}, \end{aligned}$$

so that the residual sum of squares and cross products matrix, denoted by S_{res} , is

$$S_{res} = \sum_{ijk} \left(X_{ijk} - \frac{X_{ij.}}{t} \right) \left(X_{ijk} - \frac{X_{ij.}}{t} \right)'. \quad (13.5.2)$$

All other derivations are analogous to those provided in the real scalar case. The sum of squares and cross products matrix due to interaction, denoted by S_{int} is the following:

$$\begin{aligned} S_{int} &= t \sum_{ij} \left(\frac{X_{ij.}}{t} - \frac{X_{...}}{rst} \right) \left(\frac{X_{ij.}}{t} - \frac{X_{...}}{rst} \right)' \\ &\quad - st \sum_i \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right)' \\ &\quad - rt \sum_j \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right)'. \end{aligned} \quad (13.5.3)$$

The sum of squares and cross products matrices due to the rows and columns are respectively given by

$$S_{row} = st \sum_{i=1}^r \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right)', \quad (13.5.4)$$

$$S_{col} = rt \sum_{j=1}^s \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right)'. \quad (13.5.5)$$

The sum of squares and cross products matrix for the subtotal is denoted by $S_{sub} = S_{row} + S_{col} + S_{int}$. The total sum of squares and cross products matrix, denoted by S_{tot} , is the following:

$$S_{tot} = \sum_{ijk} \left(X_{ijk} - \frac{X_{...}}{rst} \right) \left(X_{ijk} - \frac{X_{...}}{rst} \right)'. \quad (13.5.6)$$

We may now construct the MANOVA table. The following abbreviations are used: df stands for degrees of freedom of the corresponding Wishart matrix, SSP means the sum of squares and cross products matrix, MS stands for mean squares and is equal to SSP/df , and S_{row} , S_{col} , S_{res} and S_{tot} are respectively specified in (13.5.4), (13.5.5), (13.5.2) and (13.5.6).

MANOVA Table for a Two-Way Layout

Variation due to	df (1)	SSP (2)	MS (3)=(2)/(1)
rows	$r - 1$	S_{row}	$S_{row}/(r - 1)$
columns	$s - 1$	S_{col}	$S_{col}/(s - 1)$
interaction	$(r - 1)(s - 1)$	S_{int}	$S_{int}/[(r - 1)(s - 1)]$
subtotal	$rs - 1$	S_{sub}	
residuals	$rs(t - 1)$	S_{res}	$S_{res}/[rs(t - 1)]$
total	$rst - 1$	S_{tot}	

13.5.1. Likelihood ratio test for multivariate two-way layout

Under the assumption that the error or random components $E_{ijk} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$ for all i, j and k , the exponential part of the multivariate normal density excluding $-\frac{1}{2}$ is obtained as follows:

$$\begin{aligned} E'_{ijk} E_{ijk} &= (X_{ijk} - M - A_i - B_j - \Gamma_{ij})' \Sigma^{-1} (X_{ijk} - M - A_i - B_j - \Gamma_{ij}) \\ &= \text{tr}[\Sigma^{-1} (X_{ijk} - M - A_i - B_j - \Gamma_{ij})(X_{ijk} - M - A_i - B_j - \Gamma_{ij})'] \Rightarrow \\ \sum_{ijk} E'_{ijk} E_{ijk} &= \text{tr} \left\{ \Sigma^{-1} \left[\sum_{ijk} (X_{ijk} - M - A_i - B_j - \Gamma_{ij})(X_{ijk} - M - A_i - B_j - \Gamma_{ij})' \right] \right\}. \end{aligned}$$

Thus, the joint density of all the X_{ijk} 's, denoted by L , is

$$\begin{aligned} L &= \frac{1}{(2\pi)^{\frac{prst}{2}} |\Sigma|^{\frac{rst}{2}}} \\ &\quad \times e^{-\frac{1}{2} \text{tr}[\Sigma^{-1} \sum_{ijk} (X_{ijk} - M - A_i - B_j - \Gamma_{ij})(X_{ijk} - M - A_i - B_j - \Gamma_{ij})']}. \end{aligned}$$

The maximum likelihood estimates of M , A_i , B_j and Γ_{ij} are the same as the least squares estimates and hence, the maximum likelihood estimator (MLE) of Σ is the least squares

minimum which is the residual sum of squares and cross products matrix S or S_{res} (in the present notation), where

$$\begin{aligned} S_{res} &= \sum_{ijk} (X_{ijk} - \hat{M} - \hat{A}_i - \hat{B}_j - \hat{\Gamma}_{ij})(X_{ijk} - \hat{M} - \hat{A}_i - \hat{B}_j - \hat{\Gamma}_{ij})' \\ &= \sum_{ijk} \left(X_{ijk} - \frac{X_{ij.}}{t} \right) \left(X_{ijk} - \frac{X_{ij.}}{t} \right)'. \end{aligned} \tag{13.5.7}$$

This is the sample sum of squares and the cross products matrix under the general model and its determinant raised to the power of $\frac{rst}{2}$ is the quantity appearing in the numerator of the likelihood ratio criterion λ . Consider the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j . Then, under this hypothesis, the estimator of Σ is S_0 , where

$$\begin{aligned} S_0 &= \sum_{ijk} \left[\left(X_{ijk} - \frac{X_{...}}{rst} \right) - \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) - \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \right] \\ &\quad \times \left[\left(X_{ijk} - \frac{X_{...}}{rst} \right) - \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) - \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \right]' \end{aligned}$$

and $|S_0|^{\frac{rst}{2}}$ is the quantity appearing in the denominator of λ . However, $S_0 - S_{res} = S_{int}$ is the sum of squares and cross products matrix due to the interaction terms Γ_{ij} 's or to the hypothesis, so that $S_0 = S_{res} + S_{int}$. Therefore, λ is given by

$$\lambda = \frac{|S_{res}|^{\frac{rst}{2}}}{|S_{res} + S_{int}|^{\frac{rst}{2}}} \tag{13.5.8}$$

Letting $w = \lambda^{\frac{2}{rst}}$,

$$w = \frac{|S_{res}|}{|S_{res} + S_{int}|}. \tag{13.5.9}$$

It follows from results derived in Chap. 5 that $S_{res} \sim W_p(rs(t-1), \Sigma)$, $S_{int} \sim W_p((r-1)(s-1), \Sigma)$ under the hypothesis and S_{res} and S_{int} are independently distributed and hence, under H_o ,

$$W = (S_{res} + S_{int})^{-\frac{1}{2}} S_{res} (S_{res} + S_{int})^{-\frac{1}{2}} \sim \text{real } p\text{-variate type-1 beta random variable}$$

with the parameters $(\frac{rs(t-1)}{2}, \frac{(r-1)(s-1)}{2})$. As well,

$$W_1 = S_{res}^{-\frac{1}{2}} S_{int} S_{res}^{-\frac{1}{2}} \sim \text{real } p\text{-variate type-2 beta random variable}$$

with the parameters $(\frac{(r-1)(s-1)}{2}, \frac{rs(t-1)}{2})$. Under H_o , the h -th arbitrary moments of w and λ , which are readily obtained from those of a real matrix-variate type-1 beta variable, are

$$E[w]^h = \left\{ \prod_{j=1}^p \frac{\Gamma(v_1 + v_2 - \frac{j-1}{2})}{\Gamma(v_1 - \frac{j-1}{2})} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(v_1 + h - \frac{j-1}{2})}{\Gamma(v_1 + v_2 + h - \frac{j-1}{2})} \right\} \quad (13.5.10)$$

$$E[\lambda]^h = \left\{ \prod_{j=1}^p \frac{\Gamma(v_1 + v_2 - \frac{j-1}{2})}{\Gamma(v_1 - \frac{j-1}{2})} \right\} \left\{ \prod_{j=1}^p \frac{\Gamma(v_1 + \frac{rst}{2}h - \frac{j-1}{2})}{\Gamma(v_1 + v_2 + \frac{rst}{2}h - \frac{j-1}{2})} \right\} \quad (13.5.11)$$

where $v_1 = \frac{rs(t-1)}{2}$ and $v_2 = \frac{(r-1)(s-1)}{2}$. Note that we reject the null hypothesis $H_o : \Gamma_{ij} = O, i = 1, \dots, r, j = 1, \dots, s$, for small values of w and λ . As explained in Sect. 13.3, the exact general density of w in (13.5.10) can be expressed in terms of a G-function and the exact general density of λ in (13.5.11) can be written in terms of a H-function. For the theory and applications of the G-function and the H-function, the reader may respectively refer to Mathai (1993) and Mathai et al. (2010).

13.5.2. Asymptotic distribution of λ in the MANOVA two-way layout

Consider the arbitrary h -th moment specified in (13.5.11). On expanding all the gamma functions for large values of rst in the constant part and for large values of $rst(1+h)$ in the functional part by applying Stirling's formula or using the first term in the asymptotic expansion of a gamma function referring to (13.3.13), it can be verified that the h -th moment of λ behaves asymptotically as follows:

$$\lambda^h \rightarrow (1+h)^{-\frac{p(r-1)(s-1)}{2}} \Rightarrow -2 \ln \lambda \rightarrow \chi_{p(r-1)(s-1)}^2 \text{ as } rst \rightarrow \infty. \quad (13.5.12)$$

Thus, for large values of rst , one can utilize this real scalar chisquare approximation for testing the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j . We can work out a large number of exact distributions of w of (13.5.10) for special values of r, s, t, p . Observe that

$$E[w^h] = C \prod_{j=1}^p \frac{\Gamma(\frac{rs(t-1)}{2} - \frac{j-1}{2} + h)}{\Gamma(\frac{rs(t-1)}{2} + \frac{(r-1)(s-1)}{2} - \frac{j-1}{2} + h)} \quad (13.5.13)$$

where C is the normalizing constant such that when $h = 0$, $E[w^h] = 1$. Thus, when $(r-1)(s-1)$ is a positive integer or when r or s is odd, the gamma functions cancel out, leaving a number of factors in the denominator which can be written as a sum by applying the partial fractions technique. For small values of p , the exact density will then be expressible as a sum involving only a few terms. For larger values of p , there will be repeated factors in the denominator, which complicates matters.

13.5.3. Exact densities of w in some special cases

We will consider several special cases of the h -th moment of w as given in (13.5.13).

Case (1): $p = 1$. In this case, h -th moment becomes

$$E[w^h] = C_1 \frac{\Gamma(\frac{rs(t-1)}{2} + h)}{\Gamma(\frac{rs(t-1)}{2} + \frac{(r-1)(s-1)}{2} + h)}$$

where C_1 is the associated normalizing constant. This is the h -th moment of a real scalar type-1 beta random variable with the parameters $(\frac{rs(t-1)}{2}, \frac{(r-1)(s-1)}{2})$. Hence $y = \frac{1-w}{w}$ is a real scalar type-2 beta with parameters $(\frac{(r-1)(s-1)}{2}, \frac{rs(t-1)}{2})$, and

$$\frac{rs(t-1)}{(r-1)(s-1)} y \sim F_{(r-1)(s-1), rs(t-1)}.$$

Accordingly, the test can be carried out by using this F -statistic. One would reject the null hypothesis $H_o : \Gamma_{ij} = O$ if the observed $F \geq F_{(r-1)(s-1), rs(t-1), \alpha}$ where $F_{(r-1)(s-1), rs(t-1), \alpha}$ is the upper 100 $\alpha\%$ percentile of this F -density. For example, for $r = 2, s = 3, t = 3$ and $\alpha = 0.05$, we have $F_{2,12,0.05} = 19.4$ from F-tables so that H_o would be rejected if the observed value of $F_{2,12} \geq 19.4$ at the specified significance level.

Case (2): $p = 2$. In this case, we have a ratio of two gamma functions differing by $\frac{1}{2}$. Combining the gamma functions in the numerator and in the denominator by using the duplication formula and proceeding as in Sect. 13.3 for the one-way layout, the statistic $t_1 = \frac{1-\sqrt{w}}{\sqrt{w}}$, and we have

$$\frac{rs(t-1)}{(r-1)(s-1)} t_1 \sim F_{2(r-1)(s-1), 2(rs(t-1)-1)},$$

so that the decision can be made as in Case (1).

Case (3): $(r-1)(s-1) = 1 \Rightarrow r = 2, s = 2$. In this case, all the gamma functions in (13.3.13) cancel out except the last one in the numerator and the first one in the denominator. This gamma ratio is that of a real scalar type-1 beta random variable with the parameters $(\frac{rs(t-1)+1-p}{2}, \frac{p}{2})$, and hence $y = \frac{1-w}{w}$ is a real scalar type-2 beta so that

$$\frac{rs(t-1)+1-p}{p} y \sim F_{p, rs(t-1)+1-p},$$

and decision can be made by making use of this F distribution as in Case (1).

Case(4): $(r - 1)(s - 1) = 2$. In this case,

$$E[w^h] = C_1 \prod_{j=1}^p \frac{1}{\frac{rs(t-1)}{2} - \frac{j-1}{2} + h}$$

with the corresponding normalizing constant C_1 . This product of p factors can be expressed as a sum by using partial fractions. That is,

$$E[w^h] = C_1 \sum_{j=0}^{p-1} \frac{b_j}{a + h - \frac{j}{2}} \quad (i)$$

where

$$b_j = \lim_{a+h \rightarrow \frac{j}{2}} [(a+h)(a+h-\frac{1}{2}) \cdots (a+h-\frac{j-1}{2})(a+h-\frac{j+1}{2}) \cdots (a+h-\frac{p-1}{2})], \quad (ii)$$

$$a = \frac{rs(t-1)}{2}.$$

Thus, the density of w , denoted by $f_w(w)$, which is available from (i) and (ii), is the following:

$$f_w(w) = C_1 \sum_{j=0}^{p-1} b_j w^{a-\frac{j}{2}-1}, \quad 0 \leq w \leq 1,$$

and zero elsewhere. Some additional special cases could be worked out but the expressions would become complicated. For large values of rst , one can apply the asymptotic chisquare result given in (13.5.12) for testing the hypothesis $H_o : \Gamma_{ij} = O$.

Example 13.5.1. An experiment is conducted among heart patients to stabilize their systolic pressure, diastolic pressure and heart rate or pulse around the standard numbers which are 120, 80 and 60, respectively. A random sample of 24 patients who may be considered homogeneous with respect to all factors of variation, such as age, weight group, race, gender, dietary habits, and so on, are selected. These 24 individuals are randomly divided into two groups of equal size. One group of 12 subjects are given the medication combination Med-1 and the other 12 are administered the medication combination Med-2. Then, the Med-1 group is randomly divided into three subgroups of 4 subjects. These subgroups are assigned exercise routines Ex-1, Ex-2, Ex-3. Similarly, the Med-2 group is also divided at random into 3 subgroups of 4 individuals who are respectively subjected to exercise routines Ex-1, Ex-2, Ex-3. After one week, the following observations are made $x_1 =$ current

reading on systolic pressure minus 120, x_2 = current reading on diastolic pressure minus 80, x_3 = current reading on heart rate minus 60. The structure of the two-way data layout is as follows:

	Ex-1	Ex-2	Ex-3
Med-1	four 3×1 vectors	four 3×1 vectors	four 3×1 vectors
Med-2	four 3×1 vectors	four 3×1 vectors	four 3×1 vectors

Let X_{ijk} be the k -th vector in the i -th row (i -th medication) and j -th column (j -th exercise routine). For convenience, the data are presented in matrix form:

$$\begin{aligned}
 A_{11} &= [X_{111}, X_{112}, X_{113}, X_{114}], & A_{12} &= [X_{121}, X_{122}, X_{123}, X_{124}], \\
 A_{13} &= [X_{131}, X_{132}, X_{133}, X_{134}], & A_{21} &= [A_{211}, A_{212}, A_{213}, A_{214}], \\
 A_{22} &= [A_{221}, A_{222}, A_{223}, A_{224}], & A_{23} &= [X_{231}, X_{232}, X_{233}, X_{234}];
 \end{aligned}$$

$$\begin{aligned}
 A_{11} &= \begin{bmatrix} -2 & 3 & 2 & 5 \\ 1 & -1 & 1 & -1 \\ 2 & -1 & -1 & 0 \end{bmatrix}, & A_{12} &= \begin{bmatrix} 1 & 4 & -1 & 4 \\ -2 & -2 & -3 & 3 \\ 3 & -2 & -1 & 0 \end{bmatrix}, & A_{13} &= \begin{bmatrix} 4 & -3 & 3 & 4 \\ 2 & -3 & 2 & 3 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \\
 A_{21} &= \begin{bmatrix} 2 & 0 & -2 & 0 \\ 1 & 4 & 1 & 2 \\ 2 & 1 & -1 & -2 \end{bmatrix}, & A_{22} &= \begin{bmatrix} 3 & -1 & -1 & 3 \\ 4 & 4 & 0 & 0 \\ 0 & 1 & -1 & 4 \end{bmatrix}, & A_{23} &= \begin{bmatrix} -2 & -1 & 0 & -1 \\ 1 & 4 & 0 & 3 \\ -2 & 0 & -2 & 0 \end{bmatrix}.
 \end{aligned}$$

- (1) Perform a two-way ANOVA on the first component, namely, x_1 , the current reading minus 120; (2) Carry out a MANOVA on the full data.

Solution 13.5.1. We need the following quantities:

$$\begin{aligned}
 X_{11.} &= \begin{bmatrix} 8 \\ 0 \\ 0 \end{bmatrix}, & X_{12.} &= \begin{bmatrix} 8 \\ -4 \\ 0 \end{bmatrix}, & X_{13.} &= \begin{bmatrix} 8 \\ 4 \\ 0 \end{bmatrix} & \Rightarrow & X_{1..} &= \begin{bmatrix} 24 \\ 0 \\ 0 \end{bmatrix} \\
 X_{21.} &= \begin{bmatrix} 0 \\ 8 \\ 0 \end{bmatrix}, & X_{22.} &= \begin{bmatrix} 4 \\ 8 \\ 4 \end{bmatrix}, & X_{23.} &= \begin{bmatrix} -4 \\ 8 \\ -4 \end{bmatrix} & \Rightarrow & X_{2..} &= \begin{bmatrix} 0 \\ 24 \\ 0 \end{bmatrix} \\
 X_{.1.} &= \begin{bmatrix} 8 \\ 8 \\ 0 \end{bmatrix}, & X_{.2.} &= \begin{bmatrix} 12 \\ 4 \\ 4 \end{bmatrix}, & X_{.3.} &= \begin{bmatrix} 4 \\ 12 \\ -4 \end{bmatrix} & \Rightarrow & X_{...} &= \begin{bmatrix} 24 \\ 24 \\ 0 \end{bmatrix} \\
 \bar{X} &= \frac{X_{...}}{rst} = \frac{1}{24} \begin{bmatrix} 24 \\ 24 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, & r &= 2, & s &= 3, & t &= 4.
 \end{aligned}$$

By using the first elements in all these vectors, we will carry out a two-way ANOVA and answer the first question. Since these are all observations on real scalar variables, we will utilize lower-case letters to indicate scalar quantities. Thus, we have the following values:

$$\begin{aligned}
 s_{tot}^2 &= \sum_{ij} (x_{ijk} - \bar{x})^2 = (-2 - 1)^2 + (3 - 1)^2 + \cdots + (-1 - 1)^2 = 136, \\
 s_{row}^2 &= 12 \sum_{i=1}^2 \left(\frac{x_{i..}}{12} - 1 \right)^2 = 12[(2 - 1)^2 + (0 - 1)^2] = 24, \\
 s_{col}^2 &= 8 \sum_{j=1}^3 \left(\frac{x_{.j.}}{8} - 1 \right)^2 = 8 \left[(1 - 1)^2 + \left(\frac{3}{2} - 1 \right)^2 + \left(\frac{1}{2} - 1 \right)^2 \right] = 4, \\
 s_{int}^2 &= 4 \sum_{ij} \left(\frac{x_{ij.}}{4} - \frac{x_{i..}}{12} - \frac{x_{.j.}}{8} + 1 \right)^2 \\
 &= 4 \left[\left(\frac{8}{4} - \frac{24}{12} - \frac{8}{8} + 1 \right)^2 + \cdots + \left(-\frac{4}{4} - 0 - \frac{4}{8} + 1 \right)^2 \right] = 4, \\
 s_{sub}^2 &= 4 \sum_{ij} \left(\frac{x_{ij.}}{4} - \frac{x_{...}}{24} \right)^2 = 4 \left[\left(\frac{8}{4} - 1 \right)^2 + \cdots + \left(-\frac{4}{4} - 1 \right)^2 \right] = 32, \\
 s_{res}^2 &= \sum_{ijk} \left(x_{ijk} - \frac{x_{ij.}}{4} \right)^2 = (-2 - 2)^2 + \cdots + (-1 + 1)^2 = 104, \\
 s_{tot}^2 &= \sum_{ijk} \left(x_{ijk} - \frac{x_{...}}{24} \right)^2 = (-2 - 1)^2 + (3 - 1)^2 + \cdots + (-1 - 1)^2 = 136.
 \end{aligned}$$

All quantities have been calculated separately in order to verify the computations. We could have obtained the interaction sum of squares from the subtotal sum of squares minus the sum of squares due to rows and columns. Similarly, we could have obtained the residual sum of squares from the total sum of squares minus the subtotal sum of squares. We will set up the ANOVA table, where, as usual, *df* stands for degrees of freedom, *SS* means sum of squares and *MS* denotes mean squares:

ANOVA Table for a Two-Way Layout with Interaction

Variation due to	<i>df.</i> (1)	<i>SS</i> (2)	<i>MS</i> (3)=(2)/(1)	F-ratio
rows	1	24	24	24/5.78
columns	2	4	2	2/5.78
interaction	2	4	2	2/5.78
subtotal	5	32		
residuals	18	104	5.78	
total	23	136		

For testing the hypothesis of no interaction, the F -value at the 5% significance level is $F_{2,18,0.05} \approx 19$. The observed value of this $F_{2,18}$ being $\frac{2}{5.78} \approx 0.35 < 19$, the hypothesis of no interaction is not rejected. Thus, we can test for the significance of the row and column effects. Consider the hypothesis $\alpha_1 = \alpha_2 = 0$. Then under this hypothesis and no interaction hypothesis, the F -ratio for the row sum of squares is $24/5.78 \approx 4.15 < 240 = F_{1,18,0.05}$, the tabulated value of $F_{1,18}$ at $\alpha = 0.05$. Therefore, this hypothesis is not rejected. Now, consider the hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$. Since under this hypothesis and the hypothesis of no interaction, the F -ratio for the column sum of squares is $\frac{2}{5.78} = 0.35 < 19 = F_{2,18,0.05}$, it is not rejected either. Thus, the data show no significant interaction between exercise routine and medication, and no significant effect of the exercise routines or the two combinations of medications in bringing the systolic pressures closer to the standard value of 120.

We now carry out the computations needed to perform a MANOVA on the full data. We employ our standard notation by denoting vectors and matrices by capital letters. The sum of squares and cross products matrices for the rows and columns are the following, respectively denoted by S_{row} and S_{col} :

$$\begin{aligned}
 S_{row} &= st \sum_{i=1}^2 \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right) \left(\frac{X_{i..}}{st} - \frac{X_{...}}{rst} \right)' \\
 &= 12 \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} [1, -1, 0] + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} [-1, 1, 0] \right\} = 12 \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix},
 \end{aligned}$$

$$\begin{aligned}
S_{col} &= rt \sum_{j=1}^3 \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right) \left(\frac{X_{.j.}}{rt} - \frac{X_{...}}{rst} \right)' \\
&= 8 \left\{ O + \frac{1}{4} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix} \right\} = \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
S_{int} &= t \sum_{ij} \left(\frac{X_{ij.}}{t} - \frac{X_{i..}}{st} - \frac{X_{.j.}}{rt} + \frac{X_{...}}{rst} \right) \left(\frac{X_{ij.}}{t} - \frac{X_{i..}}{st} - \frac{X_{.j.}}{rt} + \frac{X_{...}}{rst} \right)' \\
&= 4 \left\{ O + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + O \right. \\
&\quad \left. + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right\} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
S_{sub} &= t \sum_{ij} \left(\frac{X_{ij.}}{t} - \bar{X} \right) \left(\frac{X_{ij.}}{t} - \bar{X} \right)' \\
&= \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 32 & -24 & 8 \\ -24 & 32 & 0 \\ 8 & 0 & 8 \end{bmatrix}.
\end{aligned}$$

We can verify the computations done so far as follows. The sum of squares and cross product matrices ought to be such that $S_{row} + S_{col} + S_{int} = S_{sub}$. These are

$$\begin{aligned}
S_{row} + S_{col} + S_{int} &= \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 32 & -24 & 8 \\ -24 & 32 & 0 \\ 8 & 0 & 8 \end{bmatrix} = S_{sub}.
\end{aligned}$$

Hence the result is verified. Now, the total and residual sums of squares and cross product matrices are

$$\begin{aligned}
 S_{tot} &= \sum_{ijk} (X_{ijk} - \bar{X})(X_{ijk} - \bar{X})' \\
 &= \begin{bmatrix} 9 & 0 & -6 \\ 0 & 0 & 0 \\ -6 & 0 & 4 \end{bmatrix} + \dots + \begin{bmatrix} 4 & -4 & 0 \\ -4 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 136 & 11 & 16 \\ 11 & 112 & 6 \\ 16 & 6 & 60 \end{bmatrix}
 \end{aligned}$$

and

$$\begin{aligned}
 S_{res} &= \sum_{ijk} \left(X_{ijk} - \frac{X_{ij.}}{t} \right) \left(X_{ijk} - \frac{X_{ij.}}{t} \right)' \\
 &= \begin{bmatrix} 16 & -14 & -8 \\ -4 & 1 & 2 \\ -8 & 2 & 4 \end{bmatrix} + \dots + \begin{bmatrix} 2 & 3 & 0 \\ 3 & 10 & 2 \\ 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix}.
 \end{aligned}$$

Then,

$$S_{res} + S_{int} = \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 108 & 39 & 12 \\ 39 & 84 & 10 \\ 12 & 10 & 56 \end{bmatrix}.$$

The above results are included in the following MANOVA table where *df* means degrees of freedom, *SSP* denotes a sum of squares and cross products matrix and *MS* is equal to *SSP* divided by the corresponding degrees of freedom:

MANOVA Table for a Two-Way Layout with Interaction

Variation due to	<i>df.</i> (1)	<i>SSP</i> (2)	<i>MS</i> (3)=(2)/(1)
rows	1	S_{row}	S_{row}
columns	2	S_{col}	$\frac{1}{2} S_{col}$
interaction	2	S_{int}	$\frac{1}{2} S_{int}$
subtotal	5	S_{sub}	
residuals	18	S_{res}	$\frac{1}{18} S_{res}$
total	23	S_{tot}	

Then, the λ -criterion is

$$\lambda = \frac{|S_{res}|^{\frac{rst}{2}}}{|S_{res} + S_{int}|^{\frac{rst}{2}}} \Rightarrow w = \frac{|S_{res}|}{|S_{res} + S_{int}|}.$$

The determinants are as follows:

$$\begin{aligned} |S_{res}| &= 104 \begin{vmatrix} 80 & 6 \\ 6 & 52 \end{vmatrix} - 35 \begin{vmatrix} 35 & 6 \\ 8 & 52 \end{vmatrix} + 8 \begin{vmatrix} 35 & 80 \\ 8 & 6 \end{vmatrix} = 363436 \\ |S_{res} + S_{int}| &= 108 \begin{vmatrix} 84 & 10 \\ 10 & 56 \end{vmatrix} - 39 \begin{vmatrix} 39 & 10 \\ 12 & 56 \end{vmatrix} + 12 \begin{vmatrix} 39 & 84 \\ 12 & 10 \end{vmatrix} = 409320. \end{aligned}$$

Therefore,

$$w = \frac{363436}{409320} = 0.888 \Rightarrow \ln w = -0.118783 \Rightarrow -2 \ln \lambda = 24(0.118783) = 2.8508.$$

We have explicit simple representations of the exact densities for the special cases $p = 1$, $p = 2$, $t = 2$, $t = 3$. However, our situation being $p = 3$, $t = 4$, they do not apply. A chisquare approximation is available for large values of rst , but our rst is only equal to 24. In this instance, $-2 \ln \lambda \rightarrow \chi_{p(r-1)(s-1)}^2 \simeq \chi_6^2$ as $rst \rightarrow \infty$. However, since the observed value of $-2 \ln \lambda = 2.8508$ happens to be much smaller than the critical value resulting from the asymptotic distribution, which is $\chi_{6,0.05}^2 = 12.59$ in this case, we can still safely decide not to reject the hypothesis $H_o : \Gamma_{ij} = O$ for all i and j , and go ahead and test for the main row and column effects, that is, the main effects of medical combinations Med-1 and Med-2 and the main effects of exercise routines Ex-1, Ex-2 and Ex-3. For testing the row effect, our hypothesis is $A_1 = A_2 = O$ and for testing the column effect, it is $B_1 = B_2 = B_3 = O$, given that $\Gamma_{ij} = O$ for all i and j . The corresponding likelihood ratio criteria are respectively,

$$\lambda_1 = \left(\frac{|S_{res}|}{|S_{res} + S_{row}|} \right)^{\frac{rst}{2}} \text{ and } \lambda_2 = \left(\frac{|S_{res}|}{|S_{res} + S_{col}|} \right)^{\frac{rst}{2}},$$

and we may utilize $w_j = \lambda_j^{\frac{2}{rst}}$, $j = 1, 2$. From previous calculations, we have

$$\begin{aligned} S_{res} + S_{row} &= \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 24 & -24 & 0 \\ -24 & 24 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 128 & 11 & 8 \\ 11 & 104 & 6 \\ 8 & 6 & 52 \end{bmatrix}, \\ S_{res} + S_{col} &= \begin{bmatrix} 104 & 35 & 8 \\ 35 & 80 & 6 \\ 8 & 6 & 52 \end{bmatrix} + \begin{bmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{bmatrix} = \begin{bmatrix} 108 & 31 & 12 \\ 31 & 84 & 2 \\ 12 & 2 & 56 \end{bmatrix}. \end{aligned}$$

The required determinants are as follows:

$$\begin{aligned}
 |S_{res}| &= 363436, \quad |S_{res} + S_{row}| = 675724, \quad |S_{res} + S_{col}| = 443176 \Rightarrow \\
 w_1 &= \frac{363436}{675724} = 0.5378468 \quad \text{and} \quad w_2 = \frac{363436}{443176} = 0.8200714, \\
 -2 \ln \lambda_1 &= -24 \ln 0.5378468 = 24(0.62018) = 14.88, \\
 \chi_{p(r-1), \alpha}^2 &= \chi_{3, 0.05}^2 = 7.81 < 14.88; \quad (i) \\
 -2 \ln \lambda_2 &= -24 \ln 0.8200714 = 24(0.19836) = 4.76, \\
 \chi_{p(s-1), \alpha}^2 &= \chi_{6, 0.05}^2 = 12.59 > 4.76. \quad (ii)
 \end{aligned}$$

When $rst \rightarrow \infty$, $-2 \ln \lambda_1 \rightarrow \chi_{p(r-1)}^2$ and $-2 \ln \lambda_2 \rightarrow \chi_{p(s-1)}^2$, referring to Exercises 13.5.9 and 13.5.10, respectively. These results follow from the asymptotic expansion provided in Sect. 12.5.2. Even though $rst = 24$ is not that large, we may use these chisquare approximations for making decisions as the exact densities of w_1 and w_2 do not fall into the special cases previously discussed. When making use of the likelihood ratio criterion, we reject the hypotheses $A_1 = A_2 = O$ and $B_1 = B_2 = B_3 = O$ for small values of λ_1 and λ_2 , respectively, which translates into large values of the approximate chisquare values. It is seen from (i) that the observed value $-2 \ln \lambda_1$ is larger than the tabulated critical value and hence we reject the hypothesis $A_1 = A_2 = O$ at the 5% level. However, the hypothesis $B_1 = B_2 = B_3 = O$ is not rejected since the observed value is less than the critical value. We may conclude that the present data does not show any evidence of interaction between the exercise routines and medication combinations, that the exercise routine does not contribute significantly to bringing the subjects' initial readings closer to the standard values (120, 80, 60), whereas there is a possibility that the medical combinations Med-1 and Med-2 are effective in significantly causing the subjects' initial readings to approach standard values.

Note 13.5.1. It may be noticed from the MANOVA table that the second stage analysis will involve one observation per cell in a two-way layout, that is, the (i, j) -th cell will contain only one observation vector X_{ij} . for the second stage analysis. Thus, $S_{sub} = S_{int} + S_{row} + S_{col}$ (the corresponding sum of squares in the real scalar case), and in this analysis with a single observation per cell, S_{int} acts as the residual sum of squares and cross products matrix (the residual sum of squares in the real scalar case). Accordingly, "interaction" cannot be tested when there is only a single observation per cell.

Exercises

13.1. In the ANOVA table obtained in Example 13.5.1, prove that (1) the sum of squares due to interaction and the residual sum of squares, (2) the sum of squares due to rows and residual sum of squares, (3) the sum of squares due to columns and residual sum of squares, are independently distributed under the normality assumption for the error variables, that is, $e_{ijk} \stackrel{iid}{\sim} N_1(0, \sigma^2)$, $\sigma^2 > 0$.

13.2. In the MANOVA table obtained in Example 13.5.1, prove that (1) S_{int} and S_{res} , (2) S_{row} and S_{res} , (3) S_{col} and S_{res} , are independently distributed Wishart matrices when $E_{ijk} \stackrel{iid}{\sim} N_p(O, \Sigma)$, $\Sigma > O$.

13.3. In a one-way layout, the following are the data on four treatments. (1) Carry out a complete ANOVA on the first component (including individual comparisons if the hypothesis of no interaction is not rejected). (2) Perform a full MANOVA on the full data.

$$\begin{aligned} \text{Treatment-1} & \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}, & \text{Treatment-2} & \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \\ \text{Treatment-3} & \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, & \text{Treatment-4} & \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 3 \end{bmatrix}. \end{aligned}$$

13.4. Carry out a full one-way MANOVA on the following data:

$$\begin{aligned} \text{Treatment-1} & \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, & \text{Treatment-2} & \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix}, \\ \text{Treatment-3} & \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}. \end{aligned}$$

13.5. The following are the data on a two-way layout where A_{ij} denotes the data on the i -th row and j -th column cell. (1) Perform a complete ANOVA on the first component. (2) Carry out a full MANOVA on the full data. (3) Verify that $S_{row} + S_{col} + S_{int} = S_{sub}$ and $S_{sub} + S_{res} = S_{tot}$, (4) Evaluate the exact density of w .

$$\begin{aligned} A_{11} &= \begin{matrix} 2 & 1 & 2 & -1 \\ 1 & 2 & 1 & 0 \\ -1 & 3 & 1 & 4 \end{matrix}, & A_{12} &= \begin{matrix} 1 & 3 & 1 & 1 \\ 4 & 4 & 1 & 1 \\ 6 & 3 & -1 & 0 \end{matrix}, & A_{13} &= \begin{matrix} 1 & -1 & 1 & -1 \\ -2 & 1 & 2 & 1 \\ 1 & -2 & 1 & -2 \end{matrix}, \\ A_{21} &= \begin{matrix} 3 & 4 & 2 & 3 \\ 2 & -1 & 2 & 3 \\ 3 & 5 & 2 & 2 \end{matrix}, & A_{22} &= \begin{matrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{matrix}, & A_{23} &= \begin{matrix} 2 & 3 & 3 & 0 \\ 3 & 2 & -1 & 2 \\ 3 & -3 & 2 & 4 \end{matrix}. \end{aligned}$$

13.6. Carry out a complete MANOVA on the following data where A_{ij} indicates the data in the i -th row and j -th column cell.

$$A_{11} = \begin{pmatrix} 1 & -1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 4 & 2 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 4 & 3 & 4 & 2 & 2 \\ 5 & 6 & 7 & 2 & 1 \end{pmatrix}, \quad A_{13} = \begin{pmatrix} 5 & 3 & 5 & 4 & 5 \\ 5 & 4 & 2 & 4 & 5 \end{pmatrix},$$

$$A_{21} = \begin{pmatrix} 0 & 1 & 1 & 0 & -2 \\ 2 & 2 & 0 & -1 & -1 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 6 & 7 & 6 & 4 & 5 \\ 4 & 5 & 2 & 3 & 4 \end{pmatrix}, \quad A_{23} = \begin{pmatrix} 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & 2 & 1 & 2 \end{pmatrix}.$$

13.7. Under the hypothesis $A_1 = \cdots = A_r = O$, prove that $U_1 = (S_{res} + S_{row})^{-\frac{1}{2}} S_{res} (S_{res} + S_{row})^{-\frac{1}{2}}$, is a real matrix-variate type-1 beta with the parameters $(\frac{rs(t-1)}{2}, \frac{r-1}{2})$ for $r \geq p$, $rs(t-1) \geq p$, when the hypothesis $\Gamma_{ij} = O$ for all i and j is not rejected or assuming that $\Gamma_{ij} = O$. The determinant of U_1 appears in the likelihood ratio criterion in this case.

13.8. Under the hypothesis $B_1 = \cdots = B_s = O$ when the hypothesis $\Gamma_{ij} = O$ is not rejected, or assuming that $\Gamma_{ij} = O$, prove that $U_2 = (S_{res} + S_{col})^{-\frac{1}{2}} S_{res} (S_{res} + S_{col})^{-\frac{1}{2}}$ is a real matrix-variate type-1 beta random variable with the parameters $(\frac{rs(t-1)}{2}, \frac{s-1}{2})$ for $s \geq p$, $rs(t-1) \geq p$. The determinant of U_2 appears in the likelihood ratio criterion for testing the main effect $B_j = O$, $j = 1, \dots, s$.

13.9. Show that when $rst \rightarrow \infty$, $-2 \ln \lambda_1 \rightarrow \chi_{p(r-1)}^2$, that is, $-2 \ln \lambda_1$ asymptotically tends to a real scalar chisquare having $p(r-1)$ degrees of freedom, where $\lambda_1 = |U_1|$ and U_1 is as defined in Exercise 13.7. [Hint: Look into the general h -th moment of λ_1 in this case, which can be evaluated by using the density of U_1]. Hence for large values of rst , one can use this approximate chisquare distribution for testing the hypothesis $A_1 = \cdots = A_r = O$.

13.10. Show that when $rst \rightarrow \infty$, $-2 \ln \lambda_2 \rightarrow \chi_{p(s-1)}^2$, that is, $-2 \ln \lambda_2$ asymptotically converges to a real scalar chisquare having $p(s-1)$ degrees of freedom, where $\lambda_2 = |U_2|$ with U_2 as defined in Exercise 13.8. [Hint: Look at the h -th moment of λ_2]. For large values of rst , one can utilize this approximate chisquare distribution for testing the hypothesis $B_1 = \cdots = B_s = O$.

References

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai and H.J. Haubold (2017): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 14

Profile Analysis and Growth Curves



14.1. Introduction

We will utilize the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

14.1.1. Profiles

The two topics that are treated in this chapter are profile analysis and growth curves. We first consider profile analysis. Only three types of profiles will be considered. As an example, take the amount of money spent at a specific grocery store by six families residing in a given neighborhood. Suppose that they all do their grocery shopping once a week every Saturday. Let us consider the sum spent on the average in the long run on four types of items: item (1): vegetables; item (2): baked goods; item (3): meat products; item (4): cleaning supplies. Let μ_{ij} denote the expected amount or the average amount spent in the long run by the j -th family on the i -th item, $j = 1, 2, 3, 4, 5, 6$, and $i = 1, 2, 3, 4$.

The monetary values of these purchases are plotted as points in Fig. 14.1.1. In order to see the pattern, these points are joined by straight lines, which are not actually part of the graph. From the pattern or profile, we may note that families F_1, F_2, F_3, F_4 have parallel profiles, that the profiles of F_1 and F_4 coincide, and that F_5 has a constant profile or is horizontal with respect to the base line profile. The profile of F_6 does not fall into any category relative to the other profiles.

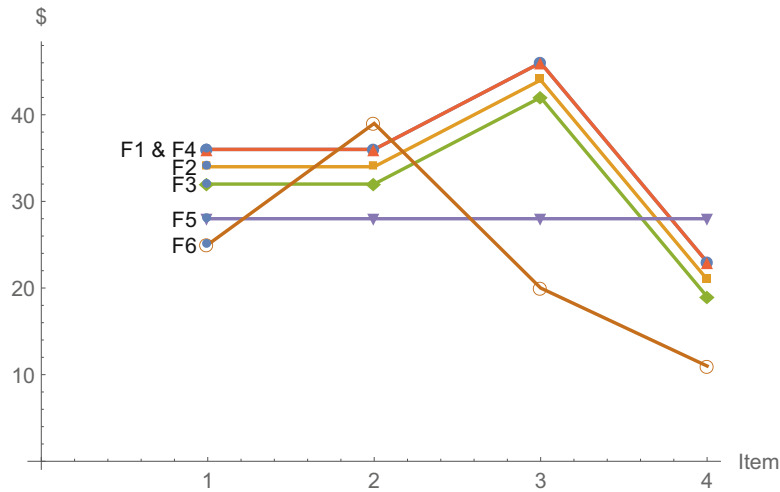


Figure 14.1.1 Profile plot for 6 families: amounts spent weekly on 4 items

In this instance, we refer to the profiles of F_1, F_2, F_3, F_4 as “parallel profiles” where the profiles of F_1 and F_4 are “coincident profiles”. F_5 is said to have a constant profile or “level profile”. We now study these patterns in more details starting with parallel profiles.

14.2. Parallel Profiles

Let μ_{ij} be the expected dollar amount of the purchase of the j -th family on the i -th item. In general, if the items are denoted by the real scalar variables x_{1j}, \dots, x_{pj} , then $\mu_{ij} = E[x_{ij}]$, $i = 1, \dots, p$ and $j = 1, \dots, k$. Thus, we have k , p -variate populations. The items or variables may be quantitative or qualitative. The p items may be p different stimulants and the response may be the expected enhancement in the performance of a sportsman; the p items could also be p different animal feeds and the response could then be the expected weight increase, and so on. The μ_{ij} 's, that is, the expected responses, may be quantitative, or qualitative but convertible to quantitative readings. If the items are

various color combinations in a lady’s dress, the response of the first observer F_1 with respect to a certain color combination may be “very pleasing”, the response of the second observer F_2 may be “indifferent”, etc. These assessments of color combinations can, for example, be quantified as readings on a 0 to 10 scale.

Let us start with $k = 2$, that is, two p -variate populations, and let the responses be

$$M_1 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix} \quad \text{and} \quad M_2 = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix}.$$

Then, what is the meaning of parallel profiles? We can impose the conditions $\mu_{12} - \mu_{11} = \mu_{22} - \mu_{21}$, $\mu_{22} - \mu_{21} = \mu_{32} - \mu_{31}$, etc., which is tantamount to saying that

$$\mu_{j-12} - \mu_{j-11} = \mu_{j2} - \mu_{j1} \Rightarrow \mu_{j1} - \mu_{j-11} = \mu_{j2} - \mu_{j-12}, \quad \text{for } j = 2, \dots, p. \tag{14.2.1}$$

The conditions specified in (14.2.1) can be expressed in matrix notation. Letting C be the following $(p - 1) \times p$ matrix:

$$C = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & -1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \Rightarrow CM_1 = \begin{bmatrix} -\mu_{11} + \mu_{21} \\ -\mu_{21} + \mu_{31} \\ \vdots \\ -\mu_{p-11} + \mu_{p1} \end{bmatrix}, \quad CM_2 = \begin{bmatrix} -\mu_{12} + \mu_{22} \\ -\mu_{22} + \mu_{32} \\ \vdots \\ -\mu_{p-12} + \mu_{p2} \end{bmatrix} \tag{14.2.2}$$

Then, $CM_1 = CM_2$. Therefore, “parallel profiles” means $CM_1 = CM_2$, and we can test the hypothesis $CM_1 = CM_2$ against $CM_1 \neq CM_2$ for determining whether the profiles are parallel or not. In order to avoid any possible confusion, let us denote the two vectors as X and Y , and let $M_1 = E[X]$ and $M_2 = E[Y]$, that is,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, \quad E[X] = M_1 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix}, \quad E[Y] = M_2 = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix}. \tag{14.2.3}$$

For testing hypotheses, we need to assume some distributions for X and Y . Let X and Y be independently distributed real Gaussian p -variate variables where $X \sim N_p(M_1, \Sigma)$ and $Y \sim N_p(M_2, \Sigma)$, $\Sigma > O$. Observe that the $(p - 1) \times p$ matrix C is of full rank $p - 1$. Then, on applying a result previously obtained in Chap. 3,

$$CX \sim N_{p-1}(CM_1, C\Sigma C') \quad \text{and} \quad CY \sim N_{p-1}(CM_2, C\Sigma C'). \tag{14.2.4}$$

Assume that we have a simple random sample of size n_1 from X and a simple random sample of size n_2 from Y . Let X_1, \dots, X_{n_1} be the sample values from X and let Y_1, \dots, Y_{n_2} be the sample values from Y . Let the sample averages be \bar{X} and \bar{Y} , the sample matrices be the boldfaced \mathbf{X} and \mathbf{Y} , the matrices of sample means be $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$, the deviation matrices be \mathbf{X}_d and \mathbf{Y}_d and the sample sum of products matrices be $S_1 = \mathbf{X}_d \mathbf{X}'_d$ and $S_2 = \mathbf{Y}_d \mathbf{Y}'_d$. Then, these are the following:

$$\mathbf{X} = [X_1, \dots, X_{n_1}] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n_1} \\ x_{21} & x_{22} & \dots & x_{2n_1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn_1} \end{bmatrix},$$

$$\bar{\mathbf{X}} = \begin{bmatrix} \sum_{j=1}^{n_1} x_{1j}/n_1 \\ \sum_{j=1}^{n_1} x_{2j}/n_1 \\ \vdots \\ \sum_{j=1}^{n_1} x_{pj}/n_1 \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix},$$

$$\bar{\mathbf{X}} = [\bar{X}, \bar{X}, \dots, \bar{X}], \mathbf{X}_d = \mathbf{X} - \bar{\mathbf{X}}, S_1 = \mathbf{X}_d \mathbf{X}'_d,$$

with similar expressions in the case of Y for which the sum of squares and cross products matrix is denoted as $S_2 = \mathbf{Y}_d \mathbf{Y}'_d$. Note that $\bar{\mathbf{X}} \sim N_p(M_1, \frac{1}{n_1} \Sigma)$, $\Sigma > O \Rightarrow C\bar{\mathbf{X}} \sim N_{p-1}(CM_1, \frac{1}{n_1} C\Sigma C')$. As well, $C\bar{\mathbf{Y}} \sim N_{p-1}(CM_2, \frac{1}{n_2} C\Sigma C')$. The sum of squares and cross product matrices of CX and CY are CS_1C' and CS_2C' , respectively. Since X and Y are independently distributed, we have

$$\begin{aligned} \bar{\mathbf{X}} - \bar{\mathbf{Y}} &\sim N_p(M_1 - M_2, (\frac{1}{n_1} + \frac{1}{n_2}) \Sigma) \Rightarrow \\ C\bar{\mathbf{X}} - C\bar{\mathbf{Y}} &\sim N_{p-1}(C(M_1 - M_2), (\frac{1}{n_1} + \frac{1}{n_2}) C\Sigma C'). \end{aligned} \quad (14.2.5)$$

An unbiased estimator for Σ is $\hat{\Sigma} = \frac{1}{n_1+n_2-2}(S_1 + S_2)$ and then, Hotelling's T^2 statistic is

$$T^2 = (\frac{1}{n_1} + \frac{1}{n_2})^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})' C' [(C\hat{\Sigma}C')^{-1}] C (\bar{\mathbf{X}} - \bar{\mathbf{Y}}). \quad (14.2.6)$$

The usual test is based on Hotelling's T^2 statistic. However, referring to Sects. 6.3.1 and 6.3.3, we are making use of a statistic based on a real type-2 beta random variable. Thus, for $n_1 + n_2 - p = (n_1 + n_2 - 2) + 1 - (p - 1)$, and under the hypothesis $H_{o1} : CM_1 = CM_2$,

$$w = [\frac{n_1+n_2-p}{p-1}] (\frac{1}{n_1} + \frac{1}{n_2})^{-1} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})' C' (C(S_1 + S_2)C')^{-1} C (\bar{\mathbf{X}} - \bar{\mathbf{Y}}) \sim F_{p-1, n_1+n_2-p}. \quad (14.2.7)$$

Accordingly, for testing the hypothesis $H_{o1} : CM_1 = CM_2$ versus $CM_1 \neq CM_2$, we reject H_{o1} if the observed value of w of (14.2.7) is greater than or equal to $F_{p-1, n_1+n_2-p, \alpha}$, the upper 100 $\alpha\%$ percentage point from an F_{p-1, n_1+n_2-p} distribution.

Example 14.2.1. The following are independent observed samples of sizes $n_1 = 4$ and $n_2 = 5$ from trivariate Gaussian populations sharing the same covariance matrix $\Sigma > O$. Test the hypothesis $H_o : CM_1 = CM_2$ at the 5% significance level:

$$\text{Observed sample-1: } X_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, X_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, X_4 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix};$$

$$\text{Observed sample-2: } Y_1 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}, Y_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, Y_3 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, Y_4 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, Y_5 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

Solution 14.2.1. Let us first evaluate the 2×3 matrix C as well as \bar{X} , \bar{Y} , \mathbf{X} , $\bar{\mathbf{X}}$, $\bar{\mathbf{Y}}$, \mathbf{X}_d , \mathbf{Y}_d , S_1 and S_2 :

$$C = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \bar{X} = \frac{1}{4} \sum_{j=1}^4 X_j = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

$$\bar{Y} = \frac{1}{5} \sum_{j=1}^5 Y_j = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ -1 & 1 & 2 & 2 \end{bmatrix}, \bar{\mathbf{X}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

$$\mathbf{X}_d = \mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ -2 & 0 & 1 & 1 \end{bmatrix}, S_1 = \mathbf{X}_d \mathbf{X}_d' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 6 \end{bmatrix};$$

$$\mathbf{Y} = \begin{bmatrix} 0 & 1 & -1 & 1 & -1 \\ -1 & 1 & 2 & 2 & 1 \\ -1 & 2 & 1 & -2 & 0 \end{bmatrix}, \mathbf{Y}_d = \begin{bmatrix} 0 & 1 & -1 & 1 & -1 \\ -2 & 0 & 1 & 1 & 0 \\ -1 & 2 & 1 & -2 & 0 \end{bmatrix},$$

$$S_2 = \mathbf{Y}_d \mathbf{Y}_d' = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 6 & 1 \\ -1 & 1 & 10 \end{bmatrix}, S_1 + S_2 = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 8 & 0 \\ -1 & 0 & 16 \end{bmatrix}, C(S_1 + S_2)C' = \begin{bmatrix} 12 & -7 \\ -7 & 24 \end{bmatrix},$$

$$[C(S_1 + S_2)C']^{-1} = \frac{1}{239} \begin{bmatrix} 24 & 7 \\ 7 & 12 \end{bmatrix}, \quad C'[C(S_1 + S_2)C']^{-1}C = \frac{1}{239} \begin{bmatrix} 24 & -17 & -7 \\ -17 & 22 & -5 \\ -7 & -5 & 12 \end{bmatrix},$$

$$\bar{X} - \bar{Y} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \quad \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} = \frac{20}{9}.$$

Therefore,

$$\begin{aligned} w &= \frac{(n_1 + n_2 - p)}{p - 1} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} (\bar{X} - \bar{Y})' C' [C(S_1 + S_2)C']^{-1} C (\bar{X} - \bar{Y}) \\ &= \frac{6}{2} \left(\frac{20}{(9)(239)}\right) [1, -1, 1] \begin{bmatrix} 24 & -17 & -7 \\ -17 & 22 & -5 \\ -7 & -5 & 12 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \\ &= 3 \left(\frac{20}{(9)(239)}\right) 88 = 2.45. \end{aligned}$$

From the tabulated values for $\alpha = 0.05$,

$$F_{p-1, n_1+n_2-p, \alpha} = F_{2, 6, 0.05} = 5.14 > 2.45.$$

Hence, the hypothesis $H_{o1} : CM_1 = CM_2$ is not rejected at the 5% significance level.

14.3. Coincident Profiles

With the notations employed in Sect. 14.2, two profiles are coincident if $\mu_{i1} = \mu_{i2}$, $i = 1, \dots, p$, which amounts to assuming $H_{o2} : M_1 = M_2$. If the $p \times 1$ real vectors X and Y are $X \sim N_p(M_1, \Sigma)$ and $Y \sim N_p(M_2, \Sigma)$, $\Sigma > O$, and the populations are independently distributed, this is the test for equality of mean value vectors in two independent Gaussian populations with common unknown covariance matrix, which was discussed in Chap. 6 where a Hotelling's T^2 is usually utilized for testing this hypothesis. We will use a statistic based on a real scalar type-2 beta variable, which has already been considered in Sects. 6.3.1 and 6.3.3. Given samples of sizes n_1 and n_2 from these normal populations, and letting \bar{X} and \bar{Y} be the sample means and S_1 and S_2 be the sample sum of squares and cross products matrices, the test statistic under the hypothesis H_{o2} , denoted by w_1 , is the following:

$$w_1 = \left[\frac{n_1 + n_2 - 1 - p}{p} \right] \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\bar{X} - \bar{Y})' (S_1 + S_2)^{-1} (\bar{X} - \bar{Y}) \sim F_{p, n_1+n_2-1-p}. \quad (14.3.1)$$

Thus, we reject the hypothesis of equality of population mean values or the hypothesis $H_{02} : M_1 = M_2$, whenever the observed value of w_1 is greater than or equal to $F_{p, n_1+n_2-1-p, \alpha}$, the upper $100\alpha\%$ percentage point of an F -distribution having p and $n_1 + n_2 - 1 - p$ degrees of freedom.

Example 14.3.1. With the data provided in Example 14.2.1 and under the same assumptions, test the hypothesis that the two profiles are coincident, that is, test $H_{02} : M_1 = M_2$.

Solution 14.3.1. From Solution 14.2.1, we have: $n_1 = 4$, $n_2 = 5$,

$$\bar{X} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \bar{Y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, S_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 6 \end{bmatrix}, S_2 = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 6 & 1 \\ -1 & 1 & 10 \end{bmatrix},$$

so that

$$S_1 + S_2 = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 8 & 0 \\ -1 & 0 & 16 \end{bmatrix}, (S_1 + S_2)^{-1} = \frac{1}{504} \begin{bmatrix} 128 & 0 & 8 \\ 0 & 63 & 0 \\ 8 & 0 & 32 \end{bmatrix}, p = 3, n_1 + n_2 - 1 - p = 5.$$

Then, an observed value of the test statistic, again denoted by w_1 , is the following:

$$\begin{aligned} w_1 &= \frac{(n_1 + n_2 - 1 - p)}{p} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} (\bar{X} - \bar{Y})' (S_1 + S_2)^{-1} (\bar{X} - \bar{Y}) \\ &= \left(\frac{5}{3} \right) \left(\frac{20}{9} \right) \frac{1}{504} [1, -1, 1] \begin{bmatrix} 128 & 0 & 8 \\ 0 & 63 & 0 \\ 8 & 0 & 32 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \\ &= \left(\frac{5}{3} \right) \left(\frac{20}{9} \right) \left(\frac{239}{504} \right) = \frac{23900}{13608} = 1.76. \end{aligned}$$

The tabulated value of $F_{p, n_1+n_2-1-p, \alpha} = F_{3, 5, 0.05}$ being $5.41 > 1.76$, the hypothesis $M_1 = M_2$ against the natural alternative $M_1 \neq M_2$ is not rejected at the 5% level of significance.

14.3.1. Conditional hypothesis for coincident profiles

If we already know that the two profiles are parallel, then the second profile is either above the first one or below the first one, and these two profiles are equal only if the sum of the elements in M_1 is equal to the sum of the elements in M_2 . There should be caution in making use of this argument. We must know that the two profiles are assuredly parallel. If the hypothesis that two profiles are parallel is not rejected, this does not necessarily imply

that the two profiles are parallel. It only entails that the departure from parallelism is not significant. Accordingly, a test for coincidence that is based on the sum of the elements is not justifiable, and the test being herein discussed is not a test for coincidence but only a test of the hypothesis that the sum of the elements in M_1 is equal to the sum of the elements in M_2 . Observe that this equality can hold whether the profiles are coincident or not. The sum of the elements in M_1 is given by $J'M_1$ and the sum of the elements in M_2 is given by $J'M_2$ where $J' = (1, 1, \dots, 1)$. Under independence and normality with $X \sim N_p(M_1, \Sigma)$ and $Y \sim N_p(M_2, \Sigma)$, $\Sigma > O$, we have $J'\bar{X} \sim N_1(J'M_1, \frac{1}{n_1}J'\Sigma J)$ and $J'\bar{Y} \sim N_1(J'M_2, \frac{1}{n_2}J'\Sigma J)$, and hence,

$$t^2 = (n_1 + n_2 - 2) \left(\frac{n_1 n_2}{n_1 + n_2} \right) [J'(\bar{X} - \bar{Y}) - J'(M_1 - M_2)] \\ \times [J'(S_1 + S_2)J]^{-1} [J'(\bar{X} - \bar{Y}) - J'(M_1 - M_2)] \quad (14.3.2)$$

where, under the hypothesis $H_{03} : J'M_1 - J'M_2 = 0$, t is a Student- t having $n_1 + n_2 - 2$ degrees of freedom and $t^2 \sim F_{1, n_1 + n_2 - 2}$. Thus, we may utilize a Student- t statistic with $n_1 + n_2 - 2$ degrees of freedom or an F -statistic with 1 and $n_1 + n_2 - 2$ degrees of freedom in order to reach a decision with respect to the hypothesis of the equality of the sum of the elements in M_1 and M_2 .

Example 14.3.2. With the data provided in Example 14.2.1, test the hypothesis $J'M_1 = J'M_2$ or the equality of the sum of elements in M_1 and M_2 .

Solution 14.3.2. In this case, $n_1 + n_2 - 2 = 7$, $\frac{n_1 n_2}{n_1 + n_2} = \frac{9}{20}$,

$$J'(\bar{X} - \bar{Y}) = [1, 1, 1] \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 1, \quad J'(S_1 + S_2)J = [1, 1, 1] \begin{bmatrix} 4 & 0 & -1 \\ 0 & 8 & 0 \\ -1 & 0 & 16 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 26.$$

Hence, the observed value of t^2 in (14.3.2) is

$$t^2 = 7 \left(\frac{9}{20} \right) \left(\frac{1}{26} \right) = 0.121 \text{ while } F_{1,7,0.05} = 5.596.61 > 0.121.$$

Thus, the hypothesis of equality of the sum of the elements in M_1 and M_2 is not rejected at the 5% significance level.

14.4. Level Profiles

A level profile occurs when the plot is horizontal with respect to the base line or when all the elements in M_1 are equal. Thus, this hypothesis involves only one population. If

there are two populations, we can consider the hypothesis of the mean value vectors M_1 and M_2 being equal or testing whether the sum of the elements in each of M_1 and M_2 are equal. When the null hypothesis is not rejected, the sum in M_1 may be different from the sum in M_2 . However, if we assume the profiles to be coincident, these two sums will be equal. We may then create a conditional hypothesis to the effect that the profiles in the two populations are level, given that the profiles are coincident. This is how one can connect the two populations and test whether the profiles are level. Let the $p \times 1$ vector $X_j \sim N_p(M, \Sigma)$, $\Sigma > O$, and let X_1, \dots, X_n be iid as X_j . In this instance, we have one real p -variate Gaussian population and a simple random sample of size n with $E[X_j] = M$ and $\text{Cov}(X_j) = \Sigma$, $j = 1, \dots, n$. Let

$$M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{bmatrix}, \bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n x_{1j}/n \\ \sum_{j=1}^n x_{2j}/n \\ \vdots \\ \sum_{j=1}^n x_{pj}/n \end{bmatrix}.$$

The hypothesis of “level profile” means $H_{o3} : \mu_1 = \mu_2 = \dots = \mu_p = \mu$ for some μ . The common value μ is estimated by $\hat{\mu} = \frac{1}{np} \sum_{i=1}^p \sum_{j=1}^n x_{ij}$, which is the maximum likelihood estimator/estimate. Then, under the hypothesis, \hat{M} is such that $\hat{M}' = [\hat{\mu}, \hat{\mu}, \dots, \hat{\mu}]$. Testing such a hypothesis has already been considered in Chap. 6, the test being based on the statistic

$$u = n(\bar{X} - \hat{M})' S^{-1} (\bar{X} - \hat{M}) \sim \text{type-2 beta} \tag{14.4.1}$$

with the parameters $(\frac{p-1}{2}, \frac{n-p+1}{2})$, under the hypothesis H_{o3} , referring more specifically to the derivations related to Eq. (6.3.14). Thus under H_{o3} ,

$$\frac{(n-p+1)}{p-1} u \sim F_{p-1, n-p+1} \tag{14.4.2}$$

where $F_{p-1, n-p+1}$ is a real scalar F -random variable with $p-1$ and $n-p+1$ degrees of freedoms. We reject H_{o3} if the observed value of the $F_{p-1, n-p+1}$ as specified in (14.4.2) is greater than or equal to $F_{p-1, n-p+1, \alpha}$, the upper 100 $\alpha\%$ percentage point of an $F_{p-1, n-p+1}$ distribution. This test has already been illustrated in Chap. 6.

14.4.1. Level profile, two independent Gaussian populations

We can make use of (14.4.1) and test for level profile in each of the two populations separately. Let the populations and simple random samples from these populations be as

follows: $X_j \stackrel{iid}{\sim} N_p(M_1, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n_1$, and $Y_j \stackrel{iid}{\sim} N_p(M_2, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n_2$, where Σ is identical for both of these independent Gaussian populations, and let

$$E[X_j] = M_1 = \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix}, \quad j = 1, \dots, n_1; \quad E[Y_j] = M_2 = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix}, \quad j = 1, \dots, n_2.$$

When population X has a level profile, $\mu_{11} = \mu_{21} = \dots = \mu_{p1} = \mu^{(1)}$ and when population Y has a level profile, $\mu_{12} = \mu_{22} = \dots = \mu_{p2} = \mu^{(2)}$, $\mu^{(1)}$ need not be equal to $\mu^{(2)}$. However, if we assume that the two populations have coincident profiles, then $\mu^{(1)} = \mu^{(2)}$. In this instance, the two populations are identical, so that we can pool the samples X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} and regard the combined sample as a simple random sample of size $n_1 + n_2$ from the population $N_p(M, \Sigma)$, $\Sigma > O$. Thus, we can apply (14.4.1) and (14.4.2) with n is replaced by $n_1 + n_2$, in order to reach a decision. Instead of letting $\mu^{(1)} = \mu^{(2)} = \mu$ for some μ and computing the estimate for μ or $\hat{\mu}$ from the combined sample and utilizing \hat{M} , $\hat{M}' = [\hat{\mu}, \dots, \hat{\mu}]$ in (14.4.1), we can make use of the matrix C as defined in Sect. 14.2 and consider the transformed vector CZ_j , $Z_j \sim N_p(M, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n_1 + n_2$, and use $C\bar{Z}$, $\bar{Z} = \frac{1}{n_1+n_2}[X_1 + \dots + X_{n_1} + Y_1 + \dots + Y_{n_2}]$ in (14.4.1). One advantage in utilizing the matrix C is that $C\hat{M} = O$, and from (13.4.1), this \hat{M} will vanish, that is, $\bar{X} - \hat{M}$ in (14.4.1) will become $C\bar{Z}$. Then, the test statistic under the hypothesis $H_{o3} : C'M = O$, again denoted by u , becomes the following:

$$\begin{aligned} u &= (n_1 + n_2)(C\bar{Z} - O)'(CS_zC')^{-1}(C\bar{Z} - O) \\ &= (n_1 + n_2)\bar{Z}'C'[(CS_zC')^{-1}]C\bar{Z} \sim \text{type-2 beta} \end{aligned} \quad (14.4.3)$$

with the parameters $(\frac{p-1}{2}, \frac{n_1+n_2-p+1}{2})$, under the hypothesis H_{o3} , where S_z is the sample sum of squares and cross products matrix from the combined sample. Then under H_{o3} ,

$$\frac{(n_1 + n_2 - p + 1)}{p - 1} u \sim F_{p-1, n_1+n_2-p+1}. \quad (14.4.4)$$

We reject the hypothesis that the profiles in the two populations are level, given that the two populations have coincident profiles, if the observed value of the F distributed statistic given in (14.4.4) is greater than or equal to $F_{p-1, n_1+n_2-p+1, \alpha}$, the upper 100 $\alpha\%$ percentage point of a real scalar F_{p-1, n_1+n_2-p+1} distribution.

Example 14.4.1. With the data provided in Example 14.2.1, test the hypothesis of level profile, given that the two populations have coincident profiles.

Solution 14.4.1. In this case, we pool the observed samples: $n_1 + n_2 = 4 + 5 = 9$,

$$\sum_{j=1}^4 X_j + \sum_{j=1}^5 Y_j = \begin{bmatrix} 4 \\ 5 \\ 4 \end{bmatrix} \Rightarrow \bar{Z} = \frac{1}{9} \begin{bmatrix} 4 \\ 5 \\ 4 \end{bmatrix}.$$

Denoting the pooled sample by a boldfaced \mathbf{Z} , the matrix of pooled sample means by $\bar{\mathbf{Z}}$, the deviation matrix by $\mathbf{Z}_d = \mathbf{Z} - \bar{\mathbf{Z}}$ and the sample sum of products matrix by $S_z = \mathbf{Z}_d \mathbf{Z}'_d$, we have

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 0 & -1 & 1 & 2 & 2 & 1 \\ -1 & 1 & 2 & 2 & -1 & 2 & 1 & -2 & 0 \end{bmatrix},$$

$$\mathbf{Z}_d = \frac{1}{9} \begin{bmatrix} 5 & 5 & 5 & 5 & -4 & 5 & -13 & 5 & -13 \\ -5 & 4 & -14 & -5 & -14 & 4 & 13 & 13 & 4 \\ -13 & 5 & 14 & 14 & -13 & 14 & 5 & -22 & -4 \end{bmatrix},$$

$$\mathbf{Z}_d \mathbf{Z}'_d = \frac{1}{9^2} \begin{bmatrix} 504 & -180 & 99 \\ -180 & 828 & -180 \\ 99 & -180 & 1476 \end{bmatrix} = S_z, \quad C = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

$$C S_z C' = \frac{1}{9^2} \begin{bmatrix} 1692 & -1287 \\ -1287 & 2664 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 188 & -143 \\ -143 & 296 \end{bmatrix},$$

$$(C S_z C')^{-1} = \frac{9}{35199} \begin{bmatrix} 296 & 143 \\ 143 & 188 \end{bmatrix}, \quad C'(C S_z C')^{-1} C = \frac{9}{35199} \begin{bmatrix} 296 & -153 & -143 \\ -153 & 198 & -45 \\ -143 & -45 & 188 \end{bmatrix},$$

$$u = \frac{n_1 + n_2 - p + 1}{p - 1} (n_1 + n_2) \bar{\mathbf{Z}}' C' (C S_z C')^{-1} C \bar{\mathbf{Z}},$$

$$= \left(\frac{7}{2}\right) 9 \left(\frac{9}{35199}\right) \left(\frac{1}{9^2}\right) [4, 5, 4] \begin{bmatrix} 296 & -153 & -143 \\ -153 & 198 & -45 \\ -143 & -45 & 188 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 4 \end{bmatrix} = 0.0197.$$

Since, under the null hypothesis H_{03} , $F_{p-1, n_1+n_2-p+1} = F_{2,7}$ and from the tabulated values for $\alpha = 0.05$, we have $F_{2,7,0.05} = 4.74$ which is larger than 0.0197, the hypothesis H_{03} of level profile in both the populations, given that the population profiles are coincident, is not rejected at 5% level of significance.

14.5. Generalizations to k Populations

Let the $p \times 1$ real vectors X_1, \dots, X_k be independently distributed where $X_j \sim N_p(M_j, \Sigma)$, $\Sigma > O$, $j = 1, \dots, k$, Σ being the same for all the populations. Consider the $(p-1) \times p$ array C , as previously defined in Sect. 14.2, whose rank is $p-1$. Letting $Z_j = CX_j$, $j = 1, \dots, k$, $Z_j \stackrel{iid}{\sim} N_{p-1}(CM_j, C\Sigma C')$, $(C\Sigma C) > O$, $j = 1, \dots, k$. Then, testing for “parallel profiles” in these k populations is equivalent to testing the hypothesis

$$H_{o1} : CM_1 = CM_2 = \dots = CM_k. \quad (14.5.1)$$

Testing for “coincident profiles” in all the k populations is tantamount to testing the hypothesis

$$H_{o2} : M_1 = M_2 = \dots = M_k. \quad (14.5.2)$$

Testing for “level profiles”, given that the profiles are coincident, in all these k populations, is the same as testing the hypothesis

$$H_{o3} : CM_1 = CM_2 = \dots = CM_k = O. \quad (14.5.3)$$

Note that H_{o1} , H_{o2} , H_{o3} are all tests involving equality of mean value vectors in multivariate populations where, in H_{o1} and H_{o3} , we have $(p-1)$ -variate Gaussian populations and, in H_{o2} , we have a p -variate Gaussian population. In all these cases, the population covariance matrices are identical but unknown. Such tests were considered in Chap. 6 and hence, further discussion is omitted. In all these tests, one point has to be kept in mind. The units of measurements must be the same in all the populations, otherwise there is no point in comparing mean value vectors. If qualitative variables are involved, they should be converted to quantitative variables using the same scale, such as changing all of them to points on a $[0, 10]$ scale.

14.6. Growth Curves or Repeated Measures Designs

For instance, suppose that the growth (measured in terms of height) of a plant seedling (for instance, a variety of corn) is monitored. Let the height be 10 cm at the onset of the observation period, that is, at time $t = 0$, and the measurements be taken weekly. After one week ($t = 1$), suppose that the height is 12 cm. Then $10 = a_0$, and, for a linear function, $12 = a_0 + a_1t = a_0 + a_1$ at $t = 1$, so that $a_0 = 10$ and $a_1 = 2$. We now have the points $(0, 10)$, $(1, 12)$. Consider a straight line passing through these points. If the third, fourth, etc. measurements taken at $t = 2$, $t = 3$, etc. weeks, fall more or less on this line, then we can say that the expected growth is linear and we can consider a model of the type $E[x] = a_0 + a_1t$ to represent the expected growth of this plant, where $E[\cdot]$ represents the expected value. If the third, fourth, etc. points are increasing away from the

straight line and the behavior is of a parabolic type then, we can put forward the model $E[x] = a_0 + a_1t + a_2t^2$. If the behavior appears to be of a polynomial type, then we may posit the model $E[x] = a_0 + a_1t + \cdots + a_{m-1}t^{m-1}$. However, we need not monitor the growth at one time unit intervals. We may monitor it at convenient time units t_1, t_2, \dots . A polynomial type model leads to the following equations:

$$\begin{aligned} E[x_1] &= a_0 + a_1t_1 + a_2t_1^2 + \cdots + a_{m-1}t_1^{m-1} \\ E[x_2] &= a_0 + a_1t_2 + a_2t_2^2 + \cdots + a_{m-1}t_2^{m-1} \\ &\vdots \\ E[x_p] &= a_0 + a_1t_p + a_2t_p^2 + \cdots + a_{m-1}t_p^{m-1} \end{aligned}$$

where $E[x_1], \dots, E[x_p]$ are expected observations at times t_1, t_2, \dots, t_p , and a_0 is the value of the expected observation at $t_j = 0$, $j = 1, \dots, p$. Thus, in matrix notation, we have a p -vector for the observations. Letting $X'_1 = [x_1, x_2, \dots, x_p]$,

$$E[X_1] = E \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & t_p^2 & \cdots & t_p^{m-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{m-1} \end{bmatrix} \equiv T A_1, \quad (14.6.1)$$

where X_1 is $p \times 1$, T is $p \times m$ and the parameter vector A_1 is $m \times 1$. In this instance, we have taken observations on the same real scalar variable x at p different occasions t_1, \dots, t_p . In this sense, it is a “repeated measures design” situation. We have monitored x over time in order to characterize the growth of this variable and, in this sense, it is a “growth curve” situation. This x may be the weight of a person under a certain weight reduction diet. The successive readings are then expected to decrease, so that it is a negative growth as well as a repeated measures design model. When x represents the spread area of a certain fungus, which is monitored over time, we have a growth curve model that depends upon the nature of spread. If x is the temperature reading of a feverish person under a certain medication, being recorded over time, we have a growth curve model with an expected negative growth. We can think of a myriad of such situations falling under repeated measures design or growth curve model.

14.6.1. Growth curve model

The model specified in (14.6.1) can be expressed as follows:

$$E[X_1] = T A_1, \quad A_1 = \begin{bmatrix} a_{10} \\ \vdots \\ a_{1m-1} \end{bmatrix}, \quad T = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & t_p^2 & \dots & t_p^{m-1} \end{bmatrix}, \quad (14.6.2)$$

where we have a p -vector X_1 with expected value $T A_1$, T being $p \times m$ and A_1 , $m \times 1$. If we are taking a simple random sample of size n_1 from this p -vector, we may denote the sample values as $X_{11}, X_{12}, \dots, X_{1n_1}$ and we may say that we have a simple random sample of size n_1 from that p -variate population or from n_1 members belonging to a group whose underlying distribution is that of X_1 , and we may write the model as

$$E[X_{1j}] = T A_1, \quad j = 1, \dots, n_1. \quad (14.6.3)$$

With reference to the weight measurements of a person subjected to a certain diet, if the diet is administered to a group of n_1 individuals who are comparable in every respect such as initial weight, age, physical condition, and so on, then the measurements on these n_1 individuals are the observations X_{11}, \dots, X_{1n_1} . Letting the same diet be administered to another group of n_2 persons whose characteristics are similar within the group but may be different from those of the first group, their measurements will be denoted by X_{2j} , $j = 1, \dots, n_2$. Consider r such groups. If these groups are men belonging to a certain age bracket and women of the same age category, $r = 2$. If there are 7 groups of women of different age categories and 5 groups of men belonging to certain age brackets in the study, r will be equal to 12, and so on. Thus, there can be variations between groups but each group ought to be homogeneous. If there are r such groups and a simple random sample of size n_i is available from group i , $i = 1, \dots, r$, we have the model

$$E[X_{ij}] = T A_i, \quad j = 1, \dots, n_i, \quad i = 1, \dots, r. \quad (14.6.4)$$

We may then compare this situation with a one-way classification MANOVA (or ANOVA when a single dependent variable is being monitored) situation involving r groups with the i -th group being of size n_i , so that $n = n_1 + n_2 + \dots + n_r = n$, as per our notation with respect to the one-way layout. There is one set of treatments and we wish to assess the expected effects in these r groups. For the analysis of the data being so modeled, we do not explicitly require a subscript to represent the order p of the observation vectors (in this case, p corresponds to the number of occasions the measurements are made) and hence, our notation (14.6.4) does not involve a subscript representing the order of the vector X_{ij} .

The model (14.6.4) can be generalized in various ways. As a second set of treatments, we may consider s different diets being administered to r distinct groups of n individuals each. “Diet” may exhibit another polynomial growth of the type $b_0 + b_1t + \cdots + b_{m_1}t^{m_1}$, where m_1 need not be equal to $m - 1$. Then, within the framework of experimental designs, we have a multivariate two-way layout with an equal number of observations per cell. In this manner, we can obtain multivariate generalizations of all standard designs where the effects are polynomial growth models, possibly of different degrees, with p , however, remaining unchanged, p being the number of occasions the measurements are made. We can also consider another type of generalization. In the example involving different types of diets, these various types may be the same diet but in different strictness levels. In this instance, the different diets are correlated among themselves. In a two-way layout, there is no correlation within one set of treatments; interaction may be present between the two sets of treatments. When different doses of the same diet is the second set of treatments, correlation is obviously present within this second set of treatments. This leads to multivariate factorial designs with growth curve model for the treatment effects. In all of the above considerations, we were only observing one item, namely weight. In another type of generalization, instead of one item, several items are measured at the same time, that is, there may be several dependent (response) variables instead of one. In this case, if y_1, \dots, y_k are the items measured on the same p occasions, then each y_j may have its own distribution with covariance matrix Σ_j , $j = 1, \dots, k$. Not only that, there may also be joint dispersions, in which case, there will be covariance matrices Σ_{ij} for all $i, j = 1, \dots, k$. Thus, several types of generalizations are possible for the basic model specified in (14.6.4).

In order to analyze the data with the model (14.6.4), the following assumptions are required:

- (1) The $n_i = n$ vectors X_{ij} 's are mutually independently distributed for all i and j with common covariance matrix $\Sigma > O$.
- (2) $E[X_{ij}] = TA_i$, $j = 1, \dots, n_i$, for each i , $i = 1, \dots, r$.
- (3) $X_{ij} \sim N_p(TA_i, \Sigma)$, $\Sigma > O$, $j = 1, \dots, n_i$, for each i , $i = 1, \dots, r$.

The normality assumption is convenient for obtaining maximum likelihood estimates (MLE) and for testing hypotheses on the parameters A_i , $i = 1, \dots, r$. Observe that the matrix T is known.

14.6.2. Maximum likelihood estimation

Under the normality assumption for the X_{ij} 's, their joint density, denoted by L , is

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i,j} (X_{ij} - TA_i)' \Sigma^{-1} (X_{ij} - TA_i)}. \quad (14.6.5)$$

Note that

$$\sum_{i,j} (X_{ij} - TA_i)' \Sigma^{-1} (X_{ij} - TA_i) = \sum_{i,j} \text{tr}[\Sigma^{-1} (X_{ij} - TA_i)(X_{ij} - TA_i)'].$$

Let $M_i = TA_i$. Then, differentiating $\ln L$ with respect to M_i and equating the result to a null vector yields the following (for vector/matrix derivatives, refer to Chap. 1):

$$\begin{aligned} \frac{\partial}{\partial M_i} \sum_{i,j} (X_{ij} - M_i)' \Sigma^{-1} (X_{ij} - M_i) &= O \Rightarrow \\ -2 \sum_j [\Sigma^{-1} (X_{ij} - M_i)] &= O \Rightarrow \\ \sum_j (X_{ij} - M_i) &= O \Rightarrow M_i = \frac{1}{n} \sum_j X_{ij} = \bar{X}_i. \end{aligned} \quad (14.6.6)$$

Thus, the MLE of M_i is $\hat{M}_i = \bar{X}_i$. At $M_i = \hat{M}_i$,

$$\begin{aligned} \ln L &= -\frac{n \cdot p}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i,j} \text{tr}[\Sigma^{-1} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'] \\ &= -\frac{n \cdot p}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{tr}[\Sigma^{-1} (S_1 + \cdots + S_r)] \end{aligned}$$

where S_i is the i -th group sample sum of products matrix or sum of squares and cross products matrix for $i = 1, \dots, r$, see also Eq. (13.2.6). Letting $S = S_1 + \cdots + S_r$ and proceeding as in Chap. 3 by differentiating $\ln L$ with respect to Σ^{-1} , equating the result to a null matrix and simplifying, we obtain the MLE of Σ as

$$\hat{\Sigma} = \frac{1}{n} S = \frac{1}{n} (S_1 + \cdots + S_r). \quad (14.6.7)$$

An unbiased estimator of Σ is $\frac{1}{n-r} (S_1 + \cdots + S_r) = \frac{1}{n-r} S$. For evaluating the MLE of the parameter vector A_i , we can proceed as follows: After substituting $\frac{1}{n} S$ to Σ in $\ln L$, consider the following for a specific i :

$$\begin{aligned}
\frac{\partial}{\partial A_i} \ln L &= \frac{\partial}{\partial A_i} \sum_{i,j} (X_{ij} - TA_i)' \left(\frac{S}{n}\right)^{-1} (X_{ij} - TA_i) = O \Rightarrow \\
&T' \left(\frac{S}{n}\right)^{-1} \sum_j (X_{ij} - TA_i) = O \Rightarrow \\
&T'S^{-1}X_i - T'S^{-1}Tn_iA_i = O \Rightarrow \\
\hat{A}_i &= (T'S^{-1}T)^{-1}T'S^{-1}\bar{X}_i, \quad i = 1, \dots, r.
\end{aligned} \tag{14.6.8}$$

Note that this equation holds if S^{-1} is replaced by Σ^{-1} or Γ^{-1} for some real positive definite matrix Γ . Thus, the maximum of the likelihood function under the general model (14.6.4) is the following:

$$\max L = \frac{e^{-\frac{n \cdot p}{2}}}{(2\pi)^{\frac{n \cdot p}{2}} \left| \frac{1}{n} \sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \right|^{\frac{n}{2}}}. \tag{14.6.9}$$

As well, observe that the matrix in the determinant of (14.6.9) is

$$\frac{1}{n} \sum_{i=1}^r \left[\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \right] = \frac{1}{n} \times (\text{the residual matrix in MANOVA})$$

in a one-way layout, referring to Sect. 13.2. As well, this quantity is Wishart distributed with $n \cdot - r$ degrees of freedom and parameter matrix Σ , that is,

$$\sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \sim W_p(n \cdot - r, \Sigma), \quad \Sigma > O. \tag{14.6.10}$$

Now, consider the MLE of A_i :

$$\hat{A}_i = (T'S^{-1}T)^{-1}T'S^{-1}\bar{X}_i = C_1\bar{X}_i, \quad C_1 = (T'S^{-1}T)^{-1}T'S^{-1}. \tag{14.6.11}$$

Observe that $n \cdot$ cancels out in \hat{A}_i . In a normal population, the sample mean and the sample sum of products matrix are independently distributed. Hence, \bar{X}_i and S_i are independently distributed, and thereby \bar{X}_i and S are independently distributed. Moreover, $E[X_{ij}] = E[\bar{X}_i] = TA_i$ for each i , and we have

$$E[\hat{A}_i] = E\{(T'S^{-1}T)^{-1}T'S^{-1}\}E[\bar{X}_i] = E\{(T'S^{-1}T)^{-1}T'S^{-1}TA_i\} = A_i.$$

Thus, \hat{A}_i is an unbiased estimator for A_i . So far, we have not made any specification on the number m or the structure of T . We can incorporate the structure of T and the specification on the degree of the polynomial in T through a hypothesis. Let

$$H_o : T = T_o = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{m-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & t_p^2 & \dots & t_p^{m-1} \end{bmatrix} \quad (14.6.12)$$

for a specified m , that is, H_o posits that it is a polynomial TA_i of degree $m - 1$ for each i and this model fits. In this instance, the condition imposed through H_o is that the degree of the polynomial is a specified number $m - 1$ because p is known as the order of the population vector. Then, the factors $e^{-\frac{n_2 p}{2}}$ and $(2\pi)^{\frac{n_2 p}{2}}$ cancel out in the likelihood ratio criterion λ , leaving

$$\lambda = \frac{\max_{H_o} L}{\max L} = \frac{|\sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'|^{\frac{n}{2}}}{|\sum_{i,j} (X_{ij} - T_o \hat{A}_i)(X_{ij} - T_o \hat{A}_i)'|^{\frac{n}{2}}}. \quad (14.6.13)$$

Note that $\sum_{i,j} (X_{ij} - T_o \hat{A}_i) = \sum_{i,j} [(X_{ij} - \bar{X}_i) + (\bar{X}_i - T_o \hat{A}_i)]$, so that

$$\begin{aligned} \sum_{i,j} (X_{ij} - T_o \hat{A}_i)(X_{ij} - T_o \hat{A}_i)' &= \sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \\ &\quad + \sum_{i,j} (\bar{X}_i - T_o \hat{A}_i)(\bar{X}_i - T_o \hat{A}_i)'. \end{aligned}$$

Since $\lambda|H_o$ is a ratio of determinants, the parameter matrix Σ can be eliminated, and we can assume that the normal population has the identity matrix I as its covariance matrix. As well, $[X_{ij} - \bar{X}_i] = [(X_{ij} - TA_i) - (\bar{X}_i - TA_i)]$ and $[\bar{X}_i - T_o \hat{A}_i] = [(\bar{X}_i - T_o A_i) - (T_o \hat{A}_i - T_o A_i)]$. We have already established that \hat{A}_i is an unbiased estimator for A_i . Hence, without any loss of generality, we can assume the population to be $N_p(O, I)$ under H_o , and we can consider

$$\begin{aligned} \bar{X}_i - T_o(T_o' S^{-1} T_o)^{-1} T_o' S^{-1} \bar{X}_i &\equiv (I - B) \bar{X}_i \\ \bar{X}_i &\sim N_p(O, \frac{1}{n_i} I) \Rightarrow \sqrt{n_i} \bar{X}_i \sim N_p(O, I) \end{aligned}$$

where $B = T_o(T_o' S^{-1} T_o)^{-1} T_o' S^{-1}$. Now, since

$$B^2 = [T_o(T_o' S^{-1} T_o)^{-1} T_o' S^{-1}][T_o(T_o' S^{-1} T_o)^{-1} T_o' S^{-1}] = T_o(T_o' S^{-1} T_o)^{-1} T_o' S^{-1} = B,$$

B is idempotent of rank m where m is equal to the rank of T_o . For convenience, let $p_1 = p - m$ and $p_2 = m$, so that $p_1 + p_2 = p$. Observe that $X_{ij} - \bar{X}_i$ and \bar{X}_i are independently distributed, and that $\bar{X}_i = \frac{1}{n_i} J J' X_{ij}$ and $X_{ij} - \bar{X}_i = [I - \frac{1}{n_i} J J'] X_{ij}$

where $J' = [1, 1, \dots, 1]$ or J is an $n_i \times 1$ vector of unities. Now, observing that $[I - \frac{1}{n_i}JJ'][\frac{1}{n_i}JJ'] = O$ and that, under normality, $\sqrt{n_i}\bar{X}_i \sim N_p(O, I)$, it follows that $U_1 = \sum_{i,j}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ and $U_2 = [(I - B)\bar{X}_i][(I - B)\bar{X}_i]'$ are independently distributed since U_1 and \bar{X}_i are independently distributed. Accordingly,

$$w = \lambda^{\frac{2}{n}} = \frac{|U_1|}{|U_1 + U_2|}. \tag{14.6.14}$$

Since $I - B$ is idempotent and of rank $p - m = p_1$, we can perform an orthonormal transformation on $\sqrt{n_i}\bar{X}_i$ and, equivalently, write U_2 as $U_2 \rightarrow \begin{bmatrix} Z & O \\ O & O \end{bmatrix}$ in distribution where $Z = [Z_1, \dots, Z_r][Z_1, \dots, Z_r]'$, $Z_j \stackrel{iid}{\sim} N_{p_1}(O, I_{p_1})$. Then, $Z \sim W_{p_1}(r, I_{p_1})$ for $r \geq p_1$. Moreover, we can reduce the $p \times p$ matrix U_1 to a $p_1 \times p_1$ submatrix in the determinant. Let us consider the following partition of U_1 where U_{11} is $p_1 \times p_1$ and U_{22} is $p_2 \times p_2$, $p_1 + p_2 = p$:

$$U_1 = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}, \quad |U_1| = |U_{22}| |U_{11} - U_{12}U_{22}^{-1}U_{21}|,$$

which follows from results included in Chap. 1 on the determinants of partitioned matrices. Then, w of (14.6.14) is such that

$$w = \frac{|U_{11} - U_{12}U_{22}^{-1}U_{21}|}{|(Z + U_{11}) - U_{12}U_{22}^{-1}U_{21}|}.$$

Now, for fixed U_{22} , make the transformation $V_{12} = U_{12}U_{22}^{-\frac{1}{2}} \Rightarrow dV_{12} = |U_{22}|^{-\frac{p_1}{2}} dU_{12}$, referring to Chap. 1. Let us take $p_1 \geq p_2$; if not, simply interchange p_1 and p_2 in the following procedure. Then, make the transformation

$$S = V_{12}V_{12}' \Rightarrow dV_{12} = \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_2}(\frac{p_1}{2})} |S|^{\frac{p_1}{2} - \frac{p_2 + 1}{2}} dS,$$

the Jacobian being given in Sect. 4.3. Observe that U_1 has a Wishart density with $n - r$ degrees of freedom and parameter matrix I_p where $n = n_1 + \dots + n_r$. Letting the density of U_1 be denoted by $f(U_1)$, we have

$$\begin{aligned}
f(U_1) &= c |U_1|^{\frac{n-r}{2} - \frac{p+1}{2}} e^{-\frac{1}{2}\text{tr}(U_1)} \\
&= c |U_{22}|^\rho |U_{11} - U_{12}U_{22}^{-1}U_{21}|^\rho e^{-\frac{1}{2}[\text{tr}(U_{11}) + \text{tr}(U_{22})]}, \quad \rho = \frac{n-r}{2} - \frac{p+1}{2}, \\
&= c |U_{22}|^{\rho + \frac{p_1}{2}} |U_{11} - V_{12}V'_{12}|^\rho e^{-\frac{1}{2}[\text{tr}(U_{11}) + \text{tr}(U_{22})]} \\
&= c \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_2}(\frac{p_1}{2})} |S|^{\frac{p_1}{2} - \frac{p_2+1}{2}} |U_{22}|^{\rho + \frac{p_1}{2}} |U_{11} - S|^\rho e^{-\frac{1}{2}[\text{tr}(U_{11}) + \text{tr}(U_{22})]} \\
&= c \frac{\pi^{\frac{p_1 p_2}{2}}}{\Gamma_{p_2}(\frac{p_1}{2})} |S|^{\frac{p_1}{2} - \frac{p_2+1}{2}} \\
&\quad \times |U_{22}|^{\rho + \frac{p_1}{2}} |V_{22}|^\rho e^{-\frac{1}{2}[\text{tr}(S) + \text{tr}(V_{22}) + \text{tr}(U_{22})]}, \quad V_{22} = U_{11} - S = U_{11} - U_{12}U_{22}^{-1}U_{21},
\end{aligned}$$

where c is the normalizing constant. Now, on integrating out U_{22} and S , we obtain the marginal density of V_{22} , denoted by $f_1(V_{22})$, as

$$f_1(V_{22}) = c_2 |V_{22}|^{\frac{n-r}{2} - \frac{p_2}{2} - \frac{p_1+1}{2}} e^{-\frac{1}{2}\text{tr}(V_{22})} \quad (14.6.15)$$

where c_2 is the normalizing constant. This shows that $V_{22} \sim W_{p_1}(n-r-p_2, I_{p_1})$ for $n-r-p_2 \geq p_1$ and that V_{22} and Z are independently distributed. Then,

$$w = \frac{|V_{22}|}{|V_{22} + Z|}$$

where V_{22} and Z are $p_1 \times p_1$ real matrices that are independently Wishart distributed with common parameter matrix I_{p_1} and degrees of freedoms $n-r-p_2$ and r , respectively. Then, $w = |W|$ where $W = (V_{22} + Z)^{-\frac{1}{2}} V_{22} (V_{22} + Z)^{-\frac{1}{2}}$ is $p_1 \times p_1$ real matrix-variate type-1 beta distributed with the parameters $(\frac{n-r-p_2}{2}, \frac{r}{2})$. Therefore, for an arbitrary h , the h -th moment of the determinant of W is the following:

$$E[w^h] = E[|W|^h] = c_3 \frac{\Gamma_{p_1}(\frac{n-r-p_2}{2} + h)}{\Gamma_{p_1}(\frac{n-r-p_2}{2} + \frac{r}{2} + h)} \quad (14.6.16)$$

where c_3 is the normalizing constant such that when $h = 0$, $E[w^h] = 1$. Since the likelihood ratio criterion is $\lambda = w^{\frac{n}{2}}$,

$$E[\lambda^h] = c_3 \frac{\Gamma_{p_1}(\frac{n-r-p_2}{2} + \frac{n h}{2})}{\Gamma_{p_1}(\frac{n-r-p_2}{2} + \frac{r}{2} + \frac{n h}{2})}. \quad (14.6.17)$$

As illustrated in earlier chapters, see for example Sect. 13.3, we can express the exact density of w in (14.6.16) as a G-function and the exact density of λ in (14.6.17) as an

H-function. Letting the densities of w and λ be denoted by $f_2(w)$ and $f_3(\lambda)$, respectively, we have

$$f_2(w) = c_3 G_{p_1, p_1}^{p_1, 0} \left[w \left| \begin{matrix} \frac{n-r-p_2}{2} - \frac{j-1}{2}, j=1, \dots, p_1 \\ \frac{n-r-p_2}{2} - \frac{j-1}{2}, j=1, \dots, p_1 \end{matrix} \right. \right] \tag{14.6.18}$$

$$f_3(\lambda) = c_3 H_{p_1, p_1}^{p_1, 0} \left[\lambda \left| \begin{matrix} (\frac{n-r-p_2-j+1}{2}, \frac{n_j}{2}), j=1, \dots, p_1 \\ (\frac{n-r-p_2-j+1}{2}, \frac{n_j}{2}), j=1, \dots, p_1 \end{matrix} \right. \right] \tag{14.6.19}$$

for $0 < w < 1$ and $0 < \lambda < 1$, and zero elsewhere, where the G and H-functions were defined earlier, for example in Chap. 13. For theoretical results on the G- and H-functions as well as applications, the reader is, for instance, referred to Mathai (1993) and Mathai, Saxena and Haubold (2010), respectively. We can express the exact density functions $f_2(w)$ and $f_3(\lambda)$ in terms of elementary functions for special values of p , m and n . For moderate values of n , p and m , numerical evaluations can be obtained by making use of the G- and H-functions programs in *Mathematica* and MAPLE. For large values of n , asymptotic results may be invoked.

Note 14.6.1. The test statistic w can also be expressed as

$$w = \frac{1}{|I + V_{22}^{-\frac{1}{2}} Z V_{22}^{-\frac{1}{2}}|} = \frac{1}{|I + W_2|}, \quad W_2 = V_{22}^{-\frac{1}{2}} Z V_{22}^{-\frac{1}{2}},$$

where W_2 is a real matrix-variate type-2 beta random variable with the parameters $(\frac{n-r-p_2}{2}, \frac{r}{2})$. As well, $V_{22} \sim W_{p_1}(n-r-p_2, I_{p_1})$, $Z \sim W_{p_1}(r, I_{p_1})$, and V_{22} and Z are independently distributed. While Roy’s largest root test is based on the largest eigenvalue of W_2 , Hotelling’s trace test relies on the trace of W_2 .

14.6.3. Some special cases

In certain particular cases, one can obtain exact densities in terms of elementary functions. Two such cases will now be considered.

Case (1) $p_2 = m = p - 1 \Rightarrow p_1 = p - m = 1$, $n-r-p_2 = n-r-p+1$. Since $p_1 = 1$, there will be only one gamma ratio containing h , and

$$E[w^h] = c_3 \frac{\Gamma(\frac{n-r-p+1}{2} + h)}{\Gamma(\frac{n-p+1}{2} + h)}, \quad \Re(h) > -\frac{n-p-r+1}{2}, \tag{i}$$

where c_3 is the normalizing constant such that the right-hand side is 1 when $h = 0$. Observe that (i) is the h -th moment of a real scalar type-1 beta random variable and hence, $w \sim$

real scalar type-1 beta $(\frac{n.-r-p+1}{2}, \frac{r}{2})$. Accordingly, $y_1 = \frac{1-w}{w}$ is a real scalar type-2 beta random variable with the parameters $(\frac{r}{2}, \frac{n.-r-p+1}{2})$, so that

$$t_1 = \frac{n.-r-p+1}{r} y_1 \sim F_{r, n.-r-p+1}, \quad n.-r-p+1 \geq 1.$$

Thus, for large values of this F statistic, we reject the null hypothesis that the growth curve has the structure of T_o of degree $m-1$ as specified in (14.6.12) and that the growth curve model fits. Equivalently, we reject H_o when the observed value of t_1 is greater than or equal to $F_{r, n.-r-p+1, \alpha}$, such that $Pr\{F_{r, n.-r-p+1} \geq F_{r, n.-r-p+1, \alpha}\} = \alpha$ for a preselected α .

Case (2) $p_2 = m = p-2 \Rightarrow p_1 = 2$, which indicates that the h -th moment will contain two gamma ratios involving h . More specifically,

$$E[w^h] = c_3 \frac{\Gamma(\frac{n.-r-p+2}{2} + h) \Gamma(\frac{n.-r-p+2}{2} - \frac{1}{2} + h)}{\Gamma(\frac{n.-p+2}{2} + h) \Gamma(\frac{n.-p+2}{2} - \frac{1}{2} + h)}, \quad \Re(h) > \frac{n.-p+2}{2}. \quad (ii)$$

Since the arguments of the gamma functions in (ii) differ by $\frac{1}{2}$, we can combine them by applying the duplication formula for gamma functions, namely

$$\Gamma(z) \Gamma(z + \frac{1}{2}) = \pi^{\frac{1}{2}} 2^{1-2z} \Gamma(2z). \quad (iii)$$

Consider $E[\sqrt{w}]^h$ so that h becomes $\frac{h}{2}$. Now, by taking $z = \frac{n.-r-p+2}{2} - \frac{1}{2} + \frac{h}{2}$ in the gamma functions containing h and $h=0$ in the constant part, we have

$$E[\sqrt{w}]^h = c_{31} \frac{\Gamma(n.-r-p+1+h)}{\Gamma(n.-p+1+h)}$$

where c_{31} is the corresponding normalizing constant. Thus, $\sqrt{w} \sim$ real scalar type-1 beta with the parameters $(n.-r-p+1, r)$ for $n.+1 > r+p$, and

$$y_2 = \frac{1-\sqrt{w}}{\sqrt{w}} \sim \text{type-2 beta}(r, n.-r-p+1), \quad t_2 = \frac{n.-r-p+1}{r} y_2 \sim F_{2r, 2(n.-r-p+1)}. \quad (iv)$$

Accordingly, we reject H_o for large values of the F statistic given in (iv) or equivalently, for an observed value of t_2 that is greater than or equal to $F_{2r, 2(n.-r-p+1), \alpha}$, such that $Pr\{F_{2r, 2(n.-r-p+1)} \geq F_{2r, 2(n.-r-p+1), \alpha}\} = \alpha$ for a given significance level α .

14.6.4. Asymptotic distribution of the likelihood ratio criterion λ

Given the gamma structure appearing in (14.6.17), we can expand the gamma functions for $\frac{n}{2}(1+h) \rightarrow \infty$ in the part containing h and $\frac{n}{2} \rightarrow \infty$ in the constant part wherein $h = 0$, by using the first term in the asymptotic expansion of the gamma function or Stirling’s formula, namely

$$\Gamma(z + \delta) \approx \sqrt{(2\pi)}z^{z+\delta-\frac{1}{2}}e^{-z} \tag{14.6.20}$$

for $|z| \rightarrow \infty$, when δ is a bounded quantity. Then, on expanding each gamma function in (14.6.17) by making use of (14.6.20), we have

$$E[\lambda^h] \rightarrow (1+h)^{-\frac{1}{2}r(p_1)} \Rightarrow E[e^{(-2\ln\lambda)h}] \rightarrow (1-2h)^{-\frac{1}{2}rp_1} \text{ for } 1-2h > 0,$$

which shows that whenever $n = n_1 + \dots + n_r \rightarrow \infty$,

$$\lambda \rightarrow \chi_{rp_1}^2 \tag{14.6.21}$$

where the χ^2 is a real scalar chisquare random variable having $rp_1 = r(p-m)$ degrees of freedom. Hence, for large values of $n = n_1 + \dots + n_r$, we reject the null hypothesis H_0 that the model specified by T_0 fits, whenever the observed $\lambda \geq \chi_{r(p-m),\alpha}^2$, with $Pr\{\chi_{r(p-m)}^2 \geq \chi_{r(p-m),\alpha}^2\} = \alpha$ for a preselected α .

Example 14.6.1. A certain exercise regimen is prescribed to a random sample of 5 men and 6 women to stabilize their weights to within normal limits. Except for being of different gender, these men and women form homogeneous groups with respect to all factors that may influence weight, such as age and health conditions. The measurements are initial weight minus weight on the reading date. The weights are recorded every 5th day and a 10-day period is taken as one time unit, the observations being made for 20 days or 2 time units. Test the fit of (1) a linear model, (2) a second degree growth curve to the following data, the columns of the matrix A being the readings on men, denoted by X_{ij} , and the columns of the matrix B , the readings on women, denoted by Y_{ij} :

$$A = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & -1 & 1 \\ 1 & 2 & 0 & 1 & 1 \end{bmatrix} \equiv \mathbf{X}, \quad B = \begin{bmatrix} -1 & 1 & 0 & -1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 2 \\ 0 & -1 & -1 & 1 & 1 & 0 \\ -1 & -1 & 1 & 0 & 0 & 1 \end{bmatrix} \equiv \mathbf{Y}.$$

Solution 14.6.1. Let the sample average vectors be denoted by \bar{X} for men and \bar{Y} for women, the deviation matrices be denoted as \mathbf{X}_d and \mathbf{Y}_d and the sample sum of products matrices, by S_1 and S_2 respectively. Then, these items are as follows:

$$\bar{X} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{X}_d = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & -1 & 1 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix}, S_1 = \mathbf{X}_d \mathbf{X}_d' = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & -1 & 2 \\ 0 & -1 & 4 & -1 \\ 0 & 2 & -1 & 2 \end{bmatrix},$$

$$\bar{Y} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{Y}_d = \begin{bmatrix} -1 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 & 1 & 0 \\ -1 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}, S_2 = \mathbf{Y}_d \mathbf{Y}_d' = \begin{bmatrix} 4 & 1 & -1 & 0 \\ 1 & 2 & -1 & 1 \\ -1 & -1 & 4 & 0 \\ 0 & 1 & 0 & 4 \end{bmatrix}.$$

Let $S = S_1 + S_2$. Consider the determinant S and its inverse S^{-1} obtained from the matrix of cofactors as that matrix divided by the determinant since it is a symmetric matrix. Thus,

$$S = S_1 + S_2 = \begin{bmatrix} 6 & 1 & -1 & 0 \\ 1 & 4 & -2 & 3 \\ -1 & -2 & 8 & -1 \\ 0 & 3 & -1 & 6 \end{bmatrix}, |S| = 580,$$

$$\text{Cof}(S) = \begin{bmatrix} 104 & -38 & 6 & 20 \\ -38 & 276 & 48 & -130 \\ 6 & 48 & 84 & -10 \\ 20 & -130 & -10 & 160 \end{bmatrix}, S^{-1} = \frac{1}{580} \begin{bmatrix} 104 & -38 & 6 & 20 \\ -38 & 276 & 48 & -130 \\ 6 & 48 & 84 & -10 \\ 20 & -130 & -10 & 160 \end{bmatrix}.$$

Take the growth matrices for the linear growth and quadratic growth as T_1 and T_2 , where

$$T_1 = \begin{bmatrix} 1 & 1/2 \\ 1 & 1 \\ 1 & 3/2 \\ 1 & 2 \end{bmatrix} \text{ and } T_2 = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1 & 1 & 1 \\ 1 & 3/2 & 9/4 \\ 1 & 2 & 4 \end{bmatrix}.$$

Now, let us compute the various quantities that are required to answer the questions:

$$\begin{aligned}
 T_1' S^{-1} &= \frac{1}{580} \begin{bmatrix} 92 & 156 & 128 & 40 \\ 63 & 69 & 157 & 185 \end{bmatrix}, \\
 T_1' S^{-1} \bar{X} &= \frac{1}{580} \begin{bmatrix} 40 \\ 185 \end{bmatrix}, \quad T_1' S^{-1} \bar{Y} = \frac{1}{580} \begin{bmatrix} 156 \\ 69 \end{bmatrix}, \\
 T_1' S^{-1} T_1 &= \frac{1}{580} \begin{bmatrix} 416 & 474 \\ 474 & 706 \end{bmatrix}, \quad (T_1' S^{-1} T_1)^{-1} = \frac{580}{69020} \begin{bmatrix} 706 & -474 \\ -474 & 416 \end{bmatrix}, \\
 T_1 (T_1' S^{-1} T_1)^{-1} T_1' S^{-1} &= \frac{1}{69020} \begin{bmatrix} 26390 & 54810 & 18270 & -30450 \\ 17690 & 32190 & 20590 & -1450 \\ 8990 & 9570 & 22910 & 27550 \\ 290 & -13050 & 25230 & 56550 \end{bmatrix}.
 \end{aligned}$$

We may now evaluate the estimates of the parameter vectors A_1 and A_2 , denoted by a hat:

$$\begin{aligned}
 \hat{A}_1 &= (T_1' S^{-1} T_1)^{-1} T_1' S^{-1} \bar{X} = \frac{1}{69020} \begin{bmatrix} -59450 \\ 58000 \end{bmatrix} = \begin{bmatrix} -0.861345 \\ 0.840336 \end{bmatrix} = \begin{bmatrix} \hat{a}_{01} \\ \hat{a}_{11} \end{bmatrix}, \\
 \hat{A}_2 &= (T_1' S^{-1} T_1)^{-1} T_1' S^{-1} \bar{Y} = \frac{1}{69020} \begin{bmatrix} 77430 \\ -45240 \end{bmatrix} = \begin{bmatrix} 1.12185 \\ -0.655462 \end{bmatrix} = \begin{bmatrix} \hat{a}_{02} \\ \hat{a}_{12} \end{bmatrix}.
 \end{aligned}$$

If the hypothesis of linear growth was not rejected, then the models for men and women, respectively denoted by $f_x(t)$ and $f_y(t)$ would be the following:

$$\text{Men: } f_x(t) = -0.861345 + 0.840336t \quad \text{Women: } f_y(t) = 1.12185 - 0.655462t.$$

As well,

$$\begin{aligned}
 T_1 \hat{A}_1 &= T_1 (T_1' S^{-1} T_1)^{-1} T_1' S^{-1} \bar{X} = \frac{1}{69020} \begin{bmatrix} -30450 \\ -1450 \\ 27550 \\ 56550 \end{bmatrix} = \begin{bmatrix} -0.441176 \\ -0.0210084 \\ 0.39916 \\ 0.819328 \end{bmatrix}, \\
 T_1 \hat{A}_2 &= T_1 (T_1' S^{-1} T_1)^{-1} T_1' S^{-1} \bar{Y} = \frac{1}{69020} \begin{bmatrix} 54810 \\ 32190 \\ 9570 \\ -13050 \end{bmatrix} = \begin{bmatrix} 0.794118 \\ 0.466387 \\ 0.138655 \\ -0.189076 \end{bmatrix}.
 \end{aligned}$$

Let us now compute the sum of products matrix under the linear growth model. The matrix $\mathbf{X} - \mathbf{T}_1 \hat{\mathbf{A}}_1$ is such that the vector $T_1 \hat{A}_1$ is subtracted from each column of the observation matrix \mathbf{X} . Letting

$$G_1 G_1' = \sum_{j=1}^{n_1} (X_{1j} - T_1 \hat{A}_1)(X_{1j} - T_1 \hat{A}_1)' \text{ with}$$

$$G_1 = \mathbf{X} - \mathbf{T}_1 \hat{\mathbf{A}}_1 = \begin{bmatrix} 1.44118 & 0.441176 & 0.441176 & -0.558824 & 0.441176 \\ 0.0210084 & 1.02101 & -0.978992 & 0.0210084 & 0.0210084 \\ -1.39916 & -0.39916 & 0.60084 & -1.39916 & 0.60084 \\ 0.180672 & 1.18067 & -0.819328 & 0.180672 & 0.180672 \end{bmatrix},$$

$$G_1 G_1' = \begin{bmatrix} 2.97318 & 0.0463421 & -0.880499 & 0.398542 \\ 0.0463421 & 2.00221 & -1.04193 & 2.01898 \\ -0.880499 & -1.04193 & 4.79664 & -1.36059 \\ 0.398542 & 2.01898 & -1.36059 & 2.16321 \end{bmatrix}.$$

We now evaluate the same matrices for the second sample, denoting the deviation matrix by G_2 and the sum of products matrix by $G_2 G_2'$, where G_2 is obtained by subtracting $T \hat{A}_2$ from each column of the observation matrix \mathbf{Y} :

$$G_2 = \mathbf{Y} - \mathbf{T}_1 \hat{\mathbf{A}}_2$$

$$= \begin{bmatrix} -1.79412 & 0.205882 & -0.794118 & -1.79412 & 0.205882 & -0.794118 \\ 0.533613 & 0.533613 & 0.533613 & -0.466387 & 0.533613 & 1.53361 \\ -0.138655 & -1.13866 & -1.13866 & 0.861345 & 0.861345 & -0.138655 \\ -0.810924 & -0.810924 & 1.18908 & 0.189076 & 0.189076 & 1.18908 \end{bmatrix},$$

$$G_2 G_2' = \begin{bmatrix} 7.78374 & -1.54251 & -0.339348 & -0.90089 \\ -1.54251 & 3.70846 & -1.44393 & 1.60536 \\ -0.339348 & -1.44393 & 4.11535 & -0.157298 \\ -0.90089 & 1.60536 & -0.157298 & 4.2145 \end{bmatrix}.$$

Thus,

$$G_1 G_1' + G_2 G_2' = \begin{bmatrix} 10.7569 & -1.49617 & -1.21985 & -0.502348 \\ -1.49617 & 5.71067 & -2.48586 & 3.62434 \\ -1.21985 & -2.48586 & 8.91199 & -1.51788 \\ -0.502348 & 3.62434 & -1.51788 & 6.37771 \end{bmatrix} \text{ and}$$

$$|G_1 G_1' + G_2 G_2'| = 1798.49,$$

so that

$$w = \frac{|S_1 + S_2|}{|G_1 G_1' + G_2 G_2'|} = \frac{580}{1798.49} = 0.322493.$$

In this case, $n - r = 11 - 2 = 9$, $m = 2$, $p = 4$, $p - m = p_1 = 2$ and $p_2 = 2$. This situation is our special Case (2). Then,

$$y_2 = \frac{1 - \sqrt{w}}{\sqrt{w}} = \frac{1 - 0.5679}{0.5679} = 0.7609,$$

$$t_2 = \frac{(n - r - p + 1)}{r} y_2 = \frac{6}{2} y_2 = 3y_2 = 2.2826,$$

which is the observed value of an $F_{2r,2(n-r-p+1)} = F_{4,12}$ random variable under the null hypothesis. From an F-table, at the 5% level, $F_{4,12,0.05} = 3.26 > 2.2826$. In the case of the F-test, we reject the hypothesis for large observed values of the statistic. Accordingly, we do not reject H_o that $m = 2$ or that the growth model is linear. Naturally, the null hypothesis would not be rejected either at a lower level of significance.

We now tackle the second part of the problem wherein $m = 3$ and $m - 1 = 2$, that is, the growth model is assumed to be a polynomial of degree 2 under the null hypothesis. Again, we take the time intervals to be $t_1 = \frac{1}{2}$, $t_2 = 1$, $t_3 = \frac{3}{2}$ and $t_4 = 2$, so that

$$T_2 = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1 & 1 & 1 \\ 1 & 3/2 & 9/4 \\ 1 & 2 & 4 \end{bmatrix}.$$

We can make use of some of the computations from the first part of the problem, that is, the calculations related to T_1 . Thus, we have

$$T_2'S^{-1} = \frac{1}{580} \begin{bmatrix} 92 & 156 & 128 & 40 \\ 63 & 69 & 157 & 185 \\ 81.5 & -145.5 & 198.5 & 492.5 \end{bmatrix}$$

$$T_2'S^{-1}T_2 = \frac{1}{580} \begin{bmatrix} 416 & 474 & 627 \\ 474 & 706 & 1178 \\ 627 & 1178 & 2291.5 \end{bmatrix}, \quad |580 \times T_2'S^{-1}T_2| = 3532200$$

$$\begin{aligned} \text{Cof}(T_2'S^{-1}T_2) &= \begin{bmatrix} 230115 & -347565 & 115710 \\ -347565 & 560135 & -192850 \\ 115710 & -192850 & 69020 \end{bmatrix} \\ (T_2'S^{-1}T_2)^{-1} &= \frac{580}{3532200} \begin{bmatrix} 230115 & -347565 & 115710 \\ -347565 & 560135 & -192850 \\ 115710 & -192850 & 69020 \end{bmatrix} \\ T_2'S^{-1}\bar{X} &= \frac{1}{580} \begin{bmatrix} 40 \\ 185 \\ 492.5 \end{bmatrix}, \quad T_2'S^{-1}\bar{Y} = \frac{1}{580} \begin{bmatrix} 156 \\ 69 \\ -145.5 \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} \hat{A}_1 &= (T_2'S^{-1}T_2)^{-1}T_2'S^{-1}\bar{X} \\ &= \frac{1}{3532200} \begin{bmatrix} 1892250 \\ -5256250 \\ 2943500 \end{bmatrix} = \begin{bmatrix} 0.535714 \\ -1.4881 \\ 0.833333 \end{bmatrix} = \begin{bmatrix} \hat{a}_{01} \\ \hat{a}_{11} \\ \hat{a}_{21} \end{bmatrix}, \\ \hat{A}_2 &= (T_2'S^{-1}T_2)^{-1}T_2'S^{-1}\bar{Y} \\ &= \frac{1}{3532200} \begin{bmatrix} -4919850 \\ 12488850 \\ -5298300 \end{bmatrix} = \begin{bmatrix} -1.39286 \\ 3.53571 \\ -1.5 \end{bmatrix} = \begin{bmatrix} \hat{a}_{02} \\ \hat{a}_{12} \\ \hat{a}_{22} \end{bmatrix}. \end{aligned}$$

If the hypothesis of a second degree growth model was not rejected, the model for men, denoted by $f_x(t)$, and that for women, denoted by $f_y(t)$, would be the following:

$$\text{Men: } f_x(t) = 0.535714 - 1.4881t + 0.83333t^2,$$

$$\text{Women: } f_y(t) = -1.39286 + 3.53571t - 1.5t^2.$$

Employing notations paralleling those utilized in the first part, we have

$$\begin{aligned} T_2\hat{A}_1 &= T_2(T_2'S^{-1}T_2)^{-1}T_2'S^{-1}\bar{X} = \frac{1}{3532200} \begin{bmatrix} 0 \\ -420500 \\ 630750 \\ 3153750 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.119048 \\ 0.178571 \\ 0.892857 \end{bmatrix}, \\ T_2\hat{A}_2 &= T_2(T_2'S^{-1}T_2)^{-1}T_2'S^{-1}\bar{Y} = \frac{1}{3532200} \begin{bmatrix} 0 \\ 2270700 \\ 1892250 \\ -1135350 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.642857 \\ 0.535714 \\ -0.321429 \end{bmatrix}, \end{aligned}$$

as well as the following sample deviation matrices and sample sum of products matrices:

$$G_1 = \mathbf{Y} - \mathbf{T}_2 \hat{\mathbf{A}}_1$$

$$= \begin{bmatrix} -1 & 1 & 0 & -1 & 1 & 0 \\ 1.11905 & 1.11905 & 1.11905 & 0.119048 & 1.11905 & 2.11905 \\ -0.178571 & -1.17857 & -1.17857 & 0.821429 & 0.821429 & -0.178571 \\ -1.89286 & -1.89286 & 0.107143 & -0.892857 & -0.892857 & 0.107143 \end{bmatrix},$$

$$G_1 G_1' = \begin{bmatrix} 4. & 1. & -1. & 0. \\ 1. & 9.51361 & -2.19898 & -4.9949 \\ -1. & -2.19898 & 4.19133 & 0.95663 \\ 0. & -4.9949 & 0.95663 & 8.78316 \end{bmatrix};$$

$$G_2 = \mathbf{Y} - \mathbf{T}_2 \hat{\mathbf{A}}_2$$

$$= \begin{bmatrix} -1 & 1 & 0 & -1 & 1 & 0 \\ 0.357143 & 0.357143 & 0.357143 & -0.642857 & 0.357143 & 1.35714 \\ -0.535714 & -1.53571 & -1.53571 & 0.464286 & 0.464286 & -0.535714 \\ -0.678571 & -0.678571 & 1.32143 & 0.321429 & 0.321429 & 1.32143 \end{bmatrix},$$

$$G_2 G_2' = \begin{bmatrix} 4. & 1. & -1. & 0. \\ 1. & 2.76531 & -2.14796 & 1.68878 \\ -1. & -2.14796 & 5.72194 & -1.03316 \\ 0. & 1.68878 & -1.03316 & 4.6199 \end{bmatrix};$$

$$G_1 G_1' + G_2 G_2' = \begin{bmatrix} 8. & 2. & -2. & 0. \\ 2. & 12.2789 & -4.34694 & -3.30612 \\ -2. & -4.34694 & 9.91326 & -0.0765339 \\ 0. & -3.30612 & -0.0765339 & 13.4031 \end{bmatrix} \text{ and}$$

$$|G_1 G_1' + G_2 G_2'| = 9462.7797.$$

Thus, for this part, the statistic w obtained from the λ -criterion, is the following:

$$w = \frac{|S_1 + S_2|}{|G_1 G_1' + G_2 G_2'|} = \frac{580}{9462.7797} = 0.0612928.$$

In this case, $m = 3$, $p = 4$, $r = 2$, $p_1 = 1$, $p_2 = 3$ and $n - r - p + 1 = 11 - 2 - 4 + 1 = 6$,

$$y_1 = \frac{1 - w}{w} = 15.3151, \quad t_1 = \frac{n - r - p + 1}{r} y_1 = \frac{6}{2} (15.3151) = 45.945,$$

$$\text{and } F_{2,6,0.01} = 10.92.$$

When implementing F-tests, we reject the null hypothesis for large observed value of the statistic. Since $45.945 > F_{2,6,0.01}$, the hypothesis that a second degree polynomial fits is rejected at the 1% significance level and, of course, at any higher level. Let us consider the asymptotic results. Since $n = 11$ is not large, the results ought to be interpreted with some caution. In the case of the linear model ($m = 2$), $rp_1 = 2(2) = 4$ and the tabulated critical values are $\chi_{4,0.05}^2 = 9.49$ and $\chi_{4,0.01}^2 = 13.26$. With this chisquare test, we reject H_0 for large observed value of the statistic. In the linear case, $w = 0.3225$ so that $-2 \ln \lambda = -2(n./2) \ln w = -11 \ln w = -11[-1.1317] = 12.45$. Hence, according to the approximate asymptotic chisquare test, the hypothesis of a linear fit should be rejected at the 5% significance level but not at the 1% significance level. In the second degree polynomial growth curve case, $p_1 = 1$ so that $rp_1 = 2$, and the tabulated values are $\chi_{2,0.05}^2 = 5.99$ and $\chi_{2,0.01}^2 = 9.49$. Since $-2 \ln \lambda = -11 \ln w = -11 \ln 0.0613 = -11[-2.79] = 30.71$, according to the approximate asymptotic chisquare test, the hypothesis of a second degree polynomial fit ought to be rejected at both the 5 and 1% levels, which corroborates the conclusion obtained with the F test. In sum, there is some evidence of a linear fit but no indication whatsoever of a quadratic fit.

14.6.5. A general structure

Potthoff and Roy (1964) considered a general format whereby several hypotheses can be tested at the same time by making use of the structure. When applied to the model specified in (14.6.4), the Potthoff-Roy procedure yields the following format:

$$E[X_{11}, X_{12}, \dots, X_{1n_1}] = [TA_1, TA_1, \dots, TA_1] = T(A_1, \dots, A_1) = TA_1J_1' \quad (14.6.22)$$

where J_1 is a $n_1 \times 1$ vector of unities. The left and right-hand sides of this equation are $p \times n_1$ matrices where T is $p \times m$ and the $m \times 1$ vector A_1 is repeated n_1 times. The sample matrix of all r groups of $n_1 + \dots + n_r = n = n$ individuals, denoted by a boldfaced \mathbf{X} , is the following:

$$E[\mathbf{X}] = [TA_1J_1', \dots, TA_rJ_r'] = T[AJ_1', \dots, A_rJ_r'] = TAJ_{(r)}, \quad A = [A_1, \dots, A_r] \quad (14.6.23)$$

where $J_{(r)}$ is an $m \times (n_1 + \dots + n_r) = m \times n$ matrix whose first n_1 columns all have unities in the first row (or J_1') and zeros as the other elements or in the first $n_1 \times n_1$ diagonal block, the first row elements are all unities, in the next $n_2 \times n_2$ diagonal block, the first row elements are all unities and the rest, all zeros, and so on, up to the last $n_r \times n_r$ diagonal block wherein the first row are all unities and the remaining elements are zeros, all non-diagonal blocks being null matrices. We can also pre-multiply \mathbf{X} by a $q \times p$ constant matrix Q such that Q is of full rank $q \leq p$. Then $QX_{ij} \sim N_q(QTA_i, Q\Sigma Q')$, $j = 1, \dots, n_i$, $i = 1, \dots, r$. In

this instance, we are making repeated measurements at q points t_1, \dots, t_q , instead of the p points previously considered, and the model is the following:

$$E[Q\mathbf{X}] = QT AJ_{(r)}. \tag{14.6.24}$$

This model enables one to test all types of hypotheses of the form

$$CAV = O \tag{14.6.25}$$

where C and V are given matrices, C is $s \times m$, A is the $m \times r$ matrix of parameters and V is $r \times t$ for some r and t . For example, let $C = I_m$, V be an $r \times (r - 1)$ matrix where the first $r - 1$ columns form an identity matrix and the last column is $(-1, -1, \dots, -1)'$. Then, $A_1 - A_r = O$, $A_2 - A_r = O, \dots, A_{r-1} - A_r = O$, or equivalently, $A_1 = \dots = A_r$. For testing the equality of the models $TA_1 = TA_2 = \dots = TA_r$, one need not apply the above procedure. One can directly employ the multivariate one-way layout analysis with $r - 1$ degrees freedom associated with the sum of squares and cross products matrix due to the general nature of the model.

Exercises 14

14.1. The following are observed random samples of sizes $n_1 = 4$ and $n_2 = 5$ from two independent real Gaussian populations, $N_3(M_1, \Sigma)$ and $N_3(M_2, \Sigma)$ sharing the same covariance matrix $\Sigma > O$ and having the mean value vectors M_1 and M_2 , respectively. Test the hypothesis at the 5% significance level that the mean values M_1 and M_2 have (1): parallel profiles; (2): coincident profiles:

$$\text{Sample 1: } \begin{bmatrix} 1 & 2 & 4 & 1 \\ -1 & 1 & 2 & 2 \\ 1 & 3 & 1 & 3 \end{bmatrix}, \quad \text{Sample 2: } \begin{bmatrix} 2 & -1 & 1 & -2 & 0 \\ 1 & 1 & 4 & 1 & 3 \\ 4 & 2 & 3 & 0 & 1 \end{bmatrix}.$$

14.2. The following are observed samples of sizes 5 and 6 from two independent real Gaussian populations with the same covariance matrix. The columns of the matrices represent the observations. Test the hypothesis at the 5% level of significance that the expected values or population mean value vectors have (1): parallel profiles; (2): coincident profiles:

$$\text{Sample 1: } \begin{bmatrix} 1 & -1 & 2 & 3 & 0 \\ 1 & 2 & 1 & 1 & 4 \\ -1 & 3 & -1 & -3 & 0 \\ 1 & 1 & 2 & -1 & 2 \end{bmatrix}, \quad \text{Sample 2: } \begin{bmatrix} 2 & 1 & 3 & -1 & 1 & 0 \\ 1 & 0 & 5 & 2 & 2 & 2 \\ -1 & 1 & 2 & -2 & 4 & -4 \\ 1 & 4 & 5 & 3 & -1 & 0 \end{bmatrix}.$$

14.3. In Exercise 14.1, determine whether the profile is level (1): in the first population; (2): in the second population; (3): in both the populations.

14.4. Repeat Exercise 14.3 with the sample of observations provided in Exercise 14.2.

14.5. Two breeds of 12 cows having identical characteristics are given a particular type of feed and their weights are monitored for four weeks. The observations are the weight readings on the date minus the initial weight. The weight readings are made at the end of every week for four successive weeks. The columns in A represent successive observations on 5 cows of breed-1 and the columns of B , successive observations on 7 cows of breed-2. Test the hypothesis at the 5% significance level that the breed effect is a polynomial growth curve of (1): degree 1, (2): degree 2 for each breed:

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 3 \\ 1 & 2 & 1 & 2 & 4 \\ 2 & 4 & 2 & 3 & 4 \\ 2 & 6 & 2 & 4 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 3 & 0 \\ 2 & 1 & 1 & 3 & 2 & 3 & 2 \\ 3 & 1 & 2 & 5 & 2 & 4 & 4 \\ 4 & 2 & 2 & 7 & 3 & 6 & 4 \end{bmatrix}.$$

References

A.M. Mathai (1993): *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*, Oxford University Press, Oxford.

A.M. Mathai and H.J. Haubold (2017): *Probability and Statistics: A Course for Physicists and Engineers*, De Gruyter, Germany.

A.M. Mathai, R.K. Saxena and H.J. Haubold (2010): *The H-function: Theory and Applications*, Springer, New York.

R.F. Potthoff and S.N. Roy (1964): A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**(3), 313–326.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 15

Cluster Analysis and Correspondence Analysis



15.1. Introduction

We will employ the same notations as in the previous chapters. Lower-case letters x, y, \dots will denote real scalar variables, whether mathematical or random. Capital letters X, Y, \dots will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C . A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. The determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, dX will indicate the wedge product of all the distinct differentials of the elements of the matrix X . Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij} 's are distinct real scalar variables, $dX = \wedge_{i=1}^p \wedge_{j=1}^q dx_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{-1}$, where X_1 and X_2 are real, $d\tilde{X} = dX_1 \wedge dX_2$.

15.1.1. Clusters

A cluster means a group or a cloud of items close together with reference to one or more characteristics. For instance, in a countryside, there are villages which are clusters of houses. In a city, there are clusters of high-rise buildings or clusters of apartment blocks. If we have 2-dimensional data points marked on a sheet of paper, then there may be several places where the points are grouped together in large crowds, at other places the points may be bunched together in smaller clumps and somewhere else, there may be singleton points. In a classification problem, we have a number of preassigned populations and we

want to assign a point at hand to one of those populations. This cannot be achieved in the context of cluster analysis as we do not know beforehand how many clusters there are in the data at hand or which data point belongs to which cluster. Cluster analysis is akin to pattern recognition whereas classification is a sort of taxonomy. Suppose that a new plant is to be classified as belonging to one of the known species of plants; if it does not fall into any of the known species, then we have a member from a new species. In cluster analysis, we are, in a manner of speaking, going to create various ‘species’. To start with, we have only a cloud of items and we do not know how many categories or clusters there exist.

Cluster analysis techniques are widely utilized in many fields such as psychiatry, sociology, anthropology, archeology, medicine, criminology, engineering and geology, to mention only a few areas. If real scalar variables are to be classified as belonging to a certain category, one way of achieving this is to ascertain their joint dispersion or joint variation as measured in terms of scale-free covariance or correlation. Those variables that are similarly correlated may be grouped together.

We will consider the problem of cluster analysis involving n points X_1, \dots, X_n where each X_j is a real p -dimensional vector, that is, we have a $p \times n$ data matrix

$$\mathbf{X} = [X_1, X_2, \dots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix}. \quad (15.1.1)$$

15.1.2. Distance measures

Two real p -vectors are close together if the “distance” between them is small. Many types of distance measures can be defined. Let X_r and X_s be two real p -vectors. These are the r -th and s -th members or columns in the data matrix (15.1.1). Then, the following are some distance measures:

$$d_m(X_r, X_s) = \left[\sum_{i=1}^p |x_{ir} - x_{is}|^m \right]^{\frac{1}{m}};$$

for $m = 2$, we have the Euclidean distance $d_2(X_r, X_s) = [\sum_{i=1}^p |x_{ir} - x_{is}|^2]^{\frac{1}{2}}$, or, denoting d_2^2 as d^2 , we have

$$d^2(X_r, X_s) = \sum_{i=1}^p (x_{ir} - x_{is})^2, \quad (15.1.2)$$

where the absolute value sign can be replaced by parentheses since we are dealing with real elements. We will utilize this convenient quantity d^2 for comparing observation vectors. There may be joint variation or covariances among the coordinates in each of the vectors, in which case, $\text{Cov}(X_r) = \Sigma > O$. If all the X_j 's, $j = 1, \dots, n$, have the same covariance matrix, then $\text{Cov}(X_j) = \Sigma$, $j = 1, \dots, n$, and a statistician might wish to consider the generalized distance between X_r and X_s , or its square, $d_{(g)}^2(X_r, X_s) = (X_r - X_s)' \Sigma^{-1} (X_r - X_s)$, the subscript g designating the generalized distance. Since Σ is unknown, we may wish to estimate it. However, if there are clusters, it may not be appropriate to make use of the entire data set of all n points, since the joint variation or the covariance within each cluster is likely to be different. And as we do not know beforehand whether clusters are present, securing a proper estimate of Σ turns out to prove problematic. As a result, this problem is usually circumvented by resorting to the ordinary Euclidean distance instead of the generalized distance.

Let us examine the effect of scaling a vector. If the unit of measurement in one vector is changed, what will be the effect on the squared distance? Consider the following vectors:

$$X_1 = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} -3 \\ 2 \\ 4 \end{bmatrix} \Rightarrow d^2(X_1, X_2) = (X_1 - X_2)'(X_1 - X_2) \\ = [(-1) - (-3)]^2 + [(0) - (2)]^2 + [(-2) - (4)]^2 = 44.$$

The squared distances between the vectors when (1) X_1 is multiplied by 2; (2) X_2 is multiplied by 2; (3) X_1 and X_2 are each multiplied by 2, are

$$d^2(2X_1, X_2) = (-2 + 3)^2 + (0 - 2)^2 + (-4 - 4)^2 = 69 \\ d^2(X_1, 2X_2) = (-1 + 6)^2 + (0 - 4)^2 + (-2 - 8)^2 = 141 \\ d^2(2X_1, 2X_2) = 4[(X_1 - X_2)'(X_1 - X_2)] = 4 \times 44 = 176.$$

Note that they are fully distorted as $69 \neq 4(44)$ and $141 \neq 4(44)$. Thus, the scaling of individual vectors can fully alter the nature of the clusters when there are clusters in the original data. As well, members of the original clusters need not be members of the same clusters in the scaled data and the number of clusters may also change. Accordingly, it is indeed inadvisable to make use of the generalized distance. Nor is re-scaling the individual vectors a good idea if we are seeking clusters. Accordingly, the recommended procedure consists of utilizing the original data without modifying them. It may also happen that the components in each p -vector are recorded in different units of measurement. Then, how

to eliminate the location and scale effect on the components in each vector? This can be achieved by standardizing them individually, that is, by subtracting the average value of the components from the components of each vector and dividing the result by the sample standard deviation. Let us see what happens in the case of our numerical example. Letting \bar{x}_1 and \bar{x}_2 be the averages of the components in X_1 and X_2 , and s_1^2 and s_2^2 be the associated sums of products, we have

$$\begin{aligned}\bar{x}_1 &= \frac{1}{3}[(-1) + (0) + (-2)] = -1, \quad \bar{x}_2 = \frac{1}{3}[(-3) + (2) + (4)] = 1, \\ s_1^2 &= \sum_{i=1}^p (x_{i1} - \bar{x}_1)^2 = [(-1) - (-1)]^2 + [(0) - (-1)]^2 + [(-2) - (-1)]^2 = 2, \\ s_2^2 &= \sum_{i=1}^p (x_{i2} - \bar{x}_2)^2 = 26.\end{aligned}$$

Thus, the standardized vectors X_1 and X_2 , denoted by Y_1 and Y_2 , are the following:

$$Y_1 = \frac{\sqrt{3}}{\sqrt{2}} \begin{bmatrix} -1 - (-1) \\ 0 - (-1) \\ -2 - (-1) \end{bmatrix} = \frac{\sqrt{3}}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad Y_2 = \frac{\sqrt{3}}{\sqrt{26}} \begin{bmatrix} -4 \\ 1 \\ 3 \end{bmatrix},$$

and $d^2(Y_1, Y_2) = (Y_1 - Y_2)'(Y_1 - Y_2) = 7.6641$. However, Y_1 and Y_2 are very distorted and the distance between X_1 and X_2 is also modified. Hence, such procedures will change the clustering aspect as well, with new clusters possibly differing from the original clusters.

Let us consider the matrix of squared distances, denoted by D :

$$D = \begin{bmatrix} 0 & d_{12}^2 & \cdots & d_{1n}^2 \\ d_{21}^2 & 0 & \cdots & d_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^2 & d_{n2}^2 & \cdots & d_{nn}^2 \end{bmatrix} = D'. \quad (15.1.3)$$

For example, letting

$$X_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad X_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad X_4 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix},$$

we have $d_{12}^2 = (1 - 2)^2 + (0 - 1)^2 + (-1 - 3)^2 = 18$, $d_{13}^2 = 11$, $d_{14}^2 = 14$, $d_{23}^2 = 5$, $d_{24}^2 = 2$, $d_{34}^2 = 9$, so that

$$D = \begin{bmatrix} 0 & 18 & 11 & 14 \\ 18 & 0 & 5 & 2 \\ 11 & 5 & 0 & 9 \\ 14 & 2 & 9 & 0 \end{bmatrix}.$$

The question of interest is the following: Given a set of n vectors of order p , how can one determine the number of clusters and then, classify them into these clusters?

15.2. Different Methods of Clustering

The main methods are hierarchical in nature, the other ones being non-hierarchical. We will begin with non-hierarchical techniques. In this category, the most popular one involves optimization or partitioning.

15.2.1. Optimization or partitioning

With this approach, we have to come up with two numbers: k , a probable number of clusters, and r , the maximum separation between the members of each prospective cluster. Based on the distances or on the dissimilarity matrix, D , one should be able to determine the likely number of clusters, that is, k . Then, one has to find a set of k vectors among the n given vectors, which will be taken as seed members or starting members within the k potential clusters. Several methods have been proposed for determining this k , including the following:

1. Examine the closeness of the original vectors as indicated by the dissimilarity matrix D and, to start with, decide on an initial numbers for k and the likely distance between members within a cluster denoted by r .
2. Examine the original data points or original p -vectors and, based on the comparative magnitudes of the components of the observed p -vectors, ascertain whether there is any grouping possible and predict a value for each of k and r .
3. Evaluate the sample sum of products matrix S from the original data matrix. Compute the two main principal components associated with this S . Substitute X_j , the j -th observation vector, in the two principal components. This provides a pair of numbers or one point in a two-dimensional space. Compute n such points for $j = 1, \dots, n$. Plot these points. From the graph, assess the clustering pattern, the number k of possible clusters, estimates for r , the maximum distance between two members within a cluster as well as the minimum distance between the clusters.

4. Choose any number k , select k vectors at random from the set of n vectors; then, preselect a number r and use it as a measure of maximum separation between vectors.
5. Take any number k and select as seed vectors the first k vectors whose separation is at least two units among the set of n vectors.
6. Look at the farthest points. Select k of them that are separated by at least r units for preselected values of k and r .

If the dissimilarity matrix D is utilized, then the separation number r must be measured in d_{ij}^2 units, whereas r should be in d_{ij} units if the actual distances d_{ij} are used. After the seed vectors are selected, the remaining $n - k$ points are to be associated to these seed points to form clusters. Assign the vectors closest to each of the seed vectors and form the initial k clusters of two or more vectors. For example, if there are three closest members at equal distance to a seed vector then, that cluster comprises 4 members, including the seed vector. Then, compute the centroids of all initial clusters. The centroid of a cluster is the simple average of the vectors included in that cluster. Thus, the centroid is a p -vector. Then, measure the distances of all the points belonging to the same cluster from each centroid, and incorporate all points within the distance of r from a centroid to that cluster. This process will create the second stage of k clusters. Now, evaluate the centroid of each of these k clusters. Again, repeat the process of computing the distances of all points from each centroid. If a member in a cluster is found to be closer to the centroid of another cluster than to its own cluster's centroid, then redirect that vector to the cluster to which it belongs. Rearrange all vectors in such a manner, assigning each one to a cluster whose centroid is the closest. Note that the number k can increase or decrease in the course of this process. Continue the procedure until no more improvement is possible. At this stage, that final k is the number of clusters in the data and the final members in each cluster are set. This procedure is also called *k-means approach*.

This k-means approach has a serious shortcoming: if one starts with a different set of seed vectors, then it is possible to end up with a different set of final clusters. On the other hand, this method has the appreciable advantage that it allows a member provisionally assigned to a cluster to be moved to another cluster where it really belongs, that is, it allows the transfer of points. The following example should clarify the procedure.

Example 15.2.1. Ten volunteers are given an exercise routine in an experiment that monitors systolic pressure, diastolic pressure and heart beat. These are measured after adhering to the exercise routine for four weeks. The data entries are systolic pressure minus

120 (SP), diastolic pressure minus 80 (DP) and heart beat minus 60 (HB), where 120, 80 and 60 are taken as the standard readings of systolic pressure, diastolic pressure and heart beat, respectively. Carry out a cluster analysis of the data. The data matrix is the following where (1), ..., (10) represent the data vectors A_1, \dots, A_{10} for the 10 volunteers, the first row represents SP, the second row, DP, and the third, HB:

$\downarrow \rightarrow$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>SP</i> :	0	1	1	2	3	4	6	8	5	10
<i>DP</i> :	1	0	-1	3	2	3	8	10	6	8
<i>HB</i> :	-1	-1	-1	-2	5	2	7	8	9	4

Solution 15.2.1. Let us compute the dissimilarity matrix D :

$$D = \begin{bmatrix} \downarrow \rightarrow & (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) & (10) \\ (1) & 0 & 2 & 5 & 9 & 46 & 29 & 149 & 226 & 150 & 174 \\ (2) & 2 & 0 & 1 & 11 & 44 & 27 & 153 & 230 & 152 & 170 \\ (3) & 5 & 1 & 0 & 18 & 49 & 34 & 170 & 251 & 165 & 187 \\ (4) & 9 & 11 & 18 & 0 & 51 & 20 & 122 & 185 & 139 & 125 \\ (5) & 46 & 44 & 49 & 51 & 0 & 11 & 49 & 98 & 36 & 86 \\ (6) & 29 & 27 & 34 & 20 & 11 & 0 & 54 & 101 & 59 & 65 \\ (7) & 149 & 153 & 170 & 122 & 49 & 54 & 0 & 9 & 9 & 25 \\ (8) & 226 & 230 & 251 & 185 & 98 & 101 & 9 & 0 & 26 & 24 \\ (9) & 150 & 152 & 165 & 139 & 36 & 59 & 9 & 26 & 0 & 54 \\ (10) & 174 & 170 & 187 & 125 & 86 & 65 & 25 & 24 & 54 & 0 \end{bmatrix}.$$

The data matrix suggests the possibility of three clusters. Accordingly, we may begin with the vectors A_2 , A_5 and A_8 as seed vectors and take the separation width as $r = 15$ units. From D , we find $d_{23}^2 = 1$, the smallest number, and hence A_2 and A_3 form the cluster: $\{A_2, A_3\}$. Note that $d_{56}^2 = 11$, so that A_6 and A_5 form the cluster: $\{A_5, A_6\}$. Since $d_{87}^2 = 9$, A_7 and A_8 form a cluster: $\{A_7, A_8\}$. Now, consider the centroids. Letting C_{11} , C_{21} and C_{31} denote the centroids, $C_{11} = \frac{1}{2}(A_2 + A_3)$, $C_{21} = \frac{1}{2}(A_5 + A_6)$ and $C_{31} = \frac{1}{2}(A_7 + A_8)$, that is,

$$C_{11} = \begin{bmatrix} 1 \\ -1/2 \\ -1 \end{bmatrix}, \quad C_{21} = \begin{bmatrix} 7/2 \\ 5/2 \\ 7/2 \end{bmatrix} \quad \text{and} \quad C_{31} = \begin{bmatrix} 7 \\ 9 \\ 15/2 \end{bmatrix}.$$

Let us calculate the distances of A_1, \dots, A_{10} from C_{11}, C_{21}, C_{31} :

$$\begin{aligned} d^2(C_{11}, A_1) &= \frac{13}{4}, \quad d^2(C_{11}, A_2) = \frac{1}{4}, \quad d^2(C_{11}, A_3) = \frac{1}{4}, \quad d^2(C_{11}, A_4) = \frac{57}{4}, \\ d^2(C_{11}, A_5) &= \frac{185}{2}, \quad d^2(C_{11}, A_6) = \frac{121}{4}, \quad d^2(C_{11}, A_7) = \frac{645}{4}, \quad d^2(C_{11}, A_8) = \frac{961}{4}, \\ d^2(C_{11}, A_9) &= \frac{633}{4}, \quad d^2(C_{11}, A_{10}) = \frac{713}{4}, \quad d^2(C_{21}, A_1) = \frac{139}{4}, \quad d^2(C_{21}, A_2) = \frac{131}{4}, \\ d^2(C_{21}, A_3) &= \frac{155}{4}, \quad d^2(C_{21}, A_4) = \frac{131}{4}, \quad d^2(C_{21}, A_5) = \frac{11}{4}, \quad d^2(C_{21}, A_6) = \frac{11}{4}, \\ d^2(C_{21}, A_7) &= \frac{195}{4}, \quad d^2(C_{21}, A_8) = \frac{387}{4}, \quad d^2(C_{21}, A_9) = \frac{179}{4}, \quad d^2(C_{21}, A_{10}) = \frac{291}{4}, \\ d^2(C_{31}, A_1) &= \frac{741}{4}, \quad d^2(C_{31}, A_2) = \frac{757}{4}, \quad d^2(C_{31}, A_3) = \frac{833}{4}, \quad d^2(C_{31}, A_4) = \frac{605}{4}, \\ d^2(C_{31}, A_5) &= \frac{285}{4}, \quad d^2(C_{31}, A_6) = \frac{301}{4}, \quad d^2(C_{31}, A_7) = \frac{9}{4}, \quad d^2(C_{31}, A_8) = \frac{9}{4}, \\ d^2(C_{31}, A_9) &= \frac{61}{4}, \quad d^2(C_{31}, A_{10}) = \frac{89}{4}. \end{aligned}$$

We include all the points located within 15 units of distance to the nearest cluster. Then, the second set of clusters are the following: Cluster 1: $\{A_1, A_2, A_3, A_4\}$, cluster 2: $\{A_5, A_6\}$, cluster 3: $\{A_7, A_8\}$. Note that A_9 is quite close to Cluster 3. We may either include it in Cluster 3 or treat it as a singleton. Since the next stage calculations do not change the composition of the clusters, we may take the final clusters as $\{A_1, A_2, A_3, A_4\}$, $\{A_5, A_6\}$, $\{A_7, A_8, A_9\}$ and $\{A_{10}\}$ where Cluster 4 consists of a single element. This completes the computations.

Let us examine the principal components of the sample sum of products matrix and plot the points to see whether any cluster can be detected. The sample matrix denoted by \mathbf{X} and the sample average, denoted by $\bar{\mathbf{X}}$, are the following:

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 1 & 2 & 3 & 4 & 6 & 8 & 5 & 10 \\ 1 & 0 & -1 & 3 & 2 & 3 & 8 & 10 & 6 & 8 \\ -1 & -1 & -1 & -2 & 5 & 2 & 7 & 8 & 9 & 4 \end{bmatrix}, \quad \bar{\mathbf{X}} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}.$$

Let the matrix of sample averages be $\bar{\mathbf{X}} = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p]$ and the deviation matrix be $\mathbf{X}_d = \mathbf{X} - \bar{\mathbf{X}}$. Then,

$$\mathbf{X}_d = \begin{bmatrix} -4 & -3 & -3 & -2 & -1 & 0 & 2 & 4 & 1 & 6 \\ -3 & -4 & -5 & -1 & -2 & -1 & 4 & 6 & 2 & 4 \\ -4 & -4 & -4 & -5 & 2 & -1 & 4 & 5 & 6 & 1 \end{bmatrix},$$

and the sample sum of products matrix is $S = \mathbf{X}_d \mathbf{X}'_d$, that is,

$$S = \begin{bmatrix} 96 & 101 & 88 \\ 101 & 128 & 112 \\ 88 & 112 & 156 \end{bmatrix}.$$

The eigenvalues of S are $\lambda_1 = 330.440$, $\lambda_2 = 40.522$ and $\lambda_3 = 9.039$. An eigenvector corresponding to $\lambda_1 = 330.440$ and an eigenvector corresponding to $\lambda_2 = 40.522$, respectively denoted by U_1 and U_2 are the following:

$$U_1 = \begin{bmatrix} 0.782 \\ 0.943 \\ 1.000 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} -0.676 \\ -0.500 \\ -1.000 \end{bmatrix}.$$

Then the first two principal components are $U'_1 Y$ and $U'_2 Y$ with $Y' = [y_1, y_2, y_3]$. We substitute our sample points A_1, \dots, A_{10} to obtain 10 pairs of numbers. For example,

$$U'_1 A_1 = [0.782, 0.943, 1] \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = -0.057, \quad U'_2 A_1 = [-0.676, -0.5, 1] \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = -1.5$$

and hence the first pair of numbers or the first point is $P_1 : (-0.057, -1.500)$. Similar calculations yield the remaining 9 points as: $P_2 : (-0.218, -1.676)$, $P_3 : (-1.161, -1.176)$, $P_4 : (2.393, -4.852)$, $P_5 : (9.232, 1.972)$, $P_6 : (7.957, -2.204)$, $P_7 : (19.236, -1.056)$, $P_8 : (23.686, -2.408)$, $P_9 : (18.568, 2.620)$, $P_{10} : (19.364, -6.760)$. It is seen that these points which are plotted in Fig. 15.2.2 form the same clusters as the original points shown in Fig. 15.2.1, that is, Cluster 1: $\{A_1, A_2, A_3, A_4\}$; Cluster 2: $\{A_5, A_6\}$; Cluster 3: $\{A_7, A_8, A_9\}$; Cluster 4: $\{A_{10}\}$.

Other non-hierarchical methods are currently in use. We will mention these procedures later, after discussing the main hierarchical technique known as *single linkage* or nearest neighbor method.

15.3. Hierarchical Methods of Clustering

Hierarchical procedures are of two categories. In one of them, we begin with all the n data points as n different clusters of one element each. Then, by applying certain rules, we start combining these single-member clusters into larger clusters, the process being halted when a desired number of clusters are obtained. If the process is continued, we ultimately end up with a single cluster containing all of the n points. In the second category, we initially consider one cluster that comprises the n elements. We then start splitting this cluster into two clusters by making use of some criteria. Next, one or both of these sub-clusters are divided again by applying the same criteria. If the process is continued, we

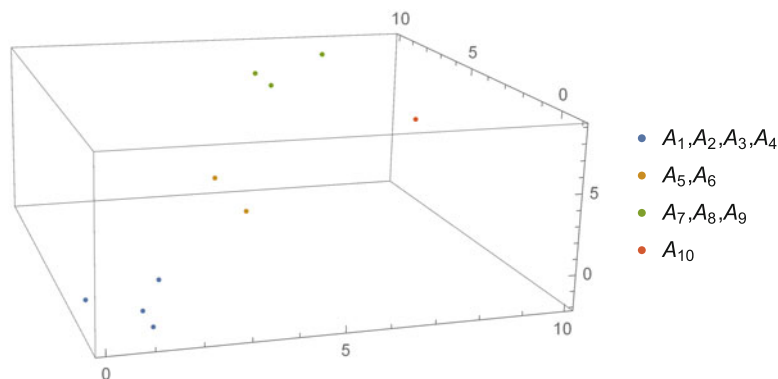


Figure 15.2.1 The original 10 data points

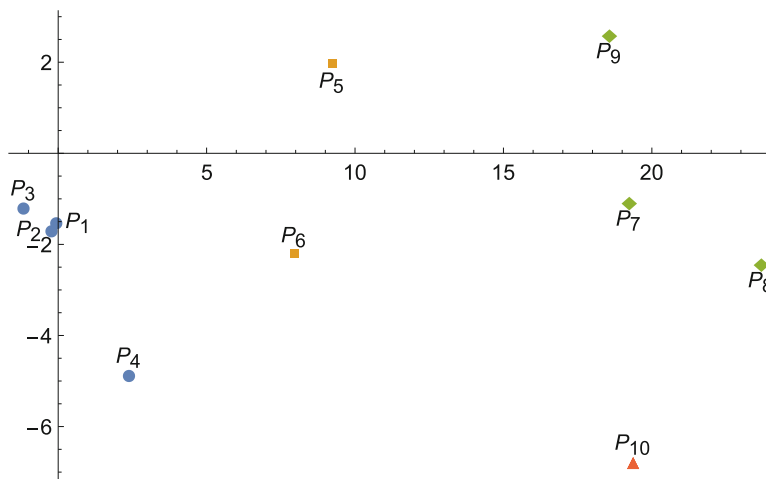


Figure 15.2.2 Second versus first principal component evaluated at the A_i 's

finally end up with n clusters of one element each. The process is halted when a desired number of clusters are obtained. In all these procedures, one cannot objectively determine when to stop the process or how many distinct clusters are present. We have to specify some stopping rules as a means of selecting a suitable number of clusters.

15.3.1. Single linkage or nearest neighbor method

In this single linkage procedure, we begin by assuming that there are n clusters consisting of one item each. We then combine these clusters by applying a minimum distance rule. At the initial stage, we have only one element in each 'cluster', but at the following steps, each cluster will potentially contain several items and hence, the rule is stated for general clusters. Consider two clusters A and B whose elements are denoted by X_j and

Y_j , that is, $X_j \in A$ and $Y_j \in B$, the X_j 's and Y_j 's being p -vectors belonging to the data set at hand. In the minimum distance rule, we define the distance between two clusters, denoted by $d(A, B)$, as follows:

$$d(A, B) = \min\{d(X_i, Y_j), \text{ for all } X_i \in A, Y_j \in B\}. \tag{15.3.1}$$

This distance is measured in the units of the definition of the distance being utilized. We will illustrate the single linkage hierarchical procedure by making use of the data set provided in Example 15.2.1 and its associated dissimilarity matrix D . We will utilize the dissimilarity matrix D to represent various "distances". Since the matrix D will be repeatedly referred to at every stage, it is duplicated next for ready reference:

$$D = \begin{bmatrix} \downarrow \rightarrow & (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8) & (9) & (10) \\ (1) & 0 & 2 & 5 & 9 & 46 & 29 & 149 & 226 & 150 & 174 \\ (2) & 2 & 0 & 1 & 11 & 44 & 27 & 153 & 230 & 152 & 170 \\ (3) & 5 & 1 & 0 & 18 & 49 & 34 & 170 & 251 & 165 & 187 \\ (4) & 9 & 11 & 18 & 0 & 51 & 20 & 122 & 185 & 139 & 125 \\ (5) & 46 & 44 & 49 & 51 & 0 & 11 & 49 & 98 & 36 & 86 \\ (6) & 29 & 27 & 34 & 20 & 11 & 0 & 54 & 101 & 59 & 65 \\ (7) & 149 & 153 & 170 & 122 & 49 & 54 & 0 & 9 & 9 & 25 \\ (8) & 226 & 230 & 251 & 185 & 98 & 101 & 9 & 0 & 26 & 24 \\ (9) & 150 & 152 & 165 & 139 & 36 & 59 & 9 & 26 & 0 & 54 \\ (10) & 174 & 170 & 187 & 125 & 86 & 65 & 25 & 24 & 54 & 0 \end{bmatrix}.$$

To start with, we have 10 clusters $\{A_j\}$, $j = 1, \dots, 10$. At the initial stage, each cluster has one element. Then $d(A, B)$ as defined in (15.3.1) is the smallest distance (dissimilarity) appearing in D , that is, 1 which occurs between the elements corresponding to A_2 and A_3 . These two clusters of one vector each are combined and replaced by B_1 by taking the smaller entries in each column of the combined representation of the dissimilarity measures corresponding A_2 and A_3 . For illustration, we now list the dissimilarity measures corresponding to the original A_2 and A_3 and the new B_1 as rows:

$$\begin{array}{l} A_2 : (2) \quad [0 \quad 1] \quad (11) \quad (44) \quad (27) \quad (153) \quad (230) \quad (152) \quad (170) \\ A_3 : 5 \quad [1 \quad 0] \quad 18 \quad 49 \quad 34 \quad 170 \quad 251 \quad 165 \quad 187 \\ B_1 : 2 \quad [0] \quad 11 \quad 44 \quad 27 \quad 153 \quad 230 \quad 152 \quad 170 \end{array}$$

The rows representing A_2 and A_3 are combined and replaced by B_1 as shown above. The second and third columns in D are combined into one column, namely, the B_1 column. The elements in B_1 are the smaller elements in each column of A_2 and A_3 . The bracketed

elements in A_2 and A_3 , namely $[0, 1]$ and $[1, 0]$, are combined into one element $[0]$ in B_1 , the updated dissimilarity matrix having one fewer row and one fewer column. These are the intersections of the two rows and columns. Other smaller elements in the two original columns, which make up B_1 , are displayed in parentheses. This process will be repeated at each stage. At the first stage of the procedure, we end up with 9 clusters: $C_1 = \{A_2, A_3\}, \{A_j\}, j = 1, 4, \dots, 10$, the resulting configuration of the dissimilarity matrix, denoted by D_1 , being

$$D_1 = \begin{bmatrix} \downarrow \rightarrow & A_1 & B_1 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ A_1 & 0 & 2 & 9 & 46 & 29 & 149 & 226 & 150 & 174 \\ B_1 & 2 & 0 & 11 & 44 & 27 & 153 & 230 & 152 & 170 \\ A_4 & 9 & 11 & 0 & 51 & 20 & 122 & 185 & 139 & 125 \\ A_5 & 46 & 44 & 51 & 0 & 11 & 49 & 98 & 36 & 86 \\ A_6 & 29 & 27 & 20 & 11 & 0 & 54 & 101 & 59 & 65 \\ A_7 & 149 & 153 & 122 & 49 & 54 & 0 & 9 & 9 & 25 \\ A_8 & 226 & 230 & 185 & 98 & 101 & 9 & 0 & 26 & 24 \\ A_9 & 150 & 152 & 139 & 36 & 59 & 9 & 26 & 0 & 54 \\ A_{10} & 174 & 170 & 125 & 86 & 65 & 25 & 24 & 54 & 0 \end{bmatrix}.$$

Now, the next smallest dissimilarity is 2 which occurs at (A_1, B_1) . Thus, the rows (columns) corresponding to A_1 and B_1 are combined into one row (column) B_2 . The original rows corresponding to A_1 and B_1 and the new row corresponding to B_2 are the following:

$$\begin{array}{l} A_1 : [0 \quad 2] \quad (9) \quad 46 \quad 29 \quad (149) \quad (226) \quad (150) \quad 174 \\ B_1 : [2 \quad 0] \quad 11 \quad (44) \quad (27) \quad 153 \quad 230 \quad 152 \quad (170) \\ B_2 : [0] \quad \quad 9 \quad 44 \quad 27 \quad 149 \quad 226 \quad 150 \quad 170. \end{array}$$

The new configuration, denoted by D_2 , is the following:

$$D_2 = \begin{bmatrix} \downarrow \rightarrow & B_2 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ B_2 & 0 & 9 & 44 & 27 & 149 & 226 & 150 & 170 \\ A_4 & 9 & 0 & 51 & 20 & 122 & 185 & 139 & 125 \\ A_5 & 44 & 51 & 0 & 11 & 49 & 98 & 36 & 86 \\ A_6 & 27 & 20 & 11 & 0 & 54 & 101 & 59 & 65 \\ A_7 & 149 & 122 & 49 & 54 & 0 & 9 & 9 & 25 \\ A_8 & 226 & 185 & 98 & 101 & 9 & 0 & 26 & 24 \\ A_9 & 150 & 139 & 36 & 59 & 9 & 26 & 0 & 54 \\ A_{10} & 170 & 125 & 86 & 65 & 25 & 24 & 54 & 0 \end{bmatrix},$$

the resulting clusters being $C_2 = \{A_1, A_2, A_3\}, \{A_j\}, j = 4, \dots, 10$. The next smallest dissimilarity is 9, which occurs at (B_2, A_4) . Hence these are combined, that is, the first

two columns (rows) are merged as explained. The combined row, denoted by B_3 , is the following, its transpose becoming the first column:

$$B_3 = [0, 44, 20, 122, 185, 139, 125],$$

and the new configuration is the following:

$$D_3 = \begin{bmatrix} \downarrow \rightarrow & B_3 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ B_3 & 0 & 44 & 20 & 122 & 185 & 139 & 125 \\ A_5 & 44 & 0 & 11 & 49 & 98 & 36 & 86 \\ A_6 & 20 & 11 & 0 & 54 & 101 & 59 & 65 \\ A_7 & 122 & 49 & 54 & 0 & 9 & 9 & 25 \\ A_8 & 185 & 98 & 101 & 9 & 0 & 26 & 24 \\ A_9 & 139 & 36 & 59 & 9 & 26 & 0 & 54 \\ A_{10} & 125 & 86 & 65 & 25 & 24 & 54 & 0 \end{bmatrix}.$$

At this stage, the clusters are $C_2 = \{A_1, A_2, A_3, A_4\}$, $\{A_j\}$, $j = 5, \dots, 10$. The next smallest number is 9, which is occurring at (A_7, A_8) , (A_7, A_9) . Accordingly, we combine A_7, A_8 and A_9 , and the resulting configuration is the following where the resultant of the replacement rows (columns) is denoted by B_4 :

$$D_4 = \begin{bmatrix} \downarrow \rightarrow & B_3 & A_5 & A_6 & B_4 & A_{10} \\ B_3 & 0 & 44 & 20 & 122 & 125 \\ A_5 & 44 & 0 & 11 & 36 & 86 \\ A_6 & 20 & 11 & 0 & 54 & 65 \\ B_4 & 122 & 36 & 54 & 0 & 24 \\ A_{10} & 125 & 86 & 65 & 24 & 0 \end{bmatrix},$$

the clusters being $C_2 = \{A_1, A_2, A_3, A_4\}$, $C_3 = \{A_7, A_8, A_9\}$, $\{A_i\}$, $i = 5, 6, 10$. The next smallest dissimilarity measure is 11 at (A_5, A_6) . Combining these, the replacement row is $B_5 = [20, 0, 36, 65]$, and the new configuration, denoted by D_5 is as follows:

$$D_5 = \begin{bmatrix} \downarrow \rightarrow & B_3 & B_5 & B_4 & A_{10} \\ B_3 & 0 & 20 & 122 & 125 \\ B_5 & 20 & 0 & 36 & 65 \\ B_4 & 122 & 36 & 0 & 24 \\ A_{10} & 125 & 65 & 24 & 0 \end{bmatrix},$$

the resulting clusters being $C_2 = \{A_1, A_2, A_3, A_4\}$, $C_3 = \{A_7, A_8, A_9\}$, $C_4 = \{A_5, A_6\}$, $C_5 = \{A_{10}\}$.

We may stop at this stage since the clusters obtained from the other methods coincide with C_2, C_3, C_4, C_5 . At the following step of the procedure, C_4 would combine with C_3 , with the next final stage resulting in a single cluster that would encompass all 10 points.

15.3.2. Average linking as a modified distance measure

An alternative distance measure involving all the items in pairs of clusters is considered in this subsection. As one proceeds from any stage to the next one in a hierarchical procedure, a decision is based on the next smallest distance between two clusters. At the initial stage, this does not pose any problem since the dissimilarity matrix D is available and each cluster contains only a single element. However, further on in the process, as there are several elements in the clusters, a more suitable definition of “distance” is required in order to proceed to the next stage. Several types of methods have been proposed in the literature. One such procedure is the *average linkage method* under which the distance between two clusters A and B , denoted again by $d(A, B)$, is defined as follows:

$$d(A, B) = \frac{1}{n_1 n_2} \sum_{j=1}^{n_2} \sum_{i=1}^{n_1} d(X_i, Y_j) \text{ for all } X_i \in A, Y_j \in B \quad (15.3.2)$$

where the X_i 's and Y_j 's are all p -vectors from the given set of data points. In this case, the rule being applied is that two clusters having the smallest distance, as measured in terms of (15.3.2), are combined before initiating the next stage.

15.3.3. The centroid method

In a hierarchical single linkage procedure, another way of determining the distance between two clusters before proceeding to the next stage is referred to as the centroid method under which the Euclidean distance between the centroids of clusters A and B is defined as follows:

$$d(A, B) = d(\bar{X}, \bar{Y}) \text{ with } \bar{X} = \frac{1}{n_1} \sum_{j=1}^{n_1} X_j \text{ and } \bar{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j, \quad (15.3.3)$$

where \bar{X} is the centroid of the cluster A and \bar{Y} is the centroid of the cluster B , $X_i \in A$, $i = 1, \dots, n_1$, $Y_j \in B$, $j = 1, \dots, n_2$. In this case, the process involves combining two clusters with the smallest $d(A, B)$ as specified in (15.3.3) into a single cluster. After combining them, or equivalently, after taking the union of A and B , the centroid of the combined cluster, denoted by \bar{Z} , is

$$\bar{Z} = \frac{n_1 \bar{X} + n_2 \bar{Y}}{n_1 + n_2} = \frac{1}{n_1 + n_2} \sum_{j=1}^{n_1+n_2} Z_j, \quad Z_j \in A \cup B,$$

where the Z_j 's are the original vectors that were included in A or B .

15.3.4. The median method

A main shortcoming of the centroid method of joining two clusters is that if n_1 is very large compared to n_2 , then \bar{Z} is likely to be closer of \bar{X} , and vice versa. In order to avoid this type of imbalance, a method based on the median is suggested, under which the median of the combined clusters A and B is defined as

$$\text{Median}_{A \cup B} = \frac{1}{2}(\bar{X} + \bar{Y}) \text{ with } X_i \in A \text{ and } Y_j \in B, \quad (15.3.4)$$

for all i, j and r . In this process, the clusters A and B for which $\text{Median}_{A \cup B}$ is the smallest are combined to form the next cluster whose elements are the Z_r 's, $Z_r \in A \cup B$.

15.3.5. The residual sum of products method

From the one-way MANOVA layout, a residual or within group (within cluster) sum of products for clusters A , B and $A \cup B$, denoted by R_A , R_B and $R_{A \cup B}$, are the following:

$$R_A = \sum_{i=1}^{n_1} (X_i - \bar{X})'(X_i - \bar{X}), \quad R_B = \sum_{j=1}^{n_2} (Y_j - \bar{Y})'(Y_j - \bar{Y})$$

$$R_{A \cup B} = \sum_{r=1}^{n_1+n_2} (Z_r - \bar{Z})'(Z_r - \bar{Z}), \quad Z_j \in A \cup B, \quad \bar{Z} = \frac{n_1 \bar{X} + n_2 \bar{Y}}{n_1 + n_2}.$$

Once those sums of squares have been evaluated, we compute the quantity

$$T_{A \cup B} = R_{A \cup B} - (R_A + R_B), \quad (15.3.5)$$

which can be interpreted as the increase in residual sum of products due to the process of merging the clusters A and B . Then, the procedure consists of combining those clusters A and B for which $T_{A \cup B}$ as defined in (15.3.5) is the minimum. This method is also called *Ward's method*.

There exist other methods for combining clusters such as the *flexible beta method*, and several comparative studies point out the merits and drawbacks of the various methods.

In the hierarchical procedures considered in Sect. 15.3, we begin with the n data points as n distinct clusters of one element each. Then, by applying certain "minimum distance" methods, "distance" being defined in different ways, we combined the clusters one by one. We may also consider a hierarchical procedure wherein the n data points are treated as one cluster of n elements. At this stage, by making use of some rules, we break up this cluster into two clusters. Then, one of these or both are split again as two clusters by applying the same rule. We continue the process and stop it when it is determined that there is a

sufficient number of clusters. If the process is not halted at a certain stage, we will end up with a single cluster containing all of the n elements or points. We will not elaborate further on such procedures.

15.3.6. Other criteria for partitioning or optimization

In Sect. 15.2, we considered a non-hierarchical procedure known as the *k-means method*, which is the most popular in this area. After discussing this, we described the most widely utilized non-hierarchical procedure in Sect. 15.3. We will now examine other non-hierarchical procedures in common use. Some of these are connected with the MANOVA or multivariate analysis of variation of a one-way classification. In a multivariate one-way layout, let X_{ij} be the j -th vector in the i -th group or i -th cluster, all vectors being p -vectors or $p \times 1$ real vectors. Let there be k groups (k clusters) of sizes n_1, \dots, n_k with $n_1 + n_2 + \dots + n_k = n. = n$, that is, the cluster sizes are n_1, \dots, n_k , respectively. Let the residual sum of products or sum of squares and cross products matrix be denoted by U , which is $p \times p$. This matrix U is also called *within group or within cluster variation matrix*. Let the *between groups or between clusters variation matrix* be V . In this setup, U and V are the following:

$$U = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)', \quad \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad (15.3.6)$$

$$V = \sum_{i,j} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})', \quad \bar{X} = \frac{1}{n.} \sum_{i,j} X_{ij}. \quad (15.3.7)$$

Then, under the hypothesis that the group effects or cluster effects are the same, and under the normality assumption on the X_{ij} 's, U and V are independently distributed Wishart matrices with $n. - k$ and $k - 1$ degrees of freedom, respectively, where $\Sigma > O$ is the parameter matrix in the Wishart densities as well as the common covariance matrix of the X_{ij} 's, referring to Chap. 5. Thus, $W_1 = (U + V)^{-\frac{1}{2}} U (U + V)^{-\frac{1}{2}}$ is a real matrix-variate type-1 beta random variable with the parameters $(\frac{n.-k}{2}, \frac{k-1}{2})$, $W_2 = U^{-\frac{1}{2}} V U^{-\frac{1}{2}}$ is a real matrix-variate type-2 beta random variable with the parameters $(\frac{k-1}{2}, \frac{n.-k}{2})$ and $W_3 = U + V$ follows a real Wishart distribution having $n. - 1$ degrees of freedom and parameter matrix $\Sigma > O$, again referring Chap. 5. Observe that both U and V are real positive definite matrices, so that all of their eigenvalues are positive. The likelihood ratio criterion λ for testing the hypothesis that the group effects are the same is the following:

$$\lambda^{\frac{2}{n}} = \frac{|U|}{|U + V|} = |W_1| = \frac{1}{|I + U^{-\frac{1}{2}} V U^{-\frac{1}{2}}|} = \frac{1}{|I + W_2|}. \quad (15.3.8)$$

We are aiming to have the within cluster variation small and the between cluster variation large, which means, in some sense, that U will be small and V will be large, in which case λ as given in (15.3.8) will be small. This also means that the trace of U must be small and trace of W_2 must be large. Accordingly, a few criteria for merging clusters are based on $\text{tr}(U)$, $|U|$ and $\text{tr}(W_2)$. The following are some commonly utilized criteria for combining clusters:

- (1) Minimizing $\text{tr}(U)$;
- (2) Minimizing $|U|$;
- (3) Maximizing $\text{tr}(W_2)$.

These criteria are applied as follows: One of the n observation vectors is moved to a selected cluster if $\text{tr}(U)$ is a minimum ($|U|$ is a minimum and $\text{tr}(W_2)$ is a maximum for the other criteria). Then, $\text{tr}(U)$ is evaluated after moving the observation vectors one by one to the selected cluster and, each time, $\text{tr}(U)$ is noted; the vector for which $\text{tr}(U)$ attains a minimum value belongs to the selected cluster, that is, it is combined with the selected cluster. Observe that

$$\begin{aligned} \text{tr}(U) &= \text{tr}\left(\sum_{i,j} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'\right) \\ &= \text{tr}\left(\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)'\right) + \cdots + \text{tr}\left(\sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)(X_{kj} - \bar{X}_k)'\right) \\ &= \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)'(X_{1j} - \bar{X}_1) + \cdots + \sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)'(X_{kj} - \bar{X}_k), \end{aligned} \quad (15.3.9)$$

owing to the property that, for two matrices P and Q , $\text{tr}(PQ) = \text{tr}(QP)$ as long as PQ and QP are defined. As well, observe that since $(X_{ij} - \bar{X}_i)'(X_{ij} - \bar{X}_i)$ is a scalar quantity for every i and j , it is equal to its trace. How does this criterion work in practice? Consider moving a member from the s -th cluster to the selected cluster, namely, the r -th cluster. The original centroids are \bar{X}_r and \bar{X}_s , and when one element is added to the r -th cluster from the s -th cluster, both centroids will respectively change to, say, \bar{X}_{r+1} and \bar{X}_{s-1} . Compute the updated sums of squares in the new r -th and s -th clusters. Then, add up all the sums of squares in all the clusters and obtain a new $\text{tr}(U)$. Carry out this process for every member in every other cluster and compute $\text{tr}(U)$ each time. Take the smallest value of $\text{tr}(U)$ thus calculated, including the original value of $\text{tr}(U)$, before considering

transferring any point. That vector for which $\text{tr}(U)$ is minimum really belongs to the r -th cluster and so, is included in it. Repeat the process until no more improvement can be made, at which point no more transfer of points is necessary.

Simplification of the computations of $\text{tr}(U)$

As will be explained, computing $\text{tr}(U)$ can be simplified. Consider the new sum of squares in the r -th cluster. Let the new and old sums of squares be denoted by $(\text{New})_r$, $(\text{New})_s$, and $(\text{Old})_r$, $(\text{Old})_s$, respectively. Let the vector transferred from the s -th cluster to the r -th cluster be denoted by Y . Then,

$$\begin{aligned}
 (\text{New})_r &= \sum_{j=1}^r (X_{rj} - \bar{X}_{r+1})'(X_{rj} - \bar{X}_{r+1}) + (Y - \bar{X}_{r+1})'(Y - \bar{X}_{r+1}) \\
 &= \sum_{j=1}^r (X_{rj} - \bar{X}_r + (\bar{X}_r - \bar{X}_{r+1}))'(X_{rj} - \bar{X}_r + (\bar{X}_r - \bar{X}_{r+1})) \\
 &\quad + (Y - \bar{X}_{r+1})'(Y - \bar{X}_{r+1}) \\
 &= \sum_{j=1}^r (X_{rj} - \bar{X}_r)'(X_{rj} - \bar{X}_r) + r(\bar{X}_r - \bar{X}_{r+1})'(\bar{X}_r - \bar{X}_{r+1}) \\
 &\quad + (Y - \bar{X}_{r+1})'(Y - \bar{X}_{r+1}) \\
 &= (\text{Old})_r + r(\bar{X}_r - \bar{X}_{r+1})'(\bar{X}_r - \bar{X}_{r+1}) + (Y - \bar{X}_{r+1})'(Y - \bar{X}_{r+1}).
 \end{aligned}$$

The difference between the new sum of squares and the old one is

$$\delta_1 = r(\bar{X}_r - \bar{X}_{r+1})'(\bar{X}_r - \bar{X}_{r+1}) + (Y - \bar{X}_{r+1})'(Y - \bar{X}_{r+1}).$$

Noting that

$$\bar{X}_r - \bar{X}_{r+1} = \bar{X}_r - \frac{r\bar{X}_r + Y}{r+1} = \frac{1}{r+1}[\bar{X}_r - Y] \text{ and } Y - \bar{X}_{r+1} = \frac{r}{r+1}[Y - \bar{X}_r],$$

δ_1 simplifies to

$$\delta_1 = \frac{r}{r+1}(Y - \bar{X}_r)'(Y - \bar{X}_r).$$

A similar procedure can be used for the s -th cluster. In that case, the new sum of squares can be written as

$$\begin{aligned} (\text{New})_s &= \sum_{j=1}^{s-1} (X_{sj} - \bar{X}_{s-1})'(X_{sj} - \bar{X}_{s-1}) \\ &= \sum_{j=1}^s (X_{ij} - \bar{X}_{s-1})'(X_{sj} - \bar{X}_{s-1}) - (Y - \bar{X}_{s-1})'(Y - \bar{X}_{s-1}). \end{aligned}$$

Then, proceeding as in the case of the r -th cluster and denoting the difference between the new and the old sums of squares as δ_2 , we have

$$\delta_2 = -\frac{s}{s-1}(Y - \bar{X}_s)'(Y - \bar{X}_s), \quad s > 1,$$

so that the sum of the differences between the new and old sums of squares, denoted by δ , is the following:

$$\delta = \delta_1 + \delta_2 = \frac{r}{r+1}(Y - \bar{X}_r)'(Y - \bar{X}_r) - \frac{s}{s-1}(Y - \bar{X}_s)'(Y - \bar{X}_s) \quad (15.3.10)$$

for $s > 1$, where \bar{X}_r and \bar{X}_s are the original centroids of the r -th and s -th clusters, respectively. As such, computing δ is very simple. Evaluate the quantity specified in (15.3.10) for all the points outside the r -th cluster and look for the minimum of δ , including the original value of $\delta = 0$. If the minimum occurs at a point Y_1 outside of the r -th cluster, then transfer that point to the r -th cluster. Continue the process for every vector in the s -th cluster and then, for $r = 1, \dots, k$, assuming there are k clusters, until $\delta = 0$. In the end, all the clusters are stabilized, and k may take on another value.

Among the three statistics $\text{tr}(U)$, $|U|$ and $\text{tr}(W_2)$, $\text{tr}(U)$ is the easiest to compute, as was just explained. However, if we consider a non-singular transformation, other than an orthonormal transformation, then $|U|$ and $\text{tr}(W_2)$ are invariant, but $\text{tr}(U)$ is not.

We have discussed one hierarchical methodology of single linkage nearest neighbor method and one non-hierarchical procedure consisting of the k -means method. These seem to be the most widely utilized. We also mentioned other hierarchical and non-hierarchical methods without going into the details. All these procedures are not well-defined mathematical procedures. None of the procedures can uniquely determine the clusters if there are some clusters in the multivariate data at hand, and none of the methods can uniquely determine the number of clusters. The advantages and shortcomings of the various methods will not be discussed so as not to confound the reader.

Exercises 15

15.1. For the $p \times 1$ vectors X_1, \dots, X_n , let the dissimilarity measures be (1) $d_{ij}^{(1)} = \sum_{k=1}^n |x_{ik} - x_{jk}|$, (2) $d_{ij}^{(2)} = \sum_{k=1}^n (x_{ik} - x_{jk})^2$, $X'_i = [x_{1i}, x_{2i}, \dots, x_{pi}]$. Compute the matrices (1) $(d_{ij}^{(1)})$; (2) $(d_{ij}^{(2)})$, for the following vectors:

$$X_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, X_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, X_3 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, X_4 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}.$$

15.2. Nine test runs $T - 1, \dots, T - 9$ are done to test the breaking strengths of three alloys. The following data are the deviations from the respective expected strengths:

$\downarrow \rightarrow$	$T - 1$	$T - 2$	$T - 3$	$T - 4$	$T - 5$	$T - 6$	$T - 7$	$T - 8$	$T - 9$
Alloy-1	0	-1	1	2	-1	2	5	4	5
Alloy-2	1	1	1	1	3	4	7	-4	8
Alloy-3	-1	0	1	2	2	3	8	4	-7

Carry out a cluster analysis by applying the following methods: (1) The single linkage or nearest neighbor method; (2) The average linkage method; (3) The centroid method; (4) The residual sum of products method.

15.3. Using the data provided in Exercise 15.2, carry out a cluster analysis by utilizing the following methods: (1) Partitioning or optimization; (2) Minimization of $\text{tr}(U)$; (3) Minimization of $|U|$; (4) Maximization of $\text{tr}(W_2)$ where U and W are given in Sect. 15.3.6.

15.4. Compare the results from the different methods in (1) Exercise 15.2; (2) Exercise 15.3, and make your observations.

15.5. Compare the results from the different methods in Exercises 15.2 and 15.3, and comment on the similarities and differences.

15.4. Correspondence Analysis

If the data at hand are classified according to two attributes, these characteristics may be of the same type, that is, both quantitative or both qualitative, or of different types, and whatever the types may be, we may construct a two-way contingency table. In a contingency table, the entries in the cells are frequencies or the number of times various combinations of the attributes appear. Correspondence Analysis is a process of identifying, quantifying, separating and plotting associations among the characteristics and relationships among the various levels. In a two-way contingency table, we identify, separate and plot associations between the two characteristics and attempt to identify relationships between row and column labels.

15.4.1. Two-way contingency table

Consider the following example. A random sample of 100 persons from a certain township are classified according to their educational level and their liberal disposition. In the frequency Table 15.4.1, the A_j 's represent their dispositions and the B_j 's, their educational levels, with $A_1 \equiv$ tolerant, $A_2 \equiv$ indifferent, $A_3 \equiv$ intolerant, $B_1 \equiv$ primary school education level, $B_2 \equiv$ high school education level, $B_3 \equiv$ bachelor's degree education level, $B_4 \equiv$ master's and higher degree education level.

Table 15.4.1: A two-way contingency table

$\downarrow \rightarrow$	B_1	B_2	B_3	B_4	Total
A_1	6	14	16	4	40
A_2	17	5	8	10	40
A_3	7	6	6	1	20
Total	30	25	30	15	100

There are 6 persons having a tolerant disposition and primary school level of education. There is one person with an intolerant disposition and a master's degree or a higher level of education, and so on. The marginal sums are also provided in the table. For example, the total number of persons having a primary school level of education is 30, the total number of persons having an intolerant disposition is 20, and so on. The corresponding relative frequencies (a given frequency divided by 100, the total frequency) are as follows (Table 15.4.2):

Table 15.4.2: Relative frequencies f_{ij} in the two-way contingency table

$\downarrow \rightarrow$	B_1	B_2	B_3	B_4	Total
A_1	0.06(f_{11})	0.14(f_{12})	0.16(f_{13})	0.04(f_{14})	0.40($f_{1.}$)
A_2	0.17(f_{21})	0.05(f_{22})	0.08(f_{23})	0.10(f_{24})	0.40($f_{2.}$)
A_3	0.07(f_{31})	0.06(f_{32})	0.06(f_{33})	0.01(f_{34})	0.20($f_{3.}$)
Total	0.30($f_{.1}$)	0.25($f_{.2}$)	0.30($f_{.3}$)	0.15($f_{.4}$)	1.00($f_{..}$)

The relative frequencies are denoted in parentheses by f_{ij} where the summation with respect to a subscript is designated by a dot, that is, $f_{i.} = \sum_j f_{ij}$, $f_{.j} = \sum_i f_{ij}$ and $f_{..} = \sum_i \sum_j f_{ij}$. Note that $f_{..} = 1$. In a general notation, a two-way contingency table and the corresponding relative frequencies are displayed as follows (Table 15.4.3):

Table 15.4.3: A two-way contingency table and a table of relative frequencies

$\downarrow \rightarrow$	B_1	B_2	\cdots	B_s	<i>Total</i>	$\downarrow \rightarrow$	B_1	B_2	\cdots	B_s	<i>Total</i>
A_1	n_{11}	n_{12}	\cdots	n_{1s}	$n_{1.}$	A_1	f_{11}	f_{12}	\cdots	f_{1s}	$f_{1.}$
A_2	n_{21}	n_{22}	\cdots	n_{2s}	$n_{2.}$	A_2	f_{21}	f_{22}	\cdots	f_{2s}	$f_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
A_r	n_{r1}	n_{r2}	\cdots	n_{rs}	$n_{r.}$	A_r	f_{r1}	f_{r2}	\cdots	f_{rs}	$f_{r.}$
<i>Total</i>	$n_{.1}$	$n_{.2}$	\cdots	$n_{.s}$	$n_{..} = n$	<i>Total</i>	$f_{.1}$	$f_{.2}$	\cdots	$f_{.s}$	$f_{..} = 1$

Letting the true probability of the occurrence of an observation in the (i, j) -th cell be p_{ij} , the following is the table of true probabilities:

Table 15.4.4: True probabilities p_{ij} in a two-way contingency table

$\downarrow \rightarrow$	B_1	B_2	\cdots	B_s	<i>Total</i>
A_1	p_{11}	p_{12}	\cdots	p_{1s}	$p_{1.}$
A_2	p_{21}	p_{22}	\cdots	p_{2s}	$p_{2.}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
A_r	p_{r1}	p_{r2}	\cdots	p_{rs}	$p_{r.}$
<i>Total</i>	$p_{.1}$	$p_{.2}$	\cdots	$p_{.s}$	$p_{..} = 1$

These are multinomial probabilities and, in this case, the n_{ij} 's become multinomial variables. An estimate of p_{ij} , denoted by \hat{p}_{ij} , is $\hat{p}_{ij} = f_{ij}$, the corresponding relative frequency. The marginal sums in Table 15.4.4 can be interpreted as follows: $p_{1.}$ = the probability of finding an item in the first row or the probability of an event will have the attribute A_1 ; $p_{.j}$ = the probability that an event will have the characteristic B_j , and so on. Thus,

$$\hat{p}_{ij} = f_{ij} = \frac{n_{ij}}{n}, \hat{p}_{i.} = \frac{n_{i.}}{n}, \hat{p}_{.j} = \frac{n_{.j}}{n}, i = 1, \dots, r, j = 1, \dots, s.$$

If A_i and B_j are respectively interpreted as the event that an observation will belong to the i -th row or the event of the occurrence of the characteristic A_i , and the event that an observation will belong to the j -th column or the event of the occurrence of the attribute B_j , and if we let $p_{i.} = P(A_i)$ and $p_{.j} = P(B_j)$, then $p_{ij} = P(A_i \cap B_j)$, where $P(A_i)$ is the probability of the event A_i , $P(B_j)$ is the probability of the event B_j , and $(A_i \cap B_j)$ is the intersection or joint occurrence of the events A_i and B_j . If A_i and B_j are independent events, $P(A_i \cap B_j) = P(A_i)P(B_j)$ or $p_{ij} = p_{i.}p_{.j}$, the product of the marginal probabilities or the marginal totals in the table of probabilities. That is,

$$P(A_i \cap B_j) = P(A_i)P(B_j) \Rightarrow p_{ij} = p_{i.}p_{.j}, \hat{p}_{ij} = \left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right) = \frac{n_{i.}n_{.j}}{n^2} \quad (15.4.1)$$

for all i and j . In a multinomial distribution, the expected frequency in the (i, j) -th cell is np_{ij} where n is the total frequency. Then, the expected frequency, denoted by $E[\cdot]$, the maximum likelihood estimate (MLE) of the expected frequency, denoted by $\hat{E}[\cdot]$, and the MLE of the expected frequency under the hypothesis H_o of independence of events A_i and B_j , are the following:

$$E[n_{ij}] = np_{ij}, \hat{E}[n_{ij}] = n\hat{p}_{ij} = n\left(\frac{n_{ij}}{n}\right), n\hat{p}_{ij|H_o} = n\hat{p}_{i.}\hat{p}_{.j} = n\left(\frac{n_{i.}}{n}\right)\left(\frac{n_{.j}}{n}\right) = \frac{n_{i.}n_{.j}}{n}. \quad (15.4.2)$$

Now, referring to our numerical example and the first row of Table 15.4.1, the estimated expected frequencies, under H_o are: $E[n_{11}|H_o] = \frac{n_{1.}n_{.1}}{n} = \frac{40 \times 30}{100} = 12$, $E[n_{12}|H_o] = \frac{n_{1.}n_{.2}}{n} = \frac{40 \times 25}{100} = 10$, $E[n_{13}|H_o] = \frac{40 \times 30}{100} = 12$, $E[n_{14}|H_o] = \frac{40 \times 15}{100} = 6$. All the estimated expected frequencies are shown in parentheses next to the observed frequencies in Table 15.4.5:

Table 15.4.5: A two-way contingency table

$\downarrow \rightarrow$	B_1	B_2	B_3	B_4	Total
A_1	6(12)	14(10)	16(12)	4(6)	40(40)
A_2	17(12)	5(10)	8(12)	10(6)	40(40)
A_3	7(6)	6(5)	6(6)	1(3)	20(20)
Total	30(30)	25(25)	30(30)	15(15)	100(100)

15.4.2. Some general computations

Let J_r and J_s be respectively $r \times 1$ and $s \times 1$ vectors of unities and P be the true probability matrix, that is,

$$J_r = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, J_s = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}. \quad (15.4.3)$$

Letting the marginal totals be denoted by R and C' , we have

$$R = PJ_s = \begin{bmatrix} p_{1.} \\ p_{2.} \\ \vdots \\ p_{r.} \end{bmatrix}, \begin{matrix} J_r'P & = [p_{.1}, p_{.2}, \dots, p_{.s}] = C' \\ J_r'R & = 1 = C'J_s. \end{matrix} \quad (15.4.4)$$

Referring to the initial numerical example, we have the following:

$$\begin{aligned}\hat{R} &= \begin{bmatrix} \hat{p}_{1.} \\ \hat{p}_{2.} \\ \vdots \\ \hat{p}_{r.} \end{bmatrix} = \begin{bmatrix} n_{1.}/n \\ n_{2.}/n \\ \vdots \\ n_{r.}/n \end{bmatrix} = \begin{bmatrix} 40/100 \\ 40/100 \\ 20/100 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.2 \end{bmatrix} \\ \hat{C}' &= [\hat{p}_{.1}, \hat{p}_{.2}, \dots, \hat{p}_{.s}] = \left[\frac{n_{.1}}{n}, \dots, \frac{n_{.s}}{n} \right] \\ &= \left[\frac{30}{100}, \frac{25}{100}, \frac{30}{100}, \frac{15}{100} \right] = [0.30, 0.25, 0.30, 0.15].\end{aligned}$$

Writing the bordered matrix P as

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} & p_{1.} \\ p_{21} & p_{22} & \cdots & p_{2s} & p_{2.} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} & p_{r.} \\ p_{.1} & p_{.2} & \cdots & p_{.s} & 1 \end{bmatrix} = \begin{bmatrix} P & R \\ C' & 1 \end{bmatrix}, \quad (15.4.5)$$

in the numerical example, these quantities are

$$\begin{bmatrix} \hat{P} & \hat{R} \\ \hat{C}' & 1 \end{bmatrix} = \begin{bmatrix} 0.06 & 0.14 & 0.16 & 0.04 & 0.40 \\ 0.17 & 0.05 & 0.08 & 0.10 & 0.40 \\ 0.07 & 0.06 & 0.06 & 0.01 & 0.20 \\ 0.30 & 0.25 & 0.30 & 0.15 & 1.00 \end{bmatrix}.$$

Let D_r and D_c be the following diagonal matrices corresponding respectively to the row and column marginal probabilities:

$$\begin{aligned}D_r &= \begin{bmatrix} p_{1.} & 0 & \cdots & 0 \\ 0 & p_{2.} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{r.} \end{bmatrix}, \quad D_c = \begin{bmatrix} p_{.1} & 0 & \cdots & 0 \\ 0 & p_{.2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{.s} \end{bmatrix} \text{ or} \\ D_r &= \text{diag}(p_{1.}, p_{2.}, \dots, p_{r.}), \quad D_c = \text{diag}(p_{.1}, p_{.2}, \dots, p_{.s}).\end{aligned} \quad (15.4.6)$$

In the numerical example, these quantities are

$$\hat{D}_r = \text{diag}(0.4, 0.4, 0.2) \quad \text{and} \quad \hat{D}_c = \text{diag}(0.30, 0.25, 0.30, 0.15).$$

Now, consider $D_r^{-1}P$ and PD_c^{-1} :

$$D_r^{-1}P = \begin{bmatrix} \frac{p_{11}}{p_1} & \frac{p_{12}}{p_1} & \cdots & \frac{p_{1s}}{p_1} \\ \frac{p_{21}}{p_2} & \frac{p_{22}}{p_2} & \cdots & \frac{p_{2s}}{p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{r1}}{p_r} & \frac{p_{r2}}{p_r} & \cdots & \frac{p_{rs}}{p_r} \end{bmatrix} \equiv \begin{bmatrix} R'_1 \\ R'_2 \\ \vdots \\ R'_r \end{bmatrix}, \quad R_j = \begin{bmatrix} \frac{p_{j1}}{p_j} \\ \frac{p_{j2}}{p_j} \\ \vdots \\ \frac{p_{js}}{p_j} \end{bmatrix}, \quad (15.4.7)$$

$$PD_c^{-1} = \begin{bmatrix} \frac{p_{11}}{p_{.1}} & \frac{p_{12}}{p_{.2}} & \cdots & \frac{p_{1s}}{p_{.s}} \\ \frac{p_{21}}{p_{.1}} & \frac{p_{22}}{p_{.2}} & \cdots & \frac{p_{2s}}{p_{.s}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{r1}}{p_{.1}} & \frac{p_{r2}}{p_{.2}} & \cdots & \frac{p_{rs}}{p_{.s}} \end{bmatrix} \equiv [C_1, \dots, C_s], \quad C_j = \begin{bmatrix} \frac{p_{1j}}{p_{.j}} \\ \frac{p_{2j}}{p_{.j}} \\ \vdots \\ \frac{p_{rj}}{p_{.j}} \end{bmatrix}. \quad (15.4.8)$$

Referring to the numerical example, we have

$$\hat{D}_r^{-1}\hat{P} = \begin{bmatrix} n_{11}/n_{.1} & n_{12}/n_{.1} & \cdots & n_{1s}/n_{.1} \\ n_{21}/n_{.1} & n_{22}/n_{.1} & \cdots & n_{2s}/n_{.1} \\ \vdots & \vdots & \ddots & \vdots \\ n_{r1}/n_{.1} & n_{r2}/n_{.1} & \cdots & n_{rs}/n_{.1} \end{bmatrix} = \begin{bmatrix} 6/40 & 14/40 & 16/40 & 4/40 \\ 17/40 & 5/40 & 8/40 & 10/40 \\ 7/20 & 6/20 & 6/20 & 1/20 \end{bmatrix}$$

$$\hat{D}_r^{-1}\hat{P}J_s = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\hat{P}\hat{D}_c^{-1} = \begin{bmatrix} n_{11}/n_{1.} & n_{12}/n_{1.} & \cdots & n_{1s}/n_{1.} \\ n_{21}/n_{1.} & n_{22}/n_{1.} & \cdots & n_{2s}/n_{1.} \\ \vdots & \vdots & \ddots & \vdots \\ n_{r1}/n_{1.} & n_{r2}/n_{1.} & \cdots & n_{rs}/n_{1.} \end{bmatrix} = \begin{bmatrix} 6/30 & 14/25 & 16/30 & 4/15 \\ 17/30 & 5/25 & 8/30 & 10/15 \\ 7/30 & 6/25 & 6/30 & 1/15 \end{bmatrix}$$

$$J'_r\hat{P}\hat{D}_c^{-1} = [1, 1, 1, 1].$$

For computing the test statistics in vector/matrix notation, we need (15.4.7) and (15.4.8).

15.5. Various Representations of Pearson's χ^2 Statistic

Now, let us consider Pearson's χ^2 statistic for testing the hypothesis that there is no association between the two characteristics of classification or the hypothesis $H_o : p_{ij} = p_{i.}p_{.j}$. The χ^2 statistic is the following:

$$\chi^2 = \sum_{ij} \frac{(\text{observed frequency} - \text{expected frequency})^2}{(\text{expected frequency})} = \sum_{ij} \frac{(n_{ij} - \frac{n_i n_j}{n})^2}{\frac{n_i n_j}{n}} \quad (15.5.1)$$

$$= \sum_{ij} n \frac{(\frac{n_{ij}}{n} - \frac{n_i}{n} \frac{n_j}{n})^2}{\frac{n_i}{n} \frac{n_j}{n}} = n \sum_{ij} \frac{(\hat{p}_{ij} - \hat{p}_i \hat{p}_j)^2}{\hat{p}_i \hat{p}_j} \quad (15.5.2)$$

$$= \sum_{i=1}^r n \hat{p}_i \sum_{j=1}^s \left[\left(\frac{\hat{p}_{ij}}{\hat{p}_i} - \hat{p}_j \right)^2 / \hat{p}_j \right] \quad (15.5.3)$$

$$= \sum_{j=1}^s n \hat{p}_j \sum_{i=1}^r \left[\left(\frac{\hat{p}_{ij}}{\hat{p}_j} - \hat{p}_i \right)^2 / \hat{p}_i \right]. \quad (15.5.4)$$

In order to simplify the notation, we shall omit placing a hat on top of the estimates of R_i , C_j , R , C , D_c and D_r . We may then express the χ^2 statistic as the following quadratic forms:

$$\chi^2 = \sum_{i=1}^r n p_i (R_i - C)' D_c^{-1} (R_i - C) \quad (15.5.5)$$

$$= \sum_{j=1}^s n p_j (C_j - R)' D_r^{-1} (C_j - R). \quad (15.5.6)$$

The forms given in (15.5.5) and (15.5.6) are very convenient for extending the theory to multi-way classifications.

It is well known that, under H_o , Pearson's χ^2 statistic is asymptotically distributed as a chisquare random variable having $(r - 1)(s - 1)$ degrees of freedom as $n \rightarrow \infty$. One can also express (15.4.8) as a generalized distance between the observed frequencies and the expected frequencies, which is a quadratic form involving the inverse of the true covariance matrix of the multinomial distribution of the n_{ij} 's. Then, on applying the multivariate version of the central limit theorem, it can be established that, as $n \rightarrow \infty$, Pearson's χ^2 statistic has a χ^2 distribution with $(r - 1)(s - 1)$ degrees of freedom. For the representation of Pearson's χ^2 goodness-of-fit statistic as a generalized distance and as a quadratic form, and for the proof of its asymptotic distribution, the reader may refer to Mathai and Haubold (2017). There exist other derivations of this result in the literature.

The quadratic forms specified in (15.5.5) and (15.5.6) can also be interpreted as comparing the generalized distance between the vectors R_i and C in (15.5.5) and between the vectors C_j and R in (15.5.6), respectively. These will also be equivalent to testing the hypothesis $H_o : p_{ij} = p_i p_j$. As well, an interpretation can be provided in terms of profile analysis: then, the test will correspond to testing the hypothesis that the weighted row profiles are similar; analogously, using (15.5.6) corresponds to testing the hypothesis that the

column profiles in a two-way contingency table are similar. Now, examine the following item:

$$\begin{aligned}
 P - RC' &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix} - \begin{bmatrix} p_{1.} \\ p_{2.} \\ \vdots \\ p_{r.} \end{bmatrix} [p_{.1}, p_{.2}, \dots, p_{.s}] \\
 &= \begin{bmatrix} p_{11} - p_{1.}p_{.1} & p_{12} - p_{1.}p_{.2} & \cdots & p_{1s} - p_{1.}p_{.s} \\ p_{21} - p_{2.}p_{.1} & p_{22} - p_{2.}p_{.2} & \cdots & p_{2s} - p_{2.}p_{.s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{rs} - p_{r.}p_{.1} & p_{r2} - p_{r.}p_{.2} & \cdots & p_{rs} - p_{r.}p_{.s} \end{bmatrix}.
 \end{aligned}$$

Referring to our numerical example, these quantities are the following:

$$\begin{aligned}
 \hat{P} - \hat{R}\hat{C}' &= \begin{bmatrix} \frac{n_{11}}{n} - \frac{n_{1.}n_{.1}}{n^2} & \cdots & \frac{n_{1s}}{n} - \frac{n_{1.}n_{.s}}{n^2} \\ \vdots & \ddots & \vdots \\ \frac{n_{r1}}{n} - \frac{n_{r.}n_{.1}}{n^2} & \cdots & \frac{n_{rs}}{n} - \frac{n_{r.}n_{.s}}{n^2} \end{bmatrix} = \frac{1}{100} \times \\
 &\begin{bmatrix} 6 - (40)(30)/100 & 14 - (40)(25)/100 & 16 - (40)(30)/100 & 4 - (40)(15)/100 \\ 17 - (40)(30)/100 & 5 - (40)(25)/100 & 8 - (40)(30)/100 & 10 - (40)(15)/100 \\ 7 - (20)(30)/100 & 6 - (20)(25)/100 & 6 - (20)(30)/100 & 1 - (20)(15)/100 \end{bmatrix} \\
 &= \begin{bmatrix} -6 & 4 & 4 & -2 \\ 5 & -5 & -4 & 4 \\ 1 & 1 & 0 & -2 \end{bmatrix}.
 \end{aligned}$$

15.5.1. Testing the hypothesis of no association in a two-way contingency table

The observed value of Pearson's χ^2 statistic is

$$\begin{aligned}
 \chi^2 &= \left[\frac{(-6)^2}{12} + \frac{(5)^2}{12} + \frac{(1)^2}{6} \right] + \left[\frac{(4)^2}{10} + \frac{(-5)^2}{10} + \frac{(1)^2}{5} \right] \\
 &\quad + \left[\frac{(4)^2}{12} + \frac{(-4)^2}{12} + \frac{(0)^2}{6} \right] + \left[\frac{(-2)^2}{6} + \frac{(4)^2}{6} + \frac{(-2)^2}{3} \right] \\
 &= 16.88.
 \end{aligned}$$

Given our data, $(r - 1)(s - 1) = (2)(3) = 6$, and the tabulated critical value is $\chi_{6,0.05}^2 = 12.59$ at the 5% significance level. Since $12.59 < 16.88$, the hypothesis of no association between the classification attributes is rejected as per the evidence provided by the data. This χ^2 approximation may be questionable since one of the expected cell frequencies is less than 5. For a proper application of this approximation, the cell frequencies ought to be at least 5.

15.6. Plot of Row and Column Profiles

Now, $(P - RC')D_c^{-1}$ means that the columns of $(P - RC')$ are multiplied by $\frac{1}{p_{.1}}, \dots, \frac{1}{p_{.s}}$, respectively. Then, $(P - RC')D_c^{-1}(P - RC)'$ is a matrix of all square and cross product terms involving $p_{ij} - p_{i \cdot} p_{\cdot j}$ for all i and j , where the s columns are weighted by $\frac{1}{p_{.j}}$, and if pre-multiplied by D_r^{-1} , the rows are weighted by $\frac{1}{p_{i \cdot}}, \dots, \frac{1}{p_{r \cdot}}$, respectively. Looking at the diagonal elements, we note that Pearson's χ^2 statistic is nothing but

$$\chi^2 = n \operatorname{tr}[D_r^{-1}(P - RC')D_c^{-1}(P - RC)'] \quad (15.6.1)$$

$$= n \sum_{ij} \frac{(p_{ij} - p_{i \cdot} p_{\cdot j})^2}{p_{i \cdot} p_{\cdot j}} \quad (15.6.2)$$

$$= n(\lambda_1^2 + \dots + \lambda_k^2) \quad (15.6.3)$$

where $\lambda_1^2, \dots, \lambda_k^2$ are the nonzero eigenvalues of the matrix $D_r^{-1}(P - RC')D_c^{-1}(P - RC)'$ or of the matrix $D_r^{-\frac{1}{2}}(P - RC')D_c^{-1}(P - RC)D_r^{-\frac{1}{2}}$ with k being the rank of $P - RC'$. For the numerical example, the observed value of the matrix $Y = (y_{ij})$ with $y_{ij} = \frac{p_{ij} - p_{i \cdot} p_{\cdot j}}{\sqrt{p_{i \cdot} p_{\cdot j}}}$, is obtained as follows, observing that

$$\sqrt{n} \frac{\hat{p}_{ij} - \hat{p}_{i \cdot} \hat{p}_{\cdot j}}{\sqrt{\hat{p}_{i \cdot} \hat{p}_{\cdot j}}} = \left[n_{ij} - \frac{n_{i \cdot} n_{\cdot j}}{n} \right] / \sqrt{n_{i \cdot} n_{\cdot j} / n}.$$

From the representation of $\hat{P} - \hat{R}\hat{C}'$, we already have the matrix $n_{ij} - n_{i \cdot} n_{\cdot j} / n$, that is,

$$\begin{aligned} & \left(\frac{(n_{ij} - n_{i \cdot} n_{\cdot j} / n)}{\sqrt{n_{i \cdot} n_{\cdot j} / n}} \right) \\ &= \begin{bmatrix} (6 - 12) / \sqrt{12} & (14 - 10) / \sqrt{10} & (16 - 12) / \sqrt{12} & (4 - 6) / \sqrt{6} \\ (17 - 12) / \sqrt{12} & (5 - 10) / \sqrt{10} & (8 - 12) / \sqrt{12} & (10 - 6) / \sqrt{6} \\ (7 - 6) / \sqrt{6} & (6 - 5) / \sqrt{5} & (6 - 6) / \sqrt{6} & (1 - 3) / \sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} -6 / \sqrt{12} & 4 / \sqrt{10} & 4 / \sqrt{12} & -2 / \sqrt{6} \\ 5 / \sqrt{12} & -5 / \sqrt{10} & -4 / \sqrt{12} & 4 / \sqrt{6} \\ 1 / \sqrt{6} & 1 / \sqrt{5} & 0 / \sqrt{6} & -2 / \sqrt{3} \end{bmatrix}. \end{aligned}$$

Then,

$$nYY' = \begin{bmatrix} \frac{33}{5} & -\frac{43}{6} & \frac{17\sqrt{2}}{30} \\ -\frac{43}{6} & \frac{103}{12} & -\frac{17\sqrt{2}}{12} \\ \frac{17\sqrt{2}}{30} & -\frac{17\sqrt{2}}{12} & \frac{17}{10} \end{bmatrix}. \quad (15.6.4)$$

The representation in (15.6.1) has the advantage that

$$\begin{aligned} & \text{tr}[D_r^{-\frac{1}{2}}(P - RC')D_c^{-1}(P - RC')'D_r^{-\frac{1}{2}}] \\ &= \text{tr}[YY'], \quad Y = D_r^{-\frac{1}{2}}(P - RC')D_c^{-\frac{1}{2}} = (y_{ij}), \\ y_{ij} &= \frac{p_{ij} - p_{i.}p_{.j}}{\sqrt{p_{i.}p_{.j}}}, \quad \sum_j n\hat{y}_{ij}^2 = \chi^2 = n\text{tr}(\hat{Y}\hat{Y}'). \end{aligned} \quad (15.6.5)$$

Note that Y is $r \times s$ and the rank of Y is equal to the rank of $P - RC'$, which is k , referring to (15.6.3). Thus, there are k nonzero eigenvalues associated with the $r \times r$ matrix YY' as well as with the $s \times s$ matrix $Y'Y$, which are $\lambda_1^2, \dots, \lambda_k^2$. Since $\text{tr}(YY') = \lambda_1^2 + \dots + \lambda_k^2$, we can represent Pearson's χ^2 statistic as follows, substituting the estimates of p_{ij} , $p_{i.}$ and $p_{.j}$, etc:

$$\begin{aligned} \frac{\chi^2}{n} &= \text{tr}(YY') = \lambda_1^2 + \dots + \lambda_k^2 \\ &= \sum_{i=1}^k \hat{p}_{i.}(\hat{R}_i - \hat{C})'\hat{D}_c^{-1}(\hat{R}_i - \hat{C}) \end{aligned} \quad (15.6.6)$$

$$= \sum_{j=1}^s \hat{p}_{.j}(\hat{C}_j - \hat{R})'\hat{D}_r^{-1}(\hat{C}_j - \hat{R}). \quad (15.6.7)$$

The expressions given in (15.6.6) and (15.6.7) and the sum of the λ_j^2 's are called the *total inertia* in a two-way contingency table. We can also define the squared distance between two rows as

$$d_{ij(r)}^2 = (R_i - R_j)'D_c^{-1}(R_i - R_j) \quad (15.6.8)$$

and the squared distance between two columns as

$$d_{ij(c)}^2 = (C_i - C_j)'D_r^{-1}(C_i - C_j). \quad (15.6.9)$$

When the distance as specified in (15.6.8) is very small, we may combine the i -th and j -th rows, if necessary. Sometimes, the cell frequencies are small and we may wish to combine the small frequencies with other cell frequencies so that the χ^2 approximation of Pearson's χ^2 statistic be more accurate. Then, one can rely on (15.6.8) and (15.6.9) to determine whether it is indicated to combine rows and columns.

For convenience, let $r \leq s$. Let U_1, \dots, U_r be the $r \times 1$ normalized eigenvectors of YY' and let the $r \times k$ matrix $U = [U_1, U_2, \dots, U_k]$, $k \leq r$. Let V_1, \dots, V_s be the normalized

eigenvectors of $Y'Y$ and let the $r \times k$ matrix $V = [V_1, \dots, V_k]$, $k \leq s$. Now, consider the singular value decomposition

$$Y = D_r^{-\frac{1}{2}}(P - RC')D_c^{-\frac{1}{2}} = U\Lambda V' \quad (15.6.10)$$

where $UU' = I_k = V'V$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$. Then, we can write

$$P - RC' = D_r^{\frac{1}{2}}U\Lambda V'D_c^{\frac{1}{2}} = W\Lambda Z' \quad (15.6.11)$$

where $W = D_r^{\frac{1}{2}}U$ and $Z = D_c^{\frac{1}{2}}V$. Let W_j , $j = 1, \dots, k$, denote the columns of $W = [W_1, W_2, \dots, W_k]$ and let Z_j , $j = 1, \dots, k$, denote the columns of $Z = [Z_1, Z_2, \dots, Z_k]$. Then, we can write

$$P - RC' = \sum_{j=1}^k \lambda_j W_j Z_j' \quad (15.6.12)$$

where $W'D_r^{-1}W = U'U = I_k = V'V = Z'D_c^{-1}Z$. Note that $P - RC'$ is the deviation matrix under the hypothesis $H_0 : p_{ij} = p_{i \cdot} p_{\cdot j}$ or

$$P - RC' = (p_{ij} - p_{i \cdot} p_{\cdot j}) \quad \text{and} \quad Y = (y_{ij}) = D_r^{-\frac{1}{2}}(P - RC')D_c^{-\frac{1}{2}} = \left(\frac{p_{ij} - p_{i \cdot} p_{\cdot j}}{\sqrt{p_{i \cdot} p_{\cdot j}}} \right).$$

Thus, the procedure is as follows: If $r \leq s$, then compute the r eigenvalues of the $r \times r$ matrix YY' . If Y is of rank r , $YY' > O$ (positive definite), otherwise YY' is positive semi-definite. Let the nonzero eigenvalues of YY' be $\lambda_1^2, \dots, \lambda_k^2$, assuming that k is the number of nonzero eigenvalues of YY' . These will also be the nonzero eigenvalues of $Y'Y$. Compute the normalized eigenvectors from YY' and denote those corresponding to the nonzero eigenvalues by $U = [U_1, \dots, U_k]$ where U_j is the j -th column of U . Letting the normalized eigenvectors obtained from $Y'Y$, which correspond to the same nonzero eigenvalues, be denoted by $V = [V_1, \dots, V_k]$, we have

$$Y = U\Lambda V', \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_k), \quad YY' = U\Lambda^2 U' \quad \text{and} \quad Y'Y = V\Lambda^2 V'. \quad (15.6.13)$$

Example 15.6.1. Construct a singular value decomposition of the following matrix Q :

$$Q = \begin{bmatrix} -1 & 1 & -1 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}.$$

Solution 15.6.1. Let us compute QQ' as well as $Q'Q$ and the eigenvalues of QQ' . Since

$$QQ' = \begin{bmatrix} -1 & 1 & -1 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ -1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix},$$

the eigenvalues of QQ' are $\lambda_1 = 3$ and $\lambda_2 = 6$. Let us determine the normalized eigenvectors of QQ' . Consider the equation $[QQ' - \lambda I]X = O$ for $\lambda = 3$ and 6 , and let $X' = [x_1, x_2]$ and $O' = [0, 0]$. Then, for $\lambda = 3$, we see that $x_2 = 0$ and for $\lambda = 6$, we note that $x_1 = 0$. Thus, the normalized solutions are

$$U_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } U_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow U = [U_1, U_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that $-U_1$ or $-U_2$ or $-U_1, -U_2$ will also satisfy all the conditions, and we could take any of these forms for convenience. Now, consider the equation $(Q'Q - \lambda I)X = O$, where $X' = [x_1, x_2, x_3, x_4]$ and $O' = [0, 0, 0, 0]$ for $\lambda = 3, 6$. For $\lambda = 3$, the coefficient matrix is

$$Q'Q - 3I = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & -1 & -1 & 2 \\ 1 & -1 & -2 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 0 & 1 & 2 \\ 0 & -1 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

by elementary transformations. Observe that $x_4 = 0$ so that $-x_1 + x_3 = 0$ and $-x_2 - x_3 = 0$. Thus, one solution or an eigenvector corresponding to $\lambda = 3$ and their normalized form are

$$\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow V_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}.$$

Now, take $\lambda = 6$ and consider the equation $(Q'Q - 6I)X = O$; the coefficient matrix and its reduced form obtained through elementary transformations are the following:

$$\begin{bmatrix} -4 & 0 & 1 & 2 \\ 0 & -4 & -1 & 2 \\ 1 & -1 & -5 & 0 \\ 2 & 2 & 0 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & -5 & 0 \\ 0 & -4 & -1 & 2 \\ 0 & 0 & -21 & 0 \\ 0 & 0 & 9 & 0 \end{bmatrix},$$

which shows that $x_3 = 0$, so that $x_1 - x_2 = 0$ and $-4x_2 + 2x_4 = 0$. Hence, an eigenvector and its normalized form are

$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix} \Rightarrow V_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}.$$

Thus, $V = [V_1, V_2]$. As mentioned earlier, we could have $-V_1$ or $-V_2$ or $-V_1, -V_2$ as the normalized eigenvectors. As per our notation,

$$\Lambda = \text{diag}(\sqrt{3}, \sqrt{6}) \text{ and } Q = U \Lambda V'.$$

Let us verify this last equality. Since

$$\begin{aligned} U \Lambda V' &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{6} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix}, \end{aligned}$$

we should take $-V_1$ to obtain Q . Then,

$$[U_1, U_2] \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{6} \end{bmatrix} \begin{bmatrix} -V_1' \\ V_2' \end{bmatrix} = Q,$$

which verifies the result and completes the computations.

Now, we shall continue with our row and column profile plots. From (15.6.4), we have

$$nYY' = \begin{bmatrix} \frac{33}{5} & -\frac{43}{6} & \frac{17\sqrt{2}}{30} \\ -\frac{43}{6} & \frac{103}{12} & -\frac{17\sqrt{2}}{12} \\ \frac{17\sqrt{2}}{30} & -\frac{17\sqrt{2}}{12} & \frac{17}{10} \end{bmatrix} \text{ and } nY'Y = \begin{bmatrix} \frac{63}{12} & -\frac{47\sqrt{30}}{60} & -\frac{11}{3} & \frac{7\sqrt{2}}{3} \\ -\frac{47\sqrt{30}}{60} & \frac{43}{10} & \frac{3\sqrt{30}}{5} & -\frac{16\sqrt{15}}{15} \\ -\frac{11}{3} & \frac{3\sqrt{30}}{5} & \frac{8}{3} & -2\sqrt{2} \\ \frac{7\sqrt{2}}{3} & -\frac{16\sqrt{15}}{15} & -2\sqrt{2} & \frac{14}{3} \end{bmatrix}.$$

The eigenvalues of nYY' are $\lambda_1 = 15.1369$, $\lambda_2 = 1.7471$ and $\lambda_3 = 0$ and the normalized eigenvectors from nYY' , corresponding to λ_1 , λ_2 and λ_3 are U_1, U_2, U_3 , so that $U = [U_1, U_2, U_3]$ where

$$U = \begin{bmatrix} 4.28 & -0.49 & 1.41 \\ -4.99 & -0.22 & 1.41 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \Lambda = \text{diag}(\sqrt{15.1369}, \sqrt{1.7471}, 0).$$

For the same eigenvalues λ_1, λ_2 and λ_3 , the normalized eigenvectors determined from $nY'Y$, which correspond to the nonzero eigenvalues, are V_1 and V_2 , with

$$V = [V_1, V_2] = \begin{bmatrix} 1.10 & -0.84 \\ -1.06 & -0.16 \\ -0.83 & 0.28 \\ 1 & 1 \end{bmatrix}.$$

Since $\lambda_3 = 0$, $k = 2$, and we can take the $r \times k$, that is, 3×2 matrix $G = (g_{ij}) = D_r^{-\frac{1}{2}} U \Lambda$ to represent the row deviation profiles and the $s \times k = 4 \times 2$ matrix $H = (h_{ij}) = D_c^{-\frac{1}{2}} V \Lambda$ to represent the column deviation profiles. For our numerical example, it follows from (15.4.6) that

$$\begin{aligned} D_r &= \text{diag}(0.4, 0.4, 0.2) \Rightarrow D_r^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{0.63}, \frac{1}{0.63}, \frac{1}{0.45}\right) \\ D_c &= \text{diag}(0.30, 0.25, 0.30, 0.15) \Rightarrow D_c^{-\frac{1}{2}} = \text{diag}\left(\frac{1}{0.55}, \frac{1}{0.50}, \frac{1}{0.55}, \frac{1}{0.39}\right) \\ \Lambda &= \text{diag}(\sqrt{15.1369}, \sqrt{1.7471}, 0) = \text{diag}(3.89, 1.32, 0). \end{aligned}$$

We only take the first two columns of U and V since $\lambda_3 = 0$; besides, only the first two vectors are required for plotting. Let $U_{(1)}$ and $V_{(1)}$ represent the first two columns of U and V , respectively. Then, $D_r^{-\frac{1}{2}} U_{(1)} \Lambda$ will be equivalent to multiplying the first and second columns by 3.89 and 1.32, respectively, and multiplying the first and second rows by $\frac{1}{0.63}$ and the third row by $\frac{1}{0.45}$. Then, we have

$$U_{(1)} = \begin{bmatrix} 4.28 & -0.49 \\ -4.99 & -0.22 \\ 1 & 1 \end{bmatrix}, \quad D_r^{-\frac{1}{2}} U_{(1)} \Lambda = \begin{bmatrix} 26.42 & -1.03 \\ -30.81 & -0.46 \\ 6.17 & 2.09 \end{bmatrix} \equiv G_2$$

where G_2 is the matrix consisting of the first two columns of G . Hence, the points required for plotting the row profile are: (26.42, -1.03), (-30.81, -0.46), (6.17, 2.09). These points being far apart, no two rows should be combined. Now, consider the column profiles: the effect of $D_c^{-\frac{1}{2}} V_{(1)} \Lambda$ is to multiply the columns of $V_{(1)}$ by 3.89 and 1.32, respectively, and to multiply the rows by $\frac{1}{0.55}, \frac{1}{0.50}, \frac{1}{0.55}, \frac{1}{0.39}$, respectively. Thus,

$$V_{(1)} = \begin{bmatrix} 1.10 & -0.84 \\ -1.06 & -0.16 \\ -0.83 & 0.28 \\ 1 & 1 \end{bmatrix}, \quad D_c^{-\frac{1}{2}} V_{(1)} \Lambda = \begin{bmatrix} 7.78 & -2.02 \\ -8.25 & -0.42 \\ -5.87 & -0.67 \\ 9.97 & 3.38 \end{bmatrix} \equiv H_2$$

where H_2 is the matrix consisting of the first two columns of H . The row profile and the column profile points are plotted in Fig. 15.6.1 where r next to a point indicates a row point and c designates a column point. That is, $i r$ indicates the i -th row point and $j c$, the j -th column point. It can be seen from this plot that the row points are far apart while the second and third column points are somewhat close; accordingly, if necessary, the second and third columns could be combined.

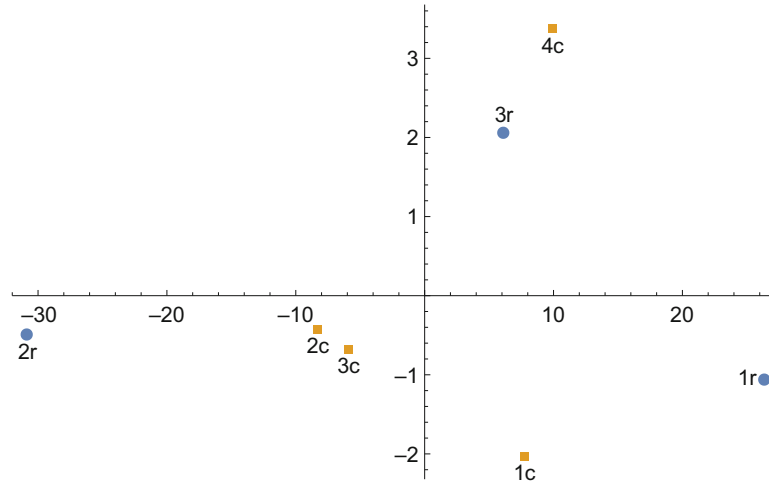


Figure 15.6.1 Row profile and column profile points

15.7. Correspondence Analysis in a Multi-way Contingency Table

When the data is classified under a number of variables, each variable having a number of categories, the resulting frequency table is referred to as a multi-way classification. Correspondence analysis for a multi-way classification involves converting data in a multi-way classification setting into a two-way classification framework and then, employing the techniques developed in Sects. 15.5 and 15.6. The first step in this regard consists of creating an indicator matrix C . In order to illustrate the steps, we will first present an example. Suppose that 10 persons selected at random from a community, are classified according to three variables. Variable 1 is gender. Under this variable, we shall consider the categories male and female. Variable 2 is weight. Under this variable, we are considering three categories: underweight, normal and overweight. The third variable is education which

is assumed to have four levels: level 1, level 2, level 3 and level 4. Thus, there are three variables and 9 categories. The actual data are provided in Table 15.7.1.

Table 15.7.1: Ten persons classified under three variables

Person #	Variables		
	Gender	Weight	Educational level
1	Female	Overweight	Level 2
2	Female	Normal	Level 4
3	Male	Underweight	Level 1
4	Female	Normal	Level 3
5	Male	Overweight	Level 1
6	Male	Normal	Level 2
7	Female	Overweight	Level 3
8	Female	Underweight	Level 4
9	Male	Normal	Level 3
10	Female	Overweight	Level 1

Table 15.7.2: Entries of the indicator matrix of the data included in Table 15.7.1

Person #	Variables								
	Gender		Weight			Educational level			
	M	F	U	N	O	L1	L2	L3	L4
1	0	1	0	0	1	0	1	0	0
2	0	1	0	1	0	0	0	0	1
3	1	0	1	0	0	1	0	0	0
4	0	1	0	1	0	0	0	1	0
5	1	0	0	0	1	1	0	0	0
6	1	0	0	1	0	0	1	0	0
7	0	1	1	0	0	0	0	1	0
8	0	1	1	0	0	0	0	0	1
9	1	0	0	1	0	0	0	1	0
10	0	1	0	0	1	1	0	0	0

Next, we construct the indicator matrix C —distinct from C as defined in (15.4.4)—of the data displayed in Table 15.7.1. If an item is present, we write 1 in the corresponding location in Table 15.7.2, and if it is absent, we write 0, thus populating this table where M \equiv Male, F \equiv Female, U \equiv underweight, N \equiv Normal, O \equiv overweight, L1 \equiv Level 1, L2 \equiv Level 2, L3 \equiv Level 3 and L4 \equiv Level 4. The resulting indicator matrix C is

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Note that since a person will belong to a single category of every variable, the row sum of every row will always be equal to the number of variables, which is 3 in the example. The sum of *all* the column entries under *each* variable is the number of items classified (10 in the example). We now convert the data into a two-way classification, which is achieved by converting C into a *Burt matrix* B , where $B = C'C$. In our example,

$$B = C'C = \begin{bmatrix} 4 & 0 & 1 & 2 & 1 & 2 & 1 & 1 & 0 \\ 0 & 6 & 1 & 2 & 3 & 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 2 & 2 & 0 & 4 & 0 & 0 & 1 & 2 & 1 \\ 1 & 3 & 0 & 0 & 4 & 2 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 & 2 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 1 & 0 & 0 & 3 & 0 \\ 0 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Observe that the diagonal blocks in $C'C$ correspond to the variables, gender, weight and educational level or gender versus gender, weight versus weight, educational level versus educational level. These blocks are the following:

$$\begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Various two-way contingency tables, namely gender versus weight, gender versus educational level, weight versus educational level, are combined into one two-way table displaying category versus category. The observed Pearson's χ^2 statistic from $C'C$ is seen to be 79.85. In this case, the number of degrees of freedom is $8 \times 8 = 64$ and at 5% level, the tabulated $\chi_{64,0.05}^2 \approx 84 > 79.85$; hence, the hypothesis of no association in $C'C$ is not rejected. Note that this χ^2 approximation is unreliable since the expected frequencies are small. The most relevant parts in the Burt matrix $C'C$ are the non-diagonal blocks of frequencies. The two non-diagonal blocks of the first two rows represent the two-way contingency tables for gender versus weight and gender versus educational level. Similarly, the non-diagonal block in the third to fifth rows represent the two-way contingency table for weight versus educational level. These are the following, denoted by A_1, A_2, A_3 respectively, where A_1 is the two-way contingency table of gender versus weight, A_2 is the contingency table of gender versus educational level and A_3 is the table of weight versus educational level:

$$A_1 = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 1 & 2 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix}.$$

The corresponding matrices of expected frequencies, under the hypothesis of no association between the characteristics of classification, denoted by $E(A_i)$, $i = 1, 2, 3$ are

$$E(A_1) = \begin{bmatrix} 0.8 & 1.6 & 1.6 \\ 1.2 & 2.4 & 2.4 \end{bmatrix}, \quad E(A_2) = \begin{bmatrix} 1.2 & 0.8 & 1.2 & 0.8 \\ 1.8 & 1.2 & 1.8 & 1.2 \end{bmatrix},$$

$$E(A_3) = \begin{bmatrix} 0.6 & 0.4 & 0.6 & 0.4 \\ 1.2 & 0.8 & 1.2 & 0.8 \\ 1.2 & 0.8 & 1.2 & 0.8 \end{bmatrix}.$$

The observed values of Pearson's χ^2 statistic under the hypothesis of no association in the contingency table, and the corresponding tabulated χ^2 critical values at the 5% significance level, are the following: $A_1 : \chi^2 = 0.63$, $\chi_{2,0.05}^2 = 5.99 > 0.63$; $A_2 : \chi^2 = 2.36$, $\chi_{3,0.05}^2 = 7.81 > 2.36$; $A_3 : \chi^2 = 5.42$, $\chi_{6,0.05}^2 = 12.59 > 5.42$; hence the hypothesis would not be rejected in any of the contingency table if Pearson's statistic were applicable. Actually, the χ^2 approximation is not appropriate in any of these cases since the expected frequencies are quite small. Hence, decisions cannot be made on the basis of Pearson's statistic in these instances.

Observe that the first column of the matrix C corresponds to the count on "Male", the second to the count on "Female", the third to "Underweight", the fourth to "Normal", the fifth to "Overweight", the sixth to "Level 1", the seventh to "Level 2", the eighth to "Level

3” and the ninth to “Level 4”. Thus, the columns represent the various characteristics or the various variables and their categories. So, if we were to plot one column as one point in the two-dimensional space, then by looking at the points we could determine which points are close to each other. For example, if the “Overweight” column point is close to the “Male” column point, then there is possibility of association between “Overweight” and “Male”. Thus, our aim will be to plot each column of C or each column of $C'C$ as a point in two dimensions. For this purpose, we may make use of the plotting technique described in Sects. 15.5 and 15.6. Consider a singular value decomposition of $C = U\Lambda V'$, $U'U = I_k$, $V'V = I_k$. If C is $r \times s$, $s < r$, then U is $r \times k$ and V is $s \times k$ where k is the number of nonzero eigenvalues of CC' as well as those of $C'C$, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$ where λ_j^2 , $j = 1, \dots, k$, are the nonzero eigenvalues of CC' and $C'C$. In the numerical example, $r = 10$ and $s = 9$. Consider the eigenvalues of $C'C$ since in this case, the order is smaller than the order of CC' . Let the nonzero eigenvalues of $C'C$ be $\lambda_1^2 \geq \dots \geq \lambda_k^2$. From $C'C$, compute the normalized eigenvectors corresponding to these nonzero eigenvalues. This $s \times k$ matrix of normalized eigenvectors is V in the singular value decomposition. By using the same nonzero eigenvalues, compute the normalized eigenvectors from CC' . This $r \times k$ matrix is U in the singular value decomposition. Since the columns of C and $C'C$ represent the various variables and their subdivisions, only the columns are useful for our geometrical representation, that is, only V will be relevant for plotting the points. Consider $H = V\Lambda$ and let $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_k^2$. Observe that $C = U\Lambda V' \Rightarrow C' = V\Lambda U' = HU'$. The rows of C' represent the various variables and their categories. Let h_1, \dots, h_s be the rows of H . Then, we have

$$\begin{aligned} h_1 U' &= \text{Men-row} \\ h_2 U' &= \text{Women-row} \\ &\vdots \\ h_s U' &= \text{Level 4-row.} \end{aligned}$$

This shows that the rows h_1, \dots, h_s represent the various variables and their categories. Since the first two eigenvalues are the largest ones and V_1, V_2 are the corresponding eigenvectors, we can take it for granted that most of the information about the various variables and their categories is contained in the first two elements in h_1, \dots, h_s or in the first two columns weighted by λ_1 and λ_2 . Accordingly, take the first two columns from H and denote this submatrix by $H_{(2)}$ where

$$H_{(2)} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ \vdots & \vdots \\ h_{s1} & h_{s2} \end{bmatrix}.$$

Plot the points $(h_{11}, h_{12}), (h_{21}, h_{22}), \dots, (h_{s1}, h_{s2})$. These s points correspond to the s columns in the $r \times s$ matrix C or the s rows in C' .

Referring to our numerical example, the eigenvalues are

$$\lambda_1^2 = 11.66, \lambda_2^2 = 5.57, \lambda_3^2 = 5.28, \lambda_4^2 = 3.47, \lambda_5^2 = 2.34, \\ \lambda_6^2 = 1.14, \lambda_7^2 = 0.54, \lambda_8 = \lambda_9 = 0,$$

so that $k = 7$ and the nonzero eigenvalues, $\sqrt{\lambda_j^2}$, $j = 1, \dots, 7$, are

$$\lambda_1 = 3.41, \lambda_2 = 2.36, \lambda_3 = 2.30, \lambda_4 = 1.86, \lambda_5 = 1.53, \lambda_6 = 1.07, \lambda_7 = 0.73.$$

Thus, the matrix Λ is

$$\Lambda = \text{diag}(3.41, 2.36, 2.30, 1.86, 1.53, 1.07, 0.73),$$

and the

$$\text{total inertia} = 11.66 + 5.57 + 5.28 + 3.47 + 2.34 + 1.14 + 0.54 = 30 = \text{tr}(C'C).$$

Noting that $\frac{11.66}{30} = 0.39$ and $\frac{(11.66+5.57+5.28)}{30} = 0.75$, we can assert that 75% of the inertia is accounted for by the first three eigenvalues of $C'C$.

The normalized eigenvectors of $C'C$, which correspond to the nonzero eigenvalues and are denoted by $V = [V_1, \dots, V_7]$, are the following:

$$V_1 = \begin{bmatrix} 0.293826 \\ 0.615045 \\ 0.138401 \\ 0.36352 \\ 0.406951 \\ 0.248836 \\ 0.173839 \\ 0.306906 \\ 0.179291 \end{bmatrix}, V_2 = \begin{bmatrix} -0.711194 \\ 0.512432 \\ -0.0995834 \\ -0.106977 \\ 0.00779839 \\ -0.386116 \\ -0.0833586 \\ 0.041766 \\ 0.228947 \end{bmatrix}, V_3 = \begin{bmatrix} 0.123362 \\ -0.0941084 \\ -0.0879546 \\ 0.638327 \\ -0.521119 \\ -0.428293 \\ 0.0446188 \\ 0.302598 \\ 0.110329 \end{bmatrix}, V_4 = \begin{bmatrix} 0.0732966 \\ 0.107206 \\ 0.601988 \\ -0.03252 \\ -0.388966 \\ 0.167134 \\ -0.164402 \\ -0.357001 \\ 0.534772 \end{bmatrix}, \\ V_5 = \begin{bmatrix} 0.0406003 \\ 0.0238456 \\ -0.124608 \\ 0.110633 \\ 0.0784214 \\ -0.206691 \\ 0.754879 \\ -0.584142 \\ 0.1004 \end{bmatrix}, V_6 = \begin{bmatrix} -0.115587 \\ -0.0128351 \\ -0.572877 \\ 0.408319 \\ 0.0361355 \\ 0.400243 \\ -0.367304 \\ -0.382451 \\ 0.22109 \end{bmatrix}, V_7 = \begin{bmatrix} 0.320998 \\ -0.262335 \\ -0.162227 \\ -0.195339 \\ 0.416228 \\ -0.427087 \\ -0.191702 \\ 0.0724589 \\ 0.604993 \end{bmatrix}.$$

Then, the first two eigenvectors weighted by λ_1 and λ_2 and the points to be plotted are

$$\lambda_1 V_1 = 3.41472 \begin{bmatrix} 0.293826 \\ 0.615045 \\ 0.138401 \\ 0.36352 \\ 0.406951 \\ 0.248836 \\ 0.173839 \\ 0.306906 \\ 0.179291 \end{bmatrix} = \begin{bmatrix} 1.00333 \\ 2.10021 \\ 0.4726 \\ 1.24132 \\ 1.38962 \\ 0.849704 \\ 0.593611 \\ 1.048 \\ 0.612228 \end{bmatrix},$$

$$\lambda_2 V_2 = 2.36098 \begin{bmatrix} -0.711194 \\ 0.512432 \\ -0.0995834 \\ -0.106977 \\ 0.00779839 \\ -0.386116 \\ -0.0833586 \\ 0.041766 \\ 0.228947 \end{bmatrix} = \begin{bmatrix} -1.67911 \\ 1.20984 \\ -0.235114 \\ -0.25257 \\ 0.0184118 \\ -0.911613 \\ -0.196808 \\ 0.0986087 \\ 0.540539 \end{bmatrix};$$

$$\text{Points to be plotted : } \begin{bmatrix} (1.00333, -1.67911) \\ (2.10021, 1.20984) \\ (0.4726, -0.235114) \\ (1.24132, -0.25257) \\ (1.38962, 0.0184118) \\ (0.849704, -0.911613) \\ (0.593611, -0.196808) \\ (1.048, 0.0986087) \\ (0.612228, 0.540539) \end{bmatrix} \leftrightarrow \begin{bmatrix} \text{Men} \\ \text{Women} \\ \text{Underweight} \\ \text{Normal} \\ \text{Overweight} \\ \text{Level 1} \\ \text{Level 2} \\ \text{Level 3} \\ \text{Level 4} \end{bmatrix}.$$

The plot of these points is displayed in Fig. 15.7.1.

It is seen from the points plotted in Fig. 15.7.1 that the categories underweight and educational level 2 are somewhat close to each other, which is indicative of a possible association, whereas the categories underweight and women are the farthest apart.

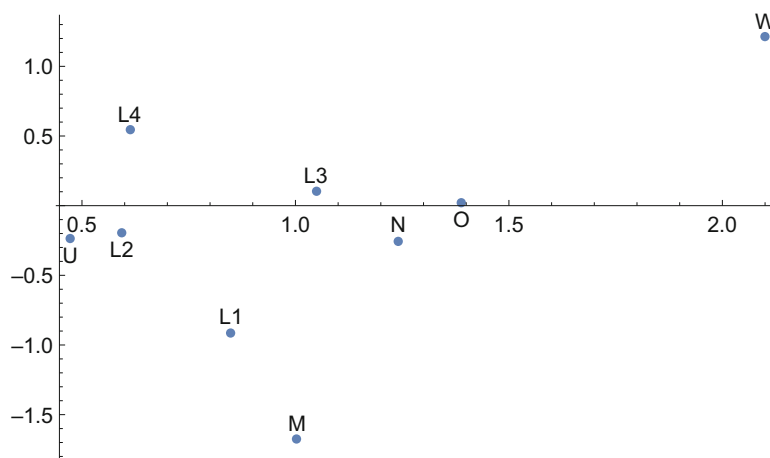


Figure 15.7.1 Multiple contingency plot

Exercises 15 (continued)

15.6. In the following two-way contingency table, where the entries in the cells are frequencies, (1) calculate Pearson's χ^2 statistic and give the representations in (15.5.1)–(15.5.6); (2) plot the row profiles; (3) plot the column profiles:

$\downarrow \rightarrow$	B_1	B_2	B_3	B_4
A_1	10	15	20	15
A_2	15	10	10	5

15.7. Repeat Exercise 15.6 for the following two-way contingency table:

$\downarrow \rightarrow$	B_1	B_2	B_3	B_4
A_1	10	5	15	5
A_2	5	10	10	20
A_3	10	5	10	5
A_4	15	10	5	10

15.8. For the data in (1) Exercise 15.6, (2) Exercise 15.7, and by using the notations defined in Sects. 15.5 and 15.6, compute the following items: Estimates of (i) $A = D_r^{-\frac{1}{2}}(P - RC')D_c^{-1}(P - RC')'D_r^{-\frac{1}{2}}$; (ii) Eigenvalues of A and $\text{tr}(A)$; (iii) Total inertia and proportions of inertia accounted for by the eigenvalues; (iv) The matrix of row-profiles; (v) The matrix of column-profiles, and make comments.

15.9. Referring to Exercises 15.6 and 15.7, plot the row profiles and column profiles and make comments.

15.10. In a used car lot, there are high price, average price and low price cars, the cars come in the following colors: red, white, blue and silver, and the paint finish is either mat or shiny. Fourteen customers bought vehicles from this car lot. Their preferences are given next. (1) Carry out a multiple correspondence analysis, plot the column profiles and make comments; (2) Create individual two-way contingency tables, analyze these tables and make comments. The following is the data where the first column indicates the customer's serial number:

1	Low price	white color	mat finish
2	Low price	red color	shiny finish
3	Average price	silver color	shiny finish
4	High price	red color	shiny finish
5	High price	blue color	shiny finish
6	Average price	white color	mat finish
7	Average price	blue color	mat finish
8	High price	blue color	shiny finish
9	High price	red color	mat finish
10	Average price	silver color	mat finish
11	Low price	white color	shiny finish
12	Average price	white color	mat finish
13	Average price	silver color	shiny finish
14	Low price	white color	shiny finish

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Correction to: Multivariate Statistical Analysis in the Real and Complex Domains



Arak M. Mathai, Serge B. Provost, Hans J. Haubold

Correction to:

Arak M. Mathai, *Multivariate Statistical Analysis in the Real and Complex Domains*,
<https://doi.org/10.1007/978-3-030-95864-0>

The original version of the book was inadvertently published with incorrect abstract. This has now been amended in the chapters.

The updated version of this book can be found at
<https://doi.org/10.1007/978-3-030-95864-0>

© The Author(s) 2022

A. M. Mathai et al., *Multivariate Statistical Analysis in the Real and Complex Domains*,
https://doi.org/10.1007/978-3-030-95864-0_16

C1

Tables of Percentage Points

Tables 1, 2, 3, 4, 5, 6 and 7 contain probabilities and percentage points that are useful for testing a variety of statistical hypotheses encountered in multivariate analysis.

Table 1: Standard normal probabilities. Table entry: probability = $\int_0^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1627	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2882	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3079	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3290	0.3315	0.3340	0.3365	0.3389
1.0	0.3414	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3888	0.3888	0.3906	0.3925	0.3943	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4146	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4278	0.4292	0.4306	0.4319

(continued)

Table 1: (continued): Standard normal probabilities

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4453	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4610	0.4625	0.4633
1.8	0.4641	0.4648	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4762	0.4767
2.0	0.4773	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.2	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4914	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4933	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4986	0.4987	0.4987	0.4988	0.4988	0.4988	0.4989	0.4989	0.4990	0.4990

Table 2: Student- t , right tail. Table entry: $t_{\nu, \alpha}$ where $\int_{t_{\nu, \alpha}}^{\infty} f(t_{\nu}) dt_{\nu} = \alpha$ and $f(t_{\nu})$ is the density of a Student- t distribution having ν degrees of freedom

ν	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.966	2.326	2.576

Table 3: Chisquare, right tail. Table entry: $\chi_{\nu, \alpha}^2$ where $\int_{\chi_{\nu, \alpha}^2}^{\infty} f(\chi_{\nu}^2) d\chi_{\nu}^2 = \alpha$ with $f(\chi_{\nu}^2)$ being the density of a chisquare random variable having ν degrees of freedom

ν	$\alpha = 0.995$	0.99	0.975	0.95	0.10	0.05	0.025	0.01	0.005	0.001
1	0.0000	0.0002	0.0010	0.0039	0.2.71	3.84	5.02	6.63	7.88	10.83
2	0.0100	0.0201	0.0506	0.1030	0.4.61	5.99	7.38	9.21	10.60	13.81
3	0.0717	0.1148	0.2160	0.3520	6.25	7.81	9.35	11.34	12.84	16.27
4	0.2070	0.2970	0.4844	0.7110	7.78	9.49	11.14	13.26	14.86	18.47
5	0.412	0.5543	0.831	1.15	9.24	11.07	12.83	15.09	16.75	20.52
6	0.676	0.872	1.24	1.64	10.64	12.59	14.45	16.81	18.55	22.46
7	0.989	1.24	1.69	2.17	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	17.28	19.68	21.92	24.73	26.76	31.26
12	3.07	3.57	4.40	5.23	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.26	7.26	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	8.90	10.28	11.59	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	33.20	36.42	39.36	42.98	45.56	51.18

(continued)

Table 3: (continued): Chisquare, right tail

ν	$\alpha = 0.995$	0.99	0.975	0.95	0.10	0.05	0.025	0.01	0.005	0.001
25	10.52	11.52	13.12	14.61	34.38	37.65	40.65	44.31	46.93	52.62
26	11.16	12.20	13.84	15.38	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	12.88	14.57	16.15	36.74	40.11	43.19	46.96	49.64	55.48
28	12.46	13.56	15.31	16.93	37.92	41.34	44.46	48.28	50.99	56.89
29	13.12	14.26	16.05	17.71	39.09	42.56	45.72	49.59	52.34	58.30
30	13.79	14.95	16.79	18.49	40.26	43.77	46.98	50.89	53.67	59.70
40	20.71	22.16	24.43	26.51	51.81	55.76	59.34	63.69	66.77	73.40
50	27.99	29.71	32.36	34.76	63.17	67.50	71.42	76.15	79.49	86.66
60	35.53	37.48	40.48	43.19	74.40	79.08	83.30	88.38	91.95	99.61
70	43.28	45.44	48.76	51.74	85.53	90.53	95.02	100.4	104.2	112.3
80	51.17	53.54	57.15	60.39	96.58	101.9	106.6	112.3	116.3	124.8
90	59.20	61.75	65.75	69.13	107.6	113.1	118.1	124.1	128.3	137.2
100	67.33	70.06	74.22	77.93	118.5	124.3	129.6	135.8	140.2	149.4

Table 4: F -distribution, right tail 5% points. Table entry: $b = F_{\nu_1, \nu_2, 0.05}$ where $\int_b^\infty f(F_{\nu_1, \nu_2}) dF_{\nu_1, \nu_2} = 0.05$ with $f(F_{\nu_1, \nu_2})$ being the density of an F -variable having ν_1 and ν_2 degrees of freedom

ν_1 :	1	2	3	4	5	6	7	8	10	12	24	∞
ν_2												
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	241.9	243.9	249.0	254.3
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5	19.5
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.38	3.37	3.29	3.23	3.14	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.61	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.27	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	1.98	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.16	1.96	1.71

(continued)

Table 4: (continued): *F*-distribution, right tail 5% points

ν_1 :	1	2	3	4	5	6	7	8	10	12	24	∞
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.15	1.95	1.69
27	4.21	3.25	2.96	2.73	2.57	2.46	2.37	2.31	2.20	2.13	1.93	1.67
28	4.20	3.34	2.96	2.71	2.56	2.45	2.36	2.29	2.19	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.18	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	1.89	1.62
32	4.15	3.29	2.90	1.67	2.51	2.40	2.31	2.24	2.14	2.07	1.86	1.59
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.12	2.05	1.84	1.57
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.11	2.03	1.82	1.55
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.09	2.02	1.81	1.53
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	2.00	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.70	1.89
100	3.92	3.07	2.63	2.45	2.29	2.18	2.09	2.02	1.91	1.83	1.61	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.52	1.00

Table 5: F -distribution, right tail 1% points. Table entry: $b = F_{\nu_1, \nu_2, 0.01}$ where $\int_b^\infty f(F_{\nu_1, \nu_2}) dF_{\nu_1, \nu_2} = 0.01$ with $f(F_{\nu_1, \nu_2})$ being the density of an F -variable having ν_1 and ν_2 degrees of freedom

$\nu_1 :$	1	2	3	4	5	6	7	8	10	12	24	∞
ν_2												
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6056	6106	6235	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.2	27.1	26.6	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.5	14.4	13.9	13.5
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.05	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.87	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.62	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.81	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.26	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.85	4.71	4.33	3.91
11	0.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.54	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.30	4.16	3.78	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.10	3.96	3.59	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	3.94	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.80	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.69	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.59	3.46	3.08	2.65
18	8.20	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.51	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.43	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.37	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.31	3.17	2.80	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.26	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.21	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.17	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.13	2.99	2.62	2.17

(continued)

Table 5: (continued): *F*-distribution, right tail 1% points

ν_1 :	1	2	3	4	5	6	7	8	10	12	24	∞
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.09	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.06	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.03	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.00	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.98	2.84	2.47	2.01
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	2.93	2.80	2.42	1.96
34	7.45	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.90	2.76	2.38	1.91
36	7.40	5.25	4.38	3.89	3.58	3.35	3.18	3.05	2.86	2.72	2.35	1.87
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.83	2.69	2.32	1.84
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.80	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.63	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.47	2.334	1.95	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.32	2.18	1.79	1.00

Table 6: Testing independence. Let $H_o : \Sigma$ is diagonal in a $N_p(\mu, \Sigma)$ population and $u = \lambda^{2/n}$ where n is the sample size and λ refers to the λ -criterion. The table entries are the 5th and 1st upper percentage points of $w = -[m + (2p + 5)/6] \ln u$ where $m = n - 1$. H_o is rejected for large values of w . The last line wherein $m = \infty$ displays the asymptotic chisquare values as $w \rightarrow \chi_{p(p-1)/2}^2$

5% points								
m	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
3	8.020							
4	7.834	15.22						
5	7.814	13.47	24.01					
6	7.811	13.03	20.44	34.30				
7	7.811	12.85	19.45	28.75	46.05			
8	7.811	12.76	19.02	27.11	38.41	59.25		
9	7.812	12.71	18.80	26.37	36.03	49.42	73.79	
10	7.812	12.68	18.67	25.96	34.91	46.22	61.76	89.92
11	7.813	12.66	18.58	25.71	34.28	44.67	57.68	75.45
12	7.813	12.65	18.52	25.55	33.89	43.78	55.65	70.43
13	7.813	12.64	18.48	25.44	33.63	43.21	54.46	67.87
14	7.813	12.63	18.45	25.36	33.44	42.82	53.69	66.34
15	7.814	12.62	18.43	25.30	33.31	42.55	53.15	65.33
16	7.814	12.62	18.41	25.25	33.20	42.34	52.77	64.63
17	7.814	12.62	18.40	25.21	33.12	42.19	52.48	64.12
18	7.814	12.61	18.38	25.19	33.06	42.06	52.26	63.73
19	7.814	12.61	18.37	25.16	33.01	41.97	52.08	63.43
20	7.814	12.61	18.37	25.14	32.97	41.89	51.94	63.19
∞	7.815	12.59	18.31	25.00	32.67	41.34	51.00	61.66

(continued)

Table 6: Testing Independence

1% points								
m	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
3	11.79							
4	11.41	21.18						
5	11.36	18.27	32.16					
6	11.34	17.54	26.50	44.65				
7	11.34	17.24	24.95	36.09	58.61			
8	11.34	17.10	24.29	33.63	47.05	74.01		
9	11.34	17.01	23.95	32.54	43.59	59.36	90.87	
10	11.34	16.96	23.75	31.95	42.00	54.83	73.03	109.53
11	11.34	16.93	23.62	31.60	41.13	52.70	67.37	88.05
12	11.34	16.90	23.53	31.36	40.59	51.49	64.64	81.20
13	11.34	16.89	23.47	31.20	40.23	50.73	63.06	77.83
14	11.34	16.87	23.42	31.09	39.97	50.22	62.05	75.84
15	11.34	16.86	23.39	31.00	39.79	49.85	61.36	74.56
16	11.34	16.86	23.36	30.94	39.65	49.59	60.86	73.66
17	11.34	16.85	23.34	30.88	39.54	49.38	60.49	73.01
18	11.34	16.85	23.32	30.84	39.46	49.22	60.21	72.52
19	11.34	16.84	23.31	30.81	39.39	49.09	59.99	72.15
20	11.34	16.84	23.30	30.78	39.33	48.99	59.81	71.85
∞	11.34	16.81	23.21	30.58	38.93	48.28	58.57	69.92

Note: For $p = 2$, $w = 1 - r^2$ where r is the sample correlation coefficient. Moreover, $\frac{mr^2}{1-r^2} \sim F_{1,m}$ where $m = n - 1$ and n is the sample size; also refer to Sect. 5.6 and Eq. (6.6.1). Thus, the case $p = 2$ is omitted from the tables as one can then readily obtain the requisite critical values from F or Student- t tables in light of these properties.

Table 7: Testing the equality of the diagonal elements, given that Σ in a $N_p(\mu, \Sigma)$ population is diagonal. Let $v = \lambda^{2/n}$ where λ is the likelihood ratio test statistic and n is the sample size. Letting $f(v)$ denote the density of v , the table entries w are the critical values of v such that $F(w) = \int_0^w f(v) dv = 0.01, 0.02, 0.025, 0.05$, for various values of p and $m = n - 1$

$p = 2$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
2	0.01990	0.03960	0.04937	0.09750
3	0.08083	0.12702	0.14675	0.22852
4	0.15874	0.22173	0.24664	0.34163
5	0.23520	0.30632	0.33318	0.43074
6	0.30387	0.37792	0.40505	0.50053
7	0.36370	0.43784	0.46442	0.55593
8	0.41540	0.48812	0.51378	0.60071
9	0.46009	0.53064	0.55524	0.63751
10	0.49889	0.56694	0.59043	0.66824
11	0.53279	0.59822	0.62062	0.69425
12	0.56258	0.62540	0.64677	0.71654
13	0.58893	0.64921	0.66961	0.73583
14	0.61238	0.67024	0.68972	0.75268
15	0.63336	0.68893	0.70756	0.76753
16	0.65224	0.70564	0.72349	0.78072
17	0.66930	0.72067	0.73778	0.79249
18	0.68479	0.73425	0.75069	0.80307
19	0.69891	0.74659	0.76239	0.81263
20	0.71185	0.75784	0.77305	0.82131
21	0.72372	0.76815	0.78280	0.82923
22	0.73467	0.77761	0.79176	0.83647
23	0.74479	0.78634	0.80000	0.84313
24	0.75417	0.79442	0.80763	0.84927
25	0.76290	0.80190	0.81469	0.85494
26	0.77103	0.80887	0.82126	0.86021
27	0.77862	0.81536	0.82738	0.86511
28	0.78753	0.82143	0.83310	0.86967
29	0.79240	0.82712	0.83845	0.87394
30	0.79867	0.83245	0.84347	0.87794

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 3$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
3	0.03171	0.05272	0.06214	0.10378
4	0.07974	0.11611	0.13109	0.19128
5	0.13649	0.18373	0.20222	0.27252
6	0.19382	0.24775	0.26816	0.34301
7	0.24796	0.30559	0.32688	0.40302
8	0.29756	0.35690	0.37844	0.45403
9	0.34240	0.40217	0.42357	0.49760
10	0.38272	0.44212	0.46314	0.53509
11	0.41895	0.47745	0.49798	0.56759
12	0.45153	0.50885	0.52881	0.59599
13	0.48092	0.53687	0.55624	0.62098
14	0.50571	0.56199	0.58075	0.64314
15	0.53164	0.58462	0.60278	0.66289
16	0.55361	0.60510	0.62267	0.68061
17	0.57369	0.62370	0.64071	0.69658
18	0.59210	0.64066	0.65713	0.71105
19	0.60903	0.65619	0.67214	0.72422
20	0.62464	0.67046	0.68592	0.73625
21	0.63908	0.68361	0.69860	0.74729
22	0.65247	0.69577	0.71031	0.75744
23	0.66492	0.70703	0.72115	0.76682
24	0.67653	0.71750	0.73121	0.77550
25	0.68737	0.72726	0.74059	0.78357
26	0.69752	0.73637	0.74933	0.79108
27	0.70704	0.74490	0.75751	0.79808
28	0.71598	0.75290	0.76518	0.80464
29	0.72440	0.76041	0.77238	0.81078
30	0.73234	0.76749	0.77916	0.81655

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 4$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
4	0.04504	0.06768	0.07726	0.11713
5	0.08692	0.11996	0.13319	0.18503
6	0.13350	0.17421	0.18994	0.24915
7	0.18050	0.22641	0.24369	0.30699
8	0.22568	0.27488	0.29305	0.35825
9	0.26804	0.31915	0.33776	0.40346
10	0.30724	0.35929	0.37801	0.44332
11	0.34328	0.39559	0.41423	0.47858
12	0.37632	0.42843	0.44684	0.50991
13	0.40659	0.45818	0.47628	0.53786
14	0.43435	0.48519	0.50294	0.56292
15	0.45983	0.50980	0.52715	0.58549
16	0.48328	0.53227	0.54921	0.60591
17	0.50490	0.55286	0.56938	0.62447
18	0.52487	0.57178	0.58768	0.64139
19	0.54336	0.58921	0.60490	0.65688
20	0.56052	0.60532	0.62061	0.67111
21	0.57649	0.62025	0.63514	0.68423
22	0.59137	0.63411	0.64863	0.69635
23	0.60527	0.64702	0.66117	0.70759
24	0.61828	0.65906	0.67286	0.71804
25	0.63048	0.67032	0.68378	0.72777
26	0.64194	0.68088	0.69401	0.73686
27	0.65273	0.69079	0.70360	0.74537
28	0.66289	0.70011	0.71262	0.75334
29	0.67249	0.70889	0.72111	0.76084
30	0.68156	0.71715	0.72912	0.76790

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 5$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
5	0.05774	0.08138	0.09103	0.12970
6	0.09523	0.12645	0.13871	0.18581
7	0.13538	0.17235	0.18648	0.23915
8	0.17568	0.21677	0.23215	0.28827
9	0.21473	0.25863	0.27481	0.33284
10	0.25181	0.29753	0.31416	0.37304
11	0.28661	0.33340	0.35026	0.40926
12	0.31908	0.36639	0.38328	0.44191
13	0.34925	0.39668	0.41350	0.47142
14	0.37725	0.42451	0.44117	0.49816
15	0.40323	0.45010	0.46654	0.52246
16	0.42735	0.47369	0.48986	0.54462
17	0.44976	0.49546	0.51134	0.56489
18	0.47061	0.51560	0.53117	0.58350
19	0.49004	0.53426	0.54952	0.60062
20	0.50818	0.55160	0.56655	0.61643
21	0.52513	0.56774	0.58237	0.63107
22	0.54101	0.58280	0.59712	0.64465
23	0.55590	0.59688	0.61089	0.65728
24	0.56989	0.61006	0.62377	0.66907
25	0.58305	0.62243	0.63584	0.68008
26	0.59546	0.63406	0.64718	0.69039
27	0.60717	0.64501	0.65784	0.70007
28	0.61824	0.65533	0.66790	0.70917
29	0.62872	0.66508	0.67738	0.71773
30	0.63864	0.67430	0.68635	0.72582

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 6$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
6	0.06938	0.09356	0.10318	0.14077
7	0.10338	0.13335	0.14494	0.18882
8	0.13892	0.17338	0.18643	0.23468
9	0.17443	0.21223	0.22631	0.27744
10	0.20897	0.24917	0.26394	0.31684
11	0.24203	0.28388	0.29909	0.35294
12	0.27336	0.31628	0.33174	0.38595
13	0.30287	0.34641	0.36198	0.41614
14	0.33057	0.37440	0.38997	0.44376
15	0.35653	0.40038	0.41587	0.46908
16	0.38082	0.42451	0.43987	0.49235
17	0.40357	0.44694	0.46213	0.51377
18	0.42487	0.46782	0.48280	0.53355
19	0.44484	0.48729	0.50204	0.55184
20	0.46357	0.50546	0.51997	0.56881
21	0.48117	0.52245	0.53672	0.58458
22	0.49772	0.53837	0.55238	0.59927
23	0.51330	0.55330	0.56706	0.61298
24	0.52799	0.56734	0.58084	0.62581
25	0.54186	0.58054	0.59379	0.63784
26	0.55497	0.59299	0.60599	0.64913
27	0.56738	0.60474	0.61750	0.65975
28	0.57914	0.61585	0.72837	0.66976
29	0.59030	0.62637	0.63865	0.67920
30	0.60089	0.63634	0.64899	0.68813

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 7$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
7	0.07993	0.10433	0.11387	0.15046
8	0.11104	0.14004	0.15112	0.19258
9	0.14306	0.17567	0.18792	0.23290
10	0.17492	0.21029	0.22340	0.27081
11	0.20598	0.24341	0.25712	0.30613
12	0.23587	0.27478	0.28890	0.33886
13	0.26439	0.30431	0.31868	0.36912
14	0.29144	0.33202	0.34653	0.39708
15	0.31703	0.35797	0.37253	0.42293
16	0.34117	0.38226	0.39680	0.44685
17	0.36394	0.40499	0.41946	0.46901
18	0.38539	0.42628	0.44063	0.48957
19	0.40562	0.44623	0.46043	0.50870
20	0.42468	0.46495	0.47898	0.52651
21	0.44267	0.48252	0.49637	0.54313
22	0.45966	0.49905	0.51270	0.55867
23	0.47572	0.51461	0.52805	0.57323
24	0.49091	0.52928	0.54251	0.58688
25	0.50529	0.54312	0.55615	0.59972
26	0.51892	0.55621	0.56902	0.61180
27	0.53186	0.56860	0.58120	0.62319
28	0.54415	0.58033	0.59272	0.63394
29	0.55584	0.59147	0.60365	0.64412
30	0.56697	0.60204	0.61402	0.65375

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 8$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
8	0.08947	0.11391	0.12334	0.15897
9	0.11816	0.14632	0.15699	0.19653
10	0.14735	0.17850	0.19012	0.23256
11	0.17632	0.20980	0.22215	0.26665
12	0.20460	0.23985	0.25274	0.29867
13	0.23192	0.26849	0.28175	0.32859
14	0.25810	0.29563	0.30913	0.35650
15	0.28309	0.32126	0.33492	0.38250
16	0.30686	0.34544	0.35916	0.40672
17	0.32942	0.36820	0.38194	0.42931
18	0.35081	0.38965	0.40335	0.45038
19	0.37108	0.40985	0.42348	0.47006
20	0.39028	0.42878	0.44241	0.48848
$p = 9$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
9	0.09813	0.12249	0.13178	0.16652
10	0.12474	0.15218	0.16250	0.20044
11	0.15160	0.18155	0.19267	0.23304
12	0.17821	0.21015	0.22189	0.26404
13	0.20421	0.23770	0.24990	0.29332
14	0.22939	0.26406	0.27660	0.32088
15	0.25362	0.28917	0.30195	0.34676
16	0.27685	0.31302	0.32596	0.37104
17	0.29904	0.33563	0.34865	0.39380
18	0.32020	0.35704	0.37010	0.41515
19	0.34036	0.37731	0.39036	0.43619
20	0.35955	0.39650	0.40950	0.45401

(continued)

Table 7: (continued): Testing the Equality of the Diagonal Elements, Given that Σ in a $N_p(\mu, \Sigma)$ Population is Diagonal

$p = 10$	$F(w) = 0.01$	$F(w) = 0.02$	$F(w) = 0.025$	$F(w) = 0.05$
m	w	w	w	w
10	0.10602	0.13021	0.13936	0.17326
11	0.13083	0.15762	0.16763	0.20420
12	0.15574	0.18467	0.19536	0.23399
13	0.18036	0.21101	0.22224	0.26243
14	0.20445	0.23646	0.24810	0.28943
15	0.22782	0.26090	0.27285	0.31498
16	0.25039	0.28428	0.29645	0.33910
17	0.27209	0.30658	0.31890	0.36184
18	0.29291	0.32781	0.34022	0.38329
19	0.31284	0.34800	0.36047	0.40351
20	0.33189	0.36721	0.37968	0.42259

Note: For large sample size n , we may utilize the approximation $-2 \ln \lambda \approx \chi_{p-1}^2$ for the null distribution of λ , where $\lambda = v^{n/2}$ is the lambda criterion, n is the sample size and χ_{p-1}^2 denotes a central chisquare random variable having $p - 1$ degrees of freedom. The null hypothesis will be rejected for large values of $-2 \ln \lambda$ since it is rejected for small values of the test statistics λ or v .

Some Additional Reading Materials

Books

- Adachi, Kohei (2016): *Matrix-based Introduction to Multivariate Data Analysis*. Singapore, Springer. <https://doi.org/10.1007/978-981-10-2341-5>
- Agresti, Alan (2013): *Categorical Data Analysis* (3rd ed.). Hoboken, Wiley-Interscience.
- Anderson, Theodore W. (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd edition. New York, Wiley-Interscience.
- Atkinson, Anthony C., Riani, Marco, and Cerioli, Andrea (2004): *Exploring Multivariate Data with the Forward Search*, New York, Springer.
- Bartholomew, David J. (2008): *Analysis of Multivariate Social Science Data* (2nd ed.). Boca Raton, CRC Press.
- Bishop, Yvonne M., Fienberg, Stephen E., and Holland, Paul W. (2007): *Discrete Multivariate Analysis: Theory and Practice* (1st ed.). New York, Springer. <https://doi.org/10.1007/978-0-387-72806-3>
- Brown, Bruce L. (2012): *Multivariate Analysis for the Biobehavioral and Social Sciences: a Graphical Approach*. Hoboken, Wiley.
- Chatfield, Chris, and Collins, A. J. (2018): *Applied Multivariate Analysis : Using Bayesian And Frequentist Methods Of Inference*. Boca Raton, Chapman & Hall/CRC.
- Cox, David R., and Wermuth, Nanny (1996): *Multivariate Dependencies: Models, Analysis and Interpretation*. London, Chapman and Hall.
- Cox, Trevor F. (2005): *An Introduction to Multivariate Data Analysis*. London, Hodder Arnold.
- Dugard, P. , Todman, J., and Staines, H. (2010). *Approaching Multivariate Analysis, a Practical Introduction*, 2nd edn. London, Routledge.

- Everitt, Brian (2011): *Cluster Analysis*, (5th ed.). Hoboken, Wiley. <https://doi.org/10.1002/9780470977811>
- Everitt, Brian, and Dunn, Graham (2001): *Applied Multivariate Data Analysis* (2nd ed.). London, Arnold.
- Everitt, Brian., and Hothorn, Torsten. (2011): *An Introduction to Applied Multivariate Analysis with R*. New York, Springer. <https://doi.org/10.1007/978-1-4419-9650-3>
- Flury, Bernhard (1997): *A First Course in Multivariate Statistics*. New York, Springer.
- Fujikoshi, Yasunori, Ulyanov, Vladimir V., and Shimizu, Ryoichi. (2010): *Multivariate Statistics: High-dimensional and Large-sample Approximations*. Hoboken, Wiley.
- Green, Paul E. (2014): *Mathematical Tools for Applied Multivariate Analysis*. Academic Press.
- Gupta, Arjun K., and Nagar, Daya K. (1999): *Matrix Variate Distributions*. Boca Raton, Chapman & Hall/ CRC.
- Hair, Joseph F. (2006): *Multivariate Data Analysis* (6th ed.). Upper Saddle River, Pearson/Prentice Hall.
- Härdle, Wolfgang K., and Simar, Léopold (2015): *Applied Multivariate Statistical Analysis*, 4th edition. Berlin, Springer.
- Hougaard, Philip (2000): *Analysis of Multivariate Survival Data*. New York, Springer.
- Huberty, Carl J., and Olejnik, Stephen (2006): *Applied MANOVA and Discriminant Analysis*. (2nd ed.). Hoboken, Wiley-Interscience. <https://doi.org/10.1002/047178947X>
- Izenman, Alan J. (2008): *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning* (1st ed.). New York, Springer. <https://doi.org/10.1007/978-0-387-78189-1>
- Johnson, Mark E. (1987): *Multivariate Statistical Simulation*. New York, Wiley.
- Johnson, Richard A. and Wichern, Dean W. (2007): *Applied Multivariate Statistical Analysis*, 6th edition. Prentice Hall, New Jersey.
- Kollo, Tõnu, and von Rosen, Dietrich (2005): *Advanced Multivariate Statistics with Matrices*. Dordrecht, Springer.
- Kotz, Samuel., Balakrishnan, N., and Johnson, Norman, L. (2000): *Continuous Multivariate Distributions. Vol. 1, Models and Applications* (2nd ed.). New York, Wiley. <https://doi.org/10.1002/0471722065>

- Krzanowski, W. J. (2000): *Principles of Multivariate Analysis: a User's Perspective* (Rev. ed.). Oxford, Oxford University Press.
- Manly, Bryan F. J., and Navarro Alberto, Jorge A. (2016): *Multivariate Statistical Methods: A Primer*, Fourth Edition. Boca Raton, Chapman & Hall/CRC Press.
- Meyers, Lawrence, Gamst, Glenn, and Guarino, A. J. (2016): *Applied Multivariate Research: Design and Interpretation*, 3rd edition. Sage Publications.
- Press, S. James (2005): *Applied Multivariate Analysis : Using Bayesian And Frequentist Methods of Inference*, 3rd edition. Dover Publications.
- Olive, David J. (2017): *Robust Multivariate Analysis* (1st ed.). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-68253-2>
- Raykov, Tenko, and Marcoulides, George A. (2008): *An Introduction to Applied Multivariate Analysis*. Abington, Taylor and Francis.
- Rencher, Alvin C., and Christensen, William F. (2012): *Methods of Multivariate Analysis*. New York, Wiley.
- Schafer, Joseph L. (1997): *Analysis of Incomplete Multivariate Data* (1st ed.). London, Chapman and Hall.
- Scott, David. W. (2014): *Multivariate Density Estimation: Theory, Practice, and Visualization* (Second edition). Hoboken, Wiley.
- Serdobolskii, Vadim. (2000): *Multivariate Statistical Analysis: a High-dimensional Approach*. Dordrecht, Kluwer Academic Publishers.
- Spencer, Neil H.(2013): *Essentials of Multivariate Data Analysis*. Boca Raton, Chapman & Hall/CRC Press.
- Srivastava, M. S., and Khatri, C. G. (1979): *An Introduction to Multivariate Statistics*. New York, North-Holland.
- Tabachnick, Barbara G., Fidell, Linda S., and Ullman, Jodie B. (2019): *Using Multivariate Statistics* (Seventh edition). Boston, Pearson.
- Timm, Neil H. (2007): *Applied Multivariate Analysis*. New York, Springer-Verlag.
- Wadsack, Peter, and Kres, Heinz. (2012): *Statistical Tables for Multivariate Analysis: A Handbook with References to Applications*. Dordrecht, Springer.
- Wei, William W. S. (2019): *Multivariate Time Series Analysis and Applications*. Hoboken, Wiley.

Some Recent Research Papers

Habti Abeida and Jean-Pierre Delmas (2020): Efficiency of subspace-based estimators for elliptical symmetric distributions. *Signal Processing*, **174**, 107644 (9 pages)

Olivier Besson (2021): On the distributions of some statistics related to adaptive filters trained with t-distributed samples. *arXiv:2101.10609v2 [math.ST]* 3 March 2021

Taras Bodnar and Yarema Okhrin (2008): Properties of the singular, inverse and generalized inverse partitioned Wishart distributions. *Journal of Multivariate Analysis*, **99**, 2389–2405.

John T. Chen and Arjun K. Gupta (2005): Matrix variate skew normal distributions. *Statistics*, **39**(3), 247–253. <https://doi.org/10.1080/02331880500108593>

Marco Chiani (2014): Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution. *Journal of Multivariate Analysis*, **129**, 69–81.

A. W. Davis (1972): On the marginal distributions of the latent roots of the multivariate beta matrix, *The Annals of Mathematical Statistics*, **43**(5), 1664–1670.

Michel Denuit and Yang Lu (2021): Wishart-gamma random effects models with applications to nonlife insurance. *The Journal of Risk and Insurance*, **88**(2), 443–481. <https://doi.org/10.1111/jori.12327>

José A. Diaz-Garcia and Ramón Gutiérrez Jáimez (2007): Singular matrix variate beta distribution. *Multivariate Analysis*, **99**, 637–648.

José A. Diaz-Garcia and Ramón Gutiérrez Jáimez (2009): Doubly singular matrix variate beta type I and II and singular inverted matrix-variate t distributions. *arXiv:0904.2147v1 [math.ST]* 14 April 2009.

Jörn Diedrichsen, Serge B. Provost and Hossein Zareamoghaddam (2021): On the distribution of cross-validated Mahalanobis distances. *arXiv: 1607.01371*

Gilles R. Ducharme, Pierre Lafaye de Micheaux and Bastien Marchina (2016): The complex multinormal distribution, quadratic forms in complex random vectors and an omnibus goodness-of-fit test for the complex normal distribution. *Annals of the Institute of Statistical Mathematics*, **68**, 77–104.

Alan Edelman (1991): The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Linear Algebra and its Applications*, **159**, 55–80.

- Junhan Fang and Grace Y. Yi (2022): Regularized matrix-variate logistic regression with response subject to misclassification. *Journal of Statistical Planning and Inference*, **217**, 106–121. <https://doi.org/10.1016/j.jspi.2021.07.001>
- Daniel R. Fuhrmann (1999): Complex random variables and stochastic processes. Section 60 of the book *Digital Signal Processing Handbook*, CRC Press LLC.
- Nina Singhal Hinrichs and Vijay S. Pande (2007): Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *The Journal of Chemical Physics*, **126**, 244101 (11 pages).
- Velimir Illić, Jan Korbel, Shamik Gupta and A. M. Scarfone (2021): An overview of generalized entropic forms. *arXiv:2102.10071v1[cond-mat.stat-mech]*, 19 Feb 2021.
- Anis Iranmanesh, M. Arashi, D.K. Nagar and S.M.M. Tabatabaey (2013): On inverted matrix variate gamma distribution. *Communications in Statistics - Theory and Methods*, **42**, 28–31.
- Oliver James and Heung-No Lee (2021): Concise probability distributions of eigenvalues of real-valued Wishart matrices. [arXiv.org/ftp/arXiv/papers/1402/1402.6757.pdf](https://arxiv.org/ftp/arXiv/papers/1402/1402.6757.pdf)
- Romuald A. Janik and Maciej A. Nowak (2003): Wishart and anti-Wishart random matrices. *Journal of Physics A: Mathematical and General*, **36**(12), 3629.
- Ian M. Johnstone (2001): On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, **29**(2), 295–327
- Tatsuya Kubokawa and Muni S. Srivastava (2008): Estimation of the precision matrix of singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, **99**, 1906–1928.
- Nojun Kwak (2008): Principal component analysis based on L1-norm maximization. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **30**(9), 1672–1680.
- Arak M. Mathai and Hans J. Haubold (2020): Mathematical aspects of Krätzel integral and Krätzel transform. *Mathematics*, **8**, 526. <https://doi.org/10.3390/math8040526>
- Shinichi Mogami, Daichi Kitamura, Yoshiki Mitsui, Narihiro Takamune, Hiroshi Saruwatari and Nobutaka Ono (2017): Independent low-rank matrix analysis based on complex Student-t distribution for blind audio source separation. *arXiv:1708.04795v1 [cs.SD]* 16 Aug 2017

- Daya K. Nagar, Alejandro Roldán-Correa and Arjun K. Gupta (2013). Extended matrix variate gamma and beta functions. *Journal of Multivariate Analysis*, **122**, 53–69.
- Felting Nie and Heng Huang (2016): Non-greedy L21-norm maximization for principal component analysis. *arXiv:1603.08293v1[cs,LG]* 28 March 2016.
- Vladimir Ostashev, D. Keith Wilson, Matthew J. Kamrath, M. J. and Chris L. Pettit, (2020): Modeling of signals scattered by turbulence on a sensor array using the matrix gamma distribution. *The Journal of the Acoustical Society of America*, **148**, 2476. <https://doi-org.proxy1.lib.uwo.ca/10.1121/1.5146860>
- Victor Pérez-Abreu and Robert Stelzer (2014): Infinitely divisible multivariate and matrix Gamma distributions. *Journal of Multivariate Analysis*, **130**, 155–175.
- Serge B. Provost and John N. Haddad (2019): A recursive approach for determining matrix inverses as applied to causal time series processes, *Metron*, **77**, 53–62.
- Tharm Ratnarajah (2005): Complex singular Wishart matrices and multiple-antenna systems, IEEE/SP 13th Workshop on Statistical Signal Processing, 1032–1037. doi: 10.1109/SSP.2005.1628747
- Tharm Ratnarajah and R. Vaillancourt (2005): Complex singular Wishart matrices and applications. *Computers and Mathematics with Applications*, **50**, 399–411.
- Maria Ribeiro, Teresa Henriques, Luisa Castro, André Souto, Luis Antunes, Cristina Costa-Santos and Andreia Teixeira (2021): The entropy universe. *Entropy*, **23**, 222 (35 pages), <https://doi.org/10.3390/e232020222>
- Yo Sheena, A. K. Gupta and Y. Fujikoshi (2004): Estimation of the eigenvalues of noncentrality parameter in matrix variate noncentral beta distribution. *Annals of the Institute of Statistical Mathematics*, **56**(1), 101–125.
- Jiarong Shi, Xiuyun Zheng and Wei Yang (2017): Survey on probabilistic models of low-rank matrix factorization. *Entropy*, **424**. <https://doi.org/10.3390/e19080424>.
- Martin Singull and Timo Koski (2012): On the distribution of matrix quadratic forms. *Communications in Statistics -Theory and Methods*, **41**(18), 3403–3415. <https://doi.org/10.1080/03610926.2011.563009>.
- Muni S. Srivastava (2003): Singular Wishart and multivariate beta distributions. *The Annals of Statistics*, **31**(5), 1537–1560.

- Tim Wirtz, G. Akemann, Thomas Guhr, M. Kieburg and R. Wegner (2015): The smallest eigenvalue distribution in the real Wishart-Laguerre ensemble with even topology. *Journal of Physics A: Mathematical and Theoretical*, **48**. <https://doi.org/10.1098/1751-8113/48/24/245202>.
- Soonyu Yu, Jaena Ryu and Kyoohong Park (2014): A derivation of anti-Wishart distribution. *Journal of Multivariate Analysis*, **131**, 121–125.
- Alberto Zanella, Marco Chiani and Moe Z. Win (2009): On the marginal distribution of the eigenvalues of Wishart matrices. *IEEE Transactions on Communications*, **57**, 1050–1060.
- Lorenzo Zaninetti (2020): New probability distributions in astrophysics: III. The truncated Maxwell-Boltzmann distributions. *International Journal of Astronomy and Astrophysics*, **10**, 191–202.
- Lorenzo Zaninetti (2021): New probability distributions in astrophysics: V. The truncated Weibull distribution. *International Journal of Astronomy and Astrophysics*, **11**, 133–149.
- Xingfu Zhang, Hyukjun Gweon and Serge B. Provost (2020): Threshold moving approaches for addressing the class imbalance problem and their application to multi-label classification, pp. 72–77. *ICAIP2020*, <https://doi.org/10.1145/3441250.3441274>

Author Index

A

Abeida, A., 910
Adachi, K., 907
Agresti, A., 907
Akemann, G., 913
Allman, E.S., 686
Anderson, T.W., 457, 566, 685, 907
Antunes, L., 912
Arashi, D.K., 911
Atkinson, A.C., 907

B

Balakrishnan, N., 510, 908
Barnes, C.A., 495
Bartholomew, D.J., 907
Besson, O., 910
Bishop, V.M., 907
Bodnar, T., 910
Brown, B.L., 907

C

Castro, L., 912
Cerioli, A., 907
Chatfield, C., 907
Chattopadhyay, A.K., 627
Chen, J.T., 910
Chen, Y., 686

Chiani, M., 627, 910, 913
Christensen, W.F., 909
Clayton, D.D., 495
Clemm, D.S., 627
Collins, A.J., 907
Costa-Santos, C., 912
Cox, D.R., 907
Cox, T.F., 907
Crichfield, C., 495

D

Davis, A.W., 435, 627, 910
de Micheaux, P.L., 910
Delmas, J.-P., 910
Denuit, M., 910
Diaz-Garcia, J.A., 910
Diedrichsen, J., 910
Ducharme, G., 910
Dugard, P., 907
Dunn, G., 908

E

Edelman, A., 627, 910
Everitt, B., 908

F

Fang, J., 911
Fidell, L.S., 909

Field, J.B., 435

Fienberg, S.E., 907

Flury, B., 908

Fowler, W.A., 495

Fuhrmann, D.R., 911

Fujikoshi, Y., 908, 912

G

Gamst, G., 909

Green, P.E., 908

Guarino, A.J., 909

Guhr, T., 913

Gupta, A.K., 908, 910, 912

Gupta, S., 911

Gweon, H., 913

H

Habti, A., 913

Haddad, J.N., 912

Hair, J.F., 908

Härdle, W.K., 908

Haubold, H.J., 1, 105, 108, 110, 113,
125, 139, 357, 359, 360, 367, 369, 377,
395, 469, 478, 495, 502, 510, 595, 599,
611, 647, 652, 657, 759, 763, 771, 833,
870, 911

Henriques, T., 912

Hinrichs, N.S., 911

Holland, P.W., 907

Hothorn, T., 908

Hougaard, P., 908

Huang, H., 620, 912

Huberty, C.J., 908

I

Ihara, M., 686

Illić, V., 911

Iranmanesh, A., 911

Izenman, A.J., 908

J

Jáimez, R.G., 910

James, A.T., 627

James, O., 627, 911

Janik, R.A., 911

Johnson, M.E., 908

Johnson, N.L., 908

Johnson, R.A., 908

Johnstone, I.M., 627, 911

Jorge, A., 909

K

Kamrath, M.J., 912

Kano, Y., 686

Katiyar, R.S., 449, 457

Khatri, C.G., 627, 909

Kieburg, M., 913

Kitamura, D., 911

Kollo, T., 908

Kondo, T., 537, 538

Korbel, J., 911

Korin, B.P., 435

Koski, T., 912

Kotz, S., 908

Kres, H., 909

Krishnaiah, P.R., 627

Krzanowski, W.J., 909

Kubokawa, T., 911

Kwak, N., 620, 911

L

Lai, C.D., 510

Lee, H.-N., 911

Lee, W.-N., 627

Li, X., 686

Lu, Y., 910

M

Manly, B.F.J., 909

Marchina, B., 910
Marcoulides, G.A., 909
Marshall, A.W., 510
Mathai, A.M., 1, 44, 77, 81, 105, 108,
110, 113, 125, 139, 357, 359, 360, 367,
369, 377, 395, 469, 478, 495, 502, 510,
595, 599, 611, 647, 652, 657, 759, 763,
771, 833, 870, 911
Matias, C., 686
Meyers, L., 909
Mitsui, Y., 911
Mogami, S., 911

N

Nagao, H., 435
Nagar, D.K., 908, 911, 912
Nagarsenker, B.N., 435
Navarro, A., 909
Nie, F., 912
Nowak, M.A., 911

O

Okhrin, Y., 910
Olejnik, S., 908
Olive, D.J., 909
Olivier, B., 910
Olkin, I., 510
Ono, N., 911
Ostashev, V., 912

P

Pais, A., 495
Pande, V.S., 911
Park, K., 913
Pérez-Abreu, V., 912
Pettit, C.L., 912
Pillai, K.C.S., 435, 627
Potthoff, R.F., 842
Press, S.J., 909

Princy, T., 495
Provost, S.B., 81, 170, 510, 533, 910,
912, 913

R

Rathie, P.N., 201, 436, 443, 449, 457
Ratnarajah, T., 912
Raykov, T., 909
Rencher, A.C., 909
Rhodes, J.A., 709
Riani, M., 907
Ribeiro, M., 912
Roldán-Correa, A., 912
Roy, S.N., 842
Rubin, H., 685
Ryu, J., 913

S

Saruwatari, H., 908, 911, 912
Saxena, R.K., 110, 113, 443, 452, 457,
478, 771, 833
Scarfone, A.M., 911
Schafer, J.L., 909
Schramm, D.N., 495
Scott, D.W., 909
Serdobolskii, V., 909
Sheena, Y., 912
Shi, J., 620, 912
Shimizu, R., 908
Simar, L., 908
Singull, M., 912
Souto, A., 912
Spencer, N.H., 909
Srivastava, M.S., 909, 911, 912
Staines, H., 907
Stelzer, R., 912
Sugiura, N., 435

T

Tabachnick, B.G., 909
Tabatabaey, S.M.M., 911
Takamune, N., 911
Teixeira, A., 912
Timm, N.H., 909
Todman, J., 907

U

Ullman, J.B., 909
Ulyanov, V.V., 908

V

Vaillancourt, R., 912
von Rosen, D., 908

W

Wadsack, P., 909
Waikar, V.B., 627
Wegge, L.L., 686

Wegner, R., 913

Wei, W.W.S., 909

Wermuth, N., 907

Wichern, D.W., 908

Wilson, D.K., 912, 913

Win, M.Z., 913

Wirtz, T., 913

Y

Yang, W., 912

Yi, G.Y., 911

Yu, S., 913

Z

Zanella, A., 913

Zaninetti, L., 913

Zareamoghaddam, H., 910

Zhang, S., 686

Zhang, X., 913

Zheng, X., 912

Subject Index

A

Analysis of variance (ANOVA), 763
Asymptotic distribution, 436
Asymptotic normality, 203

B

Bayes' rule, 714
Bessel function, 334
Beta variable, 72

C

Canonical correlation, 641
Canonical correlation matrix, 669
Cauchy density, 76
Central limit theorem, 119
Characteristic function, 212–214
Chebyshev's inequality, 115
Chisquare, 62, 65
Chisquaredness, 78, 172
Classification, 711–757
Cofactor, 21
Confidence interval, 125
Consistency, 198
Correlation coefficient, 355
Cramér-Rao inequality, 201
Critical region, 396

D

Design of experiment model, 681
Determinant, 12
Determinant, absolute value of, 43
Differential, 40
Dirichlet density, 378, 381
Discriminant function, 727
Duplication formula, 527

E

Eigenvalue, eigenvector, 32
Elliptically contoured distribution, 206
Estimation, point, 120

F

Factor analysis, 679
Factor analysis model, 682
Factor loadings, 687
Factor space, 683
Fractional integral, first kind, 371
Fractional integral, second kind, 368
F-variable, 69

G

Gamma function, multiplication formula of, 438
Gaussian, multivariate, 129
Generalized variance, 347

Geometrical probability, 355

G-function, 112, 329

Growth curve, 824

H

Hazard function, 72

Hermitian forms, independence of, 175

H-function, 112

Hotelling's T^2 , 415

I

Identification problem, 685

Interval estimation, 124

Invariance, 645

Inverse Wishart density, 349

Iterative procedure, 656–663

J

Jacobian, 43

K

Krätzel function, 334

L

$L1$, $L2$ norms, 618–620

Lambda criterion, 411

Likelihood function, 396

Linear hypothesis, 463–473

M

Matrix, 2

in the complex domain, 35

diagonal, identity, 4

elementary, 24

Hermitian, skew Hermitian, 35

idempotent, nilpotent, 34

inverse of, 23

multiplication, 5

null, square, rectangular, 4

partitioned, 29

scalar multiplication of, 4

singular values of, 38

symmetric, skew symmetric, 34

transpose of, 8

triangular, 4

Matrix-variate gamma, complex, 291

Matrix-variate gamma, real, 224

Matrix-variate, rectangular, 493

Maximum likelihood estimators (MLE), 121

Maxwell-Boltzmann distribution, 499

M-convolution, 377

Method of maximum likelihood, 120

Method of moments, 120

Misclassification, 712

Mittag-Leffler function, 334

Moment generating function (mgf), 57

Multiple correlation coefficient, 364

Multivariate analysis of variance (MANOVA), 481, 759

N

Null hypothesis, 396

O

Oblique factors, 686

Orthogonal factors, 707

P

Partial correlation, 364

Pathway models, 370

Patterned matrices, 487

Polar coordinates, 500

Poles, 322

Power of a test, 397

Power transformation, 71

Principal components, 597

Products and ratios of matrices, 366

Profiles, coincident, 814

Profiles, level, 814
Profiles, parallel, 814
Pseudo Dirichlet model, 383
 p -value, 408

Q

Quadratic form, 78
Quadratic forms, independence of, 78,
169

R

Raleigh distribution, 494
Rao-Blackwell theorem, 200
Regression, 360
Regression model, 680
Relative efficiency, 199
Reliability function, 72
Repeated measures design, 824
Residues, 322

S

Scalar quantity, 4
Singular gamma, 580

Singular gamma, complex, 588
Stirling's formula, 504
Student-t, 73
Sufficiency, 198

T

Triangularization, 340
Type-1, type-2 error, 396

U

Unbiasedness, 196

V

Vector, row, column, 6
Vectors, orthogonal, orthonormal, 10

W

Weak law of large numbers, 118
Wedge product, 40
Weyl integral, 375
Wishart density, 335

Z

Zonal polynomial, 478