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 $Y \neq O$,

$$Y'AY > 0, \ A > O \ \text{(positive definite)}$$
$$Y'AY \geq 0, \ A \geq O \ \text{(positive semi-definite)} \qquad (3.1.1)$$
$$Y'AY < 0, \ A < O \ \text{(negative definite)}$$
$$Y'AY \leq 0, \ A \leq O \ \text{(negative semi-definite)}.$$

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,

$$Y^*AY > 0, \ A > O \ \text{(Hermitian positive definite)}$$
$$Y^*AY \geq 0, \ A \geq O \ \text{(Hermitian positive semi-definite)}$$
$$Y^*AY < 0, \ A < O \ \text{(Hermitian negative definite)} \qquad (3.1.2)$$
$$Y^*AY \leq 0, \ A \leq O \ \text{(Hermitian negative semi-definite)},$$

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 $\tilde{x}_1, \ldots, \tilde{x}_p$. Let $\tilde{x}_j = x_{j1} + i x_{j2}$ 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 $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_p)'$ 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 $E(\tilde{X}) \equiv \tilde{\mu} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_p)'$. Then we have

$$\Sigma = \text{Cov}(\tilde{X}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \ldots & \sigma_p^2 \end{bmatrix}$$

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:

$$f(\tilde{X}) = \frac{1}{\pi^p |\det(\Sigma)|} e^{-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu})} \tag{3.1a.1}$$

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 $-\tilde{Y}^* \tilde{Y} = -[|\tilde{y}_1|^2 + \cdots + |\tilde{y}_p|^2]$. But

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

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

$$\int \ldots \int |\tilde{y}_1|^2 e^{-(|\tilde{y}_1|^2 + \cdots + |\tilde{y}_p|^2)} d\tilde{y}_1 \wedge \ldots \wedge d\tilde{y}_p$$

$$= \left\{ \prod_{j=2}^{p} e^{-|\tilde{y}_j|^2} d\tilde{y}_j \right\} \int_{\tilde{y}_1} |\tilde{y}_1|^2 e^{-|\tilde{y}_1|^2} d\tilde{y}_1.$$

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 $|\tilde{y}_1|^2 = y_{11}^2 + y_{12}^2$, $\tilde{y}_1 = y_{11} + i y_{12}$, $i = \sqrt{(-1)}$, 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

$$\int_{\tilde{y}_1} |\tilde{y}_1|^2 e^{-|\tilde{y}_1|^2} d\tilde{y}_1 = \left(\int_{r=0}^{\infty} r(r^2) e^{-r^2} dr \right) \left(\int_{\theta=0}^{2\pi} d\theta \right), \text{ (letting } u = r^2)$$

$$= (2\pi) \left(\frac{1}{2} \int_0^{\infty} u e^{-u} du \right) = (2\pi) \left(\frac{1}{2} \right) = \pi.$$

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:

$$\Sigma = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix}, \ \tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}, \ E[\tilde{X}] = \begin{bmatrix} 1+2i \\ 2-i \end{bmatrix} = \tilde{\mu}.$$

Show that Σ is Hermitian positive definite so that it can be a covariance matrix of $\tilde{X}$, that is, $\mathrm{Cov}(\tilde{X}) = \Sigma$. 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

$$\Sigma' = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}, \ \Sigma^* = \bar{\Sigma}' = \begin{bmatrix} 2 & 1+i \\ 1-i & 3 \end{bmatrix} = \Sigma$$

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 $\Sigma^{-1} = \frac{1}{\det(\Sigma)}(\mathrm{Cof}(\Sigma))'$ where $\mathrm{Cof}(\Sigma)$ 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,

$$\Sigma^{-1} = \frac{1}{4}\begin{bmatrix} 3 & -(1+i) \\ -(1-i) & 2 \end{bmatrix}, \ \Sigma^{-1}\Sigma = I. \tag{ii}$$

The exponent in a bivariate complex Gaussian density being $-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu})$, we have

$$-(\tilde{X} - \tilde{\mu})^* \Sigma^{-1}(\tilde{X} - \tilde{\mu}) = -\frac{1}{4}\{3\,[\tilde{x}_1 - (1+2i)]^*[\tilde{x}_1 - (1+2i)]$$
$$- (1+i)\,[\tilde{x}_1 - (1+2i)]^*[\tilde{x}_2 - (2-i)]$$
$$- (1-i)\,[\tilde{x}_2 - (2-i)]^*[\tilde{x}_1 - (1+2i)]$$
$$+ 2\,[\tilde{x}_2 - (2-i)]^*[\tilde{x}_2 - (2-i)]\}. \tag{iii}$$

Thus, the density of the $\tilde{N}_2(\tilde{\mu}, \Sigma)$ vector whose components can assume any complex value is

$$f(\tilde{X}) = \frac{e^{-(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})}}{4\pi^2} \tag{3.1a.2}$$

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 $x_1, .., x_p$ be real scalar variables and X be a $p \times 1$ vector with $x_1, \ldots, x_p$ as its elements, that is, $X' = (x_1, \ldots, x_p)$. Let $L' = (a_1, \ldots, a_p)$ where $a_1, \ldots, a_p$ are arbitrary real scalar constants. Consider the linear function $u = L'X = X'L = a_1x_1 + \cdots + a_px_p$. 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 $u = L'X$, $E[u] = L'E[X] = L'\mu$, $\mu' = (\mu_1, \ldots, \mu_p)$, $\mu_j = E[x_j]$, $j = 1, \ldots, p$, 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 $a_1, \ldots, a_p$ 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,

$$M_u(t) = M_X(tL) = e^{(tL)'\mu + \frac{1}{2}(tL)'\Sigma(tL)} = e^{T'\mu + \frac{1}{2}T'\Sigma T} = M_X(T), \ T = tL. \quad (3.2.1)$$

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 $\Sigma > O$:

$$f(X) = c\, e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}, \ -\infty < x_j < \infty, \ -\infty < \mu_j < \infty, \ \Sigma > O$$

for $j = 1, \ldots, p$. We can evaluate the normalizing constant c when $f(X)$ is a density, in which case the total integral is unity. That is,

$$1 = \int_X f(X)\mathrm{d}X = \int_X c\, e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}\mathrm{d}X.$$

Let $\Sigma^{-\frac{1}{2}}(X - \mu) = Y \Rightarrow \mathrm{d}Y = |\Sigma|^{-\frac{1}{2}}\mathrm{d}(X - \mu) = |\Sigma|^{-\frac{1}{2}}\mathrm{d}X$ since μ is a constant. The Jacobian of the transformation may be obtained from Theorem 1.6.1. Now,

$$1 = c|\Sigma|^{\frac{1}{2}} \int_Y e^{-\frac{1}{2}Y'Y}\mathrm{d}Y.$$

But $Y'Y = y_1^2 + \cdots + y_p^2$ where $y_1, \ldots, y_p$ are the real elements in Y and $\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_j^2}\mathrm{d}y_j = \sqrt{2\pi}$. Hence $\int_Y e^{-\frac{1}{2}Y'Y}\mathrm{d}Y = (\sqrt{2\pi})^p$. 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 $\Sigma > O$, $-\infty < x_j < \infty$, $-\infty < \mu_j < \infty$, $j = 1, \ldots, p$. 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?

$$E[X] = E[X - \mu] + E[\mu] = \mu + \int_X (X - \mu) f(X) \mathrm{d}X$$

$$= \mu + \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \int_X (X - \mu) \mathrm{e}^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} \mathrm{d}X$$

$$= \mu + \frac{\Sigma^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \int_Y Y \mathrm{e}^{-\frac{1}{2}Y'Y} \mathrm{d}Y, \ Y = \Sigma^{-\frac{1}{2}}(X - \mu).$$

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 $Y' = (y_1, \ldots, y_p)$ is

$$E[y_j] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_j \mathrm{e}^{-\frac{1}{2}y_j^2} \mathrm{d}y_j \Big\{ \prod_{i \neq j=1}^{p} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \Big\}.$$

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,

$$\mathrm{Cov}(X) = E[X - E(X)][X - E(X)]' = E[(X - \mu)(X - \mu)']$$

$$= \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \int_X (X - \mu)(X - \mu)' \mathrm{e}^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} \mathrm{d}X$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}} \Sigma^{\frac{1}{2}} \Big[\int_Y YY' \mathrm{e}^{-\frac{1}{2}Y'Y} \mathrm{d}Y \Big] \Sigma^{\frac{1}{2}}, \ Y = \Sigma^{-\frac{1}{2}}(X - \mu).$$

But

$$YY' = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} [y_1, \ldots, y_p] = \begin{bmatrix} y_1^2 & y_1 y_2 & \cdots & y_1 y_p \\ y_2 y_1 & y_2^2 & \cdots & y_2 y_p \\ \vdots & \vdots & \ddots & \vdots \\ y_p y_1 & y_p y_2 & \cdots & y_p^2 \end{bmatrix}.$$

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 $y_2, \ldots, y_p$ will give the following:

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}y_j^2} \mathrm{d}y_j = \sqrt{2\pi}, \ j = 2, \ldots, p \ \Rightarrow (2\pi)^{\frac{p-1}{2}}$$

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

$$\mu = E[X] \text{ and } \Sigma = \text{Cov}(X). \tag{3.2.3}$$

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

$$\rho = \frac{\text{Cov}(x_1, x_2)}{\sqrt{\text{Var}(x_1)\text{Var}(x_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}, \ \sigma_1 \neq 0, \ \sigma_2 \neq 0,$$

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 $u = a_1 x + b_1$ and $v = a_2 y + b_2$ 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

$$|\Sigma| = \begin{vmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2 \sigma_2^2 (1 - \rho^2), \ -1 < \rho < 1, \ \sigma_1 > 0, \ \sigma_2 > 0.$$

The inverse is as follows, taking the inverse as the transpose of the matrix of cofactors divided by the determinant:

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}. \tag{3.2.4}$$

Then,

$$(X - \mu)' \Sigma^{-1} (X - \mu) = \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \equiv Q. \tag{3.2.5}$$

Hence, the real bivariate normal density is

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{(1 - \rho^2)}} e^{-\frac{Q}{2(1 - \rho^2)}} \tag{3.2.6}$$

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

$$\text{Var}(u) = \text{Var}(y_1) + \text{Var}(y_2) - 2\text{Cov}(y_1, y_2) = 1 + 1 - 2\rho = 2(1 - \rho). \tag{3.2.7}$$

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:

$$Q = (X-\mu)'\Sigma^{-1}(X-\mu) = \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right) = c$$

(3.2.8)

where c is positive describes an ellipse in two-dimensional Euclidean space, and for a general p,

$$(X - \mu)'\Sigma^{-1}(X - \mu) = c > 0, \ \Sigma > O, \tag{3.2.9}$$

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

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \ \mu = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 3 & 0 & -1 \\ 0 & 3 & 1 \\ -1 & 1 & 2 \end{bmatrix}.$$

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:

$$\text{Cof}(\Sigma) = \begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix}, \ \Sigma^{-1} = \frac{1}{12}\begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix}. \tag{i}$$

Thus the exponent of the trivariate real Gaussian density is $-\frac{1}{2}Q$ where

$$Q = \frac{1}{12}[x_1 - 1, x_2 + 1, x_3 + 2]\begin{bmatrix} 5 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 9 \end{bmatrix}\begin{bmatrix} x_1 - 1 \\ x_2 + 1 \\ x_3 + 2 \end{bmatrix}$$

$$= \frac{1}{12}\{5(x_1 - 1)^2 + 5(x_2 + 1)^2 + 9(x_3 + 2)^2 - 2(x_1 - 1)(x_2 + 1)$$

$$+ 6(x_1 - 1)(x_3 + 2) - 6(x_2 + 1)(x_3 + 2)\}. \tag{ii}$$

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 $T' = (t_1, \ldots, t_p)$, we have

$$M_X(T) = E[e^{T'X}] = \int_X e^{T'X} f(X)\mathrm{d}X = e^{T'\mu}E[e^{T'(X-\mu)}]$$

$$= \frac{e^{T'\mu}}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \int_X e^{T'(X-\mu)-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}\mathrm{d}X.$$

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 $Y = \Sigma^{-\frac{1}{2}}(X - \mu) \Rightarrow \mathrm{d}Y = |\Sigma|^{-\frac{1}{2}}\mathrm{d}X$. The exponent can be simplified as follows:

$$T'(X - \mu) - \frac{1}{2}(X - \mu)'\Sigma^{-1}(X - \mu) = -\frac{1}{2}\{-2T'\Sigma^{\frac{1}{2}}Y + Y'Y\}$$

$$= -\frac{1}{2}\{(Y - \Sigma^{\frac{1}{2}}T)'(Y - \Sigma^{\frac{1}{2}}T) - T'\Sigma T\}.$$

Hence

$$M_X(T) = e^{T'\mu+\frac{1}{2}T'\Sigma T} \frac{1}{(2\pi)^{\frac{p}{2}}} \int_Y e^{-\frac{1}{2}(Y-\Sigma^{\frac{1}{2}}T)'(Y-\Sigma^{\frac{1}{2}}T)}\mathrm{d}Y.$$

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

$$\frac{\partial}{\partial T} M_X(T)|_{T=O} = \frac{\partial}{\partial T} e^{T'\mu + \frac{1}{2}T'\Sigma T}|_{T=O}$$
$$= [e^{T'\mu + \frac{1}{2}T'\Sigma T}[\mu + \Sigma T]|_{T=O}] \Rightarrow \mu = E[X]. \qquad (i)$$

Then,

$$\frac{\partial}{\partial T'} M_X(T) = e^{T'\mu + \frac{1}{2}T'\Sigma T}[\mu' + T'\Sigma]. \qquad (ii)$$

Remember to write the scalar quantity, $M_X(T)$, on the left for scalar multiplication of matrices. Now,

$$\frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T) = \frac{\partial}{\partial T} e^{T'\mu + \frac{1}{2}T'\Sigma T}[\mu' + T'\Sigma]$$
$$= M_X(T)[\mu + \Sigma T][\mu' + T'\Sigma] + M_X(T)[\Sigma].$$

Hence,

$$E[XX'] = \left[\frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T)|_{T=O}\right] = \Sigma + \mu\mu'. \qquad (iii)$$

But

$$\text{Cov}(X) = E[XX'] - E[X]E[X'] = (\Sigma + \mu\mu') - \mu\mu' = \Sigma. \qquad (iv)$$

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:

$$E[\cdots X'XX'] = \cdots \frac{\partial}{\partial T'} \frac{\partial}{\partial T} \frac{\partial}{\partial T'} M_X(T)|_{T=O} . \qquad (v)$$

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 $u = L'X$, $X \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $L' = (a_1, \ldots, a_p)$ where $a_1, \ldots, a_p$ 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

$$M_u(t) = E[e^{tu}] = E[e^{tL'X}] = E[e^{(tL')X}].$$

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,

$$M_u(t) = e^{t(L'\mu) + \frac{t^2}{2}L'\Sigma L} \Rightarrow u \sim N_1(L'\mu, L'\Sigma L). \qquad (3.2.11)$$

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*

$$U = AX \sim N_q(A\mu, A\Sigma A'), \ A\Sigma A > O. \qquad (3.2.12)$$

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:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mu = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

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

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

$$E[Y] = AE[X] = A \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \qquad (i)$$

$$\text{Cov}(Y) = A \, \text{Cov}(X)A' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 3 & 11 \end{bmatrix}. \quad (ii)$$

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

$$Q = \frac{1}{68} \left\{ [y_1 - 1, y_2 - 1] \begin{bmatrix} 11 & -3 \\ -3 & 7 \end{bmatrix} \begin{bmatrix} y_1 - 1 \\ y_2 - 1 \end{bmatrix} \right\}$$

$$= \frac{1}{68} \{ 11(y_1 - 1)^2 + 7(y_2 - 1)^2 - 6(y_1 - 1)(y_2 - 1) \}. \qquad (iii)$$

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

$$f(Y) = \frac{1}{4\sqrt{17}\pi}e^{-\frac{1}{2}Q} \qquad (iv)$$

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

$$\mathrm{Var}(y_1) = A_1 \mathrm{Cov}(X)A_1' = [1, 1, 1] \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 7.$$

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 $\mathrm{Var}(y_1)$. 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:

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad (v)$$

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

$$T'\mu + \frac{1}{2}T'\Sigma T = [T_1', O'] \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix} + \frac{1}{2}[T_1', O'] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} T_1 \\ O \end{bmatrix}$$

$$= T_1'\mu_{(1)} + \frac{1}{2}T_1'\Sigma_{11}T_1,$$

which is the structure of the mgf of a real Gaussian distribution with mean value vector $E[X_1] = \mu_{(1)}$ and covariance matrix $\mathrm{Cov}(X_1) = \Sigma_{11}$. 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 $\tilde{T}' = (\tilde{t}_1, \ldots, \tilde{t}_p)$ 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,

$$E[e^{\Re(\tilde{T}^* \tilde{Y})}] = \frac{1}{\pi^p} \int_{\tilde{Y}} e^{\Re(\tilde{T}^* \Sigma^{\frac{1}{2}} \tilde{Y}) - \tilde{Y}^* \tilde{Y}} d\tilde{Y}. \tag{i}$$

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*

$$2 \Re(\tilde{U}^* \tilde{V}) = U^* \tilde{V} + \tilde{V}^* \tilde{U} = 2 \Re(\tilde{V}^* \tilde{U}).$$

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

$$M_{\tilde{X}}(\tilde{T}) = e^{\Re(\tilde{T}^* \tilde{\mu}) + \frac{1}{4} \tilde{T}^* \Sigma \tilde{T}}. \tag{3.2a.1}$$

Example 3.2a.1. Let $\tilde{X}$, $E[\tilde{X}] = \tilde{\mu}$, $\mathrm{Cov}(\tilde{X}) = \Sigma$ 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:

$$\Re(\tilde{T}^* \tilde{\mu}) = \Re \left\{ [t_{11} - it_{12}, t_{21} - it_{22}] \begin{bmatrix} 1 - i \\ 2 - 3i \end{bmatrix} \right\}$$
$$= \Re \{ (t_{11} - it_{12})(1 - i) + (t_{21} - it_{22})(2 - 3i) \}$$
$$= t_{11} - t_{12} + 2t_{21} - 3t_{22}.$$

The second term in the exponent is the following:

$$\frac{1}{4}\tilde{T}^*\Sigma\tilde{T} = \frac{1}{4}\left\{[\tilde{t}_1^*, \tilde{t}_2^*]\begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}\begin{bmatrix} \tilde{t}_1 \\ \tilde{t}_2 \end{bmatrix}\right\}$$

$$= \frac{1}{4}\{3\tilde{t}_1^*\tilde{t}_1 + 2\tilde{t}_2^*\tilde{t}_2 + (1+i)\tilde{t}_1^*\tilde{t}_2 + (1-i)\tilde{t}_2^*\tilde{t}_1.\}.$$

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,

$$\phi = [t_{11} - t_{12} + 2t_{21} - 3t_{22}] + \frac{1}{4}\{3(t_{11}^2 + t_{12}^2) + 2(t_{21}^2 + t_{22}^2)$$

$$+ 2(t_{11}t_{21} + t_{12}t_{22} + t_{12}t_{21} - t_{11}t_{22})\}. \tag{i}$$

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 (3.2*a*.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

$$\frac{\partial}{\partial \tilde{X}} = \frac{\partial}{\partial X_1} + i\frac{\partial}{\partial X_2}. \tag{ii}$$

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, $\mu_{(1)}, \mu_{(2)}$, Σ_1, Σ_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

$$T_j'\Sigma T_j = T_j'(\Sigma_1 + i\Sigma_2)T_j = T_j'\Sigma_1 T_j + 0 \tag{iii}$$

for $j = 1, 2$ since Σ_2 is skew symmetric. The exponent in the mgf in (3.2*a*.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,

$$u = \Re(\tilde{T}^*\tilde{\mu}) + \frac{1}{4}\tilde{T}^*\Sigma\tilde{T} = \Re(T_1' - iT_2')(\mu_{(1)} + i\mu_{(2)}) + \frac{1}{4}(T_1' - iT_2')\Sigma(T_1 + iT_2)$$

$$= T_1'\mu_{(1)} + T_2'\mu_{(2)} + \frac{1}{4}[T_1'\Sigma T_1 + T_2'\Sigma T_2] + \frac{1}{4}u_1, \ u_1 = i(T_1'\Sigma T_2 - T_2'\Sigma T_1)$$

$$= T_1'\mu_{(1)} + T_2'\mu_{(2)} + \frac{1}{4}[T_1'\Sigma_1 T_1 + T_2'\Sigma_1 T_2] + \frac{1}{4}u_1. \tag{iv}$$

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*

$$u_1 = i(T_1'\Sigma T_2 - T_2'\Sigma T_1) = -2T_1'\Sigma_2 T_2 = 2T_2'\Sigma_2 T_1$$

$$\Rightarrow \frac{\partial}{\partial T_1}u_1 = -2\Sigma_2 T_2 \ \text{and} \ \frac{\partial}{\partial T_2}u_1 = 2\Sigma_2 T_1. \tag{v}$$

***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,

$$u_1 = i(T_1'\Sigma T_2 - T_2'\Sigma T_1) = i[T_1'(\Sigma_1 + i\Sigma_2)T_2 - T_2'(\Sigma_1 + i\Sigma_2)T_1],$$
$$= iT_1'\Sigma_1 T_2 - T_1'\Sigma_2 T_2 - iT_2'\Sigma_1 T_1 + T_2'\Sigma_2 T_1 = -T_1'\Sigma_2 T_2 + T_2'\Sigma_2 T_1$$
$$= -2T_1'\Sigma_2 T_2 = 2T_2'\Sigma_2 T_1. \tag{vi}$$

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

$$\frac{\partial}{\partial T_1}(T_1'\Sigma_1 T_1) = 2\Sigma_1 T_1, \ \frac{\partial}{\partial T_1}(-2T_1'\Sigma_2 T_2) = -2\Sigma_2 T_2, \ \frac{\partial}{\partial T_1}\Re(\tilde{T}^*\tilde{\mu}) = \mu_{(1)},$$

$$\frac{\partial}{\partial T_2}(T_2'\Sigma_1 T_2) = 2\Sigma_1 T_2, \ \frac{\partial}{\partial T_2}(2T_2'\Sigma_2 T_1) = 2\Sigma_2 T_1, \ \frac{\partial}{\partial T_2}\Re(\tilde{T}^*\tilde{\mu}) = \mu_{(2)}.$$

Thus, given *(ii)–(vi)*, the operator applied to the exponent of the mgf gives the following result:

$$\left(\frac{\partial}{\partial T_1} + i\frac{\partial}{\partial T_2}\right)u = \mu_{(1)} + i\mu_{(2)} + \frac{1}{4}[2\Sigma_1 T_1 - 2\Sigma_2 T_2 + 2\Sigma_1 i T_2 + 2\Sigma_2 i T_1]$$

$$= \tilde{\mu} + \frac{1}{4}[2(\Sigma_1 + i\Sigma_2)T_1 + 2(\Sigma_1 + i\Sigma_2)i T_2 = \tilde{\mu} + \frac{1}{4}[2\Sigma\tilde{T}] = \tilde{\mu} + \frac{1}{2}\Sigma\tilde{T},$$

so that

$$\frac{\partial}{\partial\tilde{T}}M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \left(\frac{\partial}{\partial T_1} + i\frac{\partial}{\partial T_2}\right)M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O}$$

$$= [M_{\tilde{X}}(\tilde{T})[\tilde{\mu} + \frac{1}{2}\Sigma\tilde{T}]|_{T_1=O,T_2=O} = \tilde{\mu}, \qquad\qquad (vii)$$

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

$$\frac{\partial}{\partial\tilde{T}}M_{\tilde{X}}(\tilde{T}) = M_{\tilde{X}}(\tilde{T})[\tilde{\mu} + \frac{1}{2}\Sigma\tilde{T}],$$

$$\frac{\partial}{\partial\tilde{T}^*}M_{\tilde{X}}(\tilde{T}) = [\tilde{\mu}^* + \frac{1}{2}\tilde{T}^*\Sigma]M_{\tilde{X}}(\tilde{T}).$$

Now, observe that

$$\tilde{T}^*\Sigma = (T_1' - iT_2')\Sigma = T_1'\Sigma - iT_2'\Sigma \Rightarrow$$

$$\frac{\partial}{\partial T_1}(\tilde{T}^*\Sigma) = \Sigma, \ \frac{\partial}{\partial T_2}(\tilde{T}^*\Sigma) = -i\Sigma,$$

$$\left(\frac{\partial}{\partial T_1} + i\frac{\partial}{\partial T_2}\right)(\tilde{T}^*\Sigma) = \Sigma - i(i)\Sigma = 2\Sigma,$$

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,

$$\frac{\partial}{\partial\tilde{T}}M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \tilde{\mu} \ \text{ and } \ \frac{\partial}{\partial\tilde{T}}\frac{\partial}{\partial\tilde{T}^*}M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} = \tilde{\Sigma} + \tilde{\mu}\tilde{\mu}^*,$$

and then $\text{Cov}(\tilde{X}) = \tilde{\Sigma}$. In general, for higher order moments, one would have

$$E[\cdots \tilde{X}^*\tilde{X}\tilde{X}^*] = \cdots \frac{\partial}{\partial \tilde{T}^*}\frac{\partial}{\partial \tilde{T}}\frac{\partial}{\partial \tilde{T}^*}M_{\tilde{X}}(\tilde{T})|_{\tilde{T}=O} \, .$$

3.2a.2. Linear functions

Let $\tilde{w} = L^*\tilde{X}$ where $L^* = (a_1, \ldots, a_p)$ and $a_1, \ldots, a_p$ 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,

$$M_{\tilde{w}}(\tilde{t}) = E[e^{\Re(\tilde{t}\tilde{w})}] = E[e^{(\Re(\tilde{t}L^*\tilde{X}))}]. \tag{3.2a.2}$$

Note that this expected value is available from (3.2*a*.1) by replacing $\tilde{T}^*$ by $\tilde{t}L^*$. Hence

$$M_{\tilde{w}}(\tilde{t}) = e^{\Re(\tilde{t}(L^*\tilde{\mu}))+\frac{1}{4}\tilde{t}\tilde{t}^*(L^*\Sigma L)}. \tag{3.2a.3}$$

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,

$$M_{\tilde{Y}}(\tilde{U}) = E[e^{\Re(\tilde{U}^*\tilde{Y})}] = E[e^{\Re(\tilde{U}^*A\tilde{X})}] = E[e^{\Re[(\tilde{U}^*A)\tilde{X}]}]. \tag{3.2a.4}$$

On comparing this expected value with (3.2*a*.1), we can write down the mgf of $\tilde{Y}$ as the following:

$$M_{\tilde{Y}}(\tilde{U}) = e^{\Re(\tilde{U}^*A\tilde{\mu})+\frac{1}{4}(\tilde{U}^*A)\Sigma(A^*\tilde{U})} = e^{\Re(\tilde{U}^*(A\tilde{\mu}))+\frac{1}{4}\tilde{U}^*(A\Sigma A^*)\tilde{U}}, \tag{3.2a.5}$$

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,*

$$\tilde{Y} \sim \tilde{N}_q(A\tilde{\mu}, A\Sigma A^*), \ A\Sigma A^* > O. \tag{3.2a.6}$$

Let us consider the following partitioning of $\tilde{T}$, $\tilde{X}$, Σ where $\tilde{T}$ is $p \times 1$, $\tilde{T}_1$ is $r \times 1$, $r \le p$, $\tilde{X}_1$ is $r \times 1$, Σ_{11} is $r \times r$, $\tilde{\mu}_{(1)}$ is $r \times 1$:

$$\tilde{T} = \begin{bmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{bmatrix}, \quad \tilde{T}_1 = \begin{bmatrix} \tilde{t}_1 \\ \vdots \\ \tilde{t}_r \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{bmatrix}.$$

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

$$M_{\tilde{X}_1}(\tilde{T}_1) = e^{\Re(\tilde{T}_1^* \tilde{\mu}_{(1)}) + \frac{1}{4} \tilde{T}_1^* \Sigma_{11} \tilde{T}_1}. \tag{3.2a.7}$$

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 $r = 1, 2, \ldots, p$ 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

$$\tilde{X} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix}, \quad \tilde{\mu} = \begin{pmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \tilde{T} = \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix}.$$

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

$$\tilde{T}^* \Sigma \tilde{T} = (\tilde{T}_1^*, \tilde{T}_2^*) \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix} = \tilde{T}_1^* \Sigma_{11} \tilde{T}_1 + \tilde{T}_2^* \Sigma_{22} \tilde{T}_2. \tag{i}$$

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:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \ \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \ \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

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

$$\begin{aligned} (X - \mu)' \Sigma^{-1} (X - \mu) &= [(X_1 - \mu_{(1)})', (X_2 - \mu_{(2)})'] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} X_1 - \mu_{(1)} \\ X_2 - \mu_{(2)} \end{bmatrix} \\ &= (X_1 - \mu_{(1)})' \Sigma^{11} (X_1 - \mu_{(1)}) + (X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) \\ &\quad + (X_1 - \mu_{(1)})' \Sigma^{12} (X_2 - \mu_{(2)}) + (X_2 - \mu_{(2)})' \Sigma^{21} (X_1 - \mu_{(1)}). \end{aligned}$$

$$(i)$$

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

$$(X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) + 2(X_2 - \mu_{(2)})' \Sigma^{21} (X_1 - \mu_{(1)}).$$

$$(ii)$$

If we expand a quadratic form of the type $(X_2 - \mu_{(2)} + C)' \Sigma^{22} (X_2 - \mu_{(2)} + C)$, we have

$$\begin{aligned} (X_2 - \mu_{(2)} + C)' \Sigma^{22} (X_2 - \mu_{(2)} + C) &= (X_2 - \mu_{(2)})' \Sigma^{22} (X_2 - \mu_{(2)}) \\ &\quad + (X_2 - \mu_{(2)})' \Sigma^{22} C + C' \Sigma^{22} (X_2 - \mu_{(2)}) + C' \Sigma^{22} C. \end{aligned}$$

$$(iii)$$

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:

$$f_1(X_1) = \frac{1}{|\Sigma_{11}|^{\frac{1}{2}} (2\pi)^{\frac{r}{2}}} e^{-\frac{1}{2}(X_1 - \mu_{(1)})' \Sigma_{11}^{-1}(X_1 - \mu_{(1)})}, \ \Sigma_{11} > O, \qquad (3.3.1)$$

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:

$$f_2(X_2) = \frac{1}{|\Sigma_{22}|^{\frac{1}{2}} (2\pi)^{\frac{p-r}{2}}} e^{-\frac{1}{2}(X_2 - \mu_{(2)})' \Sigma_{22}^{-1}(X_2 - \mu_{(2)})}, \ \Sigma_{22} > O, \qquad (3.3.2)$$

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,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}, \ \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{vmatrix} = \sigma_{11}\sigma_{22} - (\sigma_{12})^2.$$

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_1, 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,

$$\Sigma^{-1} = \frac{1}{|\Sigma|}[\text{cof}(\Sigma)]' = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$= \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}, \quad -1 < \rho < 1.$$

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:

$$f(x_1, x_2) = \frac{1}{(2\pi)\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2)} Q \right\} \qquad (3.3.3)$$

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:

$$g_1(X_1|X_2) = \frac{f(X)}{f_2(X_2)} = \frac{|\Sigma_{22}|^{\frac{1}{2}}}{(2\pi)^{\frac{r}{2}}|\Sigma|^{\frac{1}{2}}}$$

$$\times \exp\left\{ -\frac{1}{2}[(X - \mu)'\Sigma^{-1}(X - \mu) - (X_2 - \mu_{(2)})'\Sigma_{22}^{-1}(X_2 - \mu_{(2)})] \right\}.$$

We can simplify the exponent, excluding $-\frac{1}{2}$, as follows:

$$(X - \mu)'\Sigma^{-1}(X - \mu) - (X_2 - \mu_{(2)})'\Sigma_{22}^{-1}(X_2 - \mu_{(2)})$$

$$= (X_1 - \mu_{(1)})'\Sigma^{11}(X_1 - \mu_{(1)}) + 2(X_1 - \mu_{(1)})'\Sigma^{12}(X_2 - \mu_{(2)})$$

$$+ (X_2 - \mu_{(2)})'\Sigma^{22}(X_2 - \mu_{(2)}) - (X_2 - \mu_{(2)})'\Sigma_{22}^{-1}(X_2 - \mu_{(2)}).$$

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:

$$g_1(X_1|X_2) = \frac{1}{(2\pi)^{\frac{r}{2}}|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|^{\frac{1}{2}}}$$
$$\times \exp\{-\frac{1}{2}(X_1 - \mu_{(1)} + C)'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(X_1 - \mu_{(1)} + C)$$

(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

$$E[X_1|X_2] = \mu_{(1)} - C = \mu_{(1)} - (\Sigma^{11})^{-1}\Sigma^{12}(X_2 - \mu_{(2)})$$
$$= \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_{(2)}), \text{ which is linear in } X_2.$$
$$\text{Cov}(X_1|X_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \text{ which is free of } X_2. \tag{3.3.5}$$

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

$$g_2(X_2|X_1) = \frac{1}{(2\pi)^{\frac{p-r}{2}}|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}}}$$
$$\times \exp\left\{-\frac{1}{2}(X_2 - \mu_{(2)} + C_1)'(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}(X_2 - \mu_{(2)} + C_1)\right\}$$

(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

$$E[X_2|X_1] = \mu_{(2)} - C_1 = \mu_{(2)} - (\Sigma^{22})^{-1}\Sigma^{21}(X_1 - \mu_{(1)})$$
$$= \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_{(1)}), \text{ which linear in } X_1$$
$$\text{Cov}(X_2|X_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \text{ which is free of } X_1, \tag{3.3.7}$$

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

$$E[X_1|X_2] = \mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_{(2)}) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$$

$$= \mu_1 + \frac{\sigma_1\sigma_2\rho}{\sigma_2^2}(x_2 - \mu_2) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2) = E[x_1|x_2], \qquad (3.3.8)$$

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

$$E[x_2|x_1] = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x_1 - \mu_1), \text{ which is linear in } x_1 \qquad (3.3.9)$$

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:

$$\Sigma_{22}^{-1} = \frac{1}{5} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = 3 - [-2, 0]\left(\frac{1}{5}\right) \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \frac{3}{5}$$

$$\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} - \left(\frac{1}{3}\right) \begin{bmatrix} -2 \\ 0 \end{bmatrix} [-2, 0] = \begin{bmatrix} \frac{2}{3} & 1 \\ 1 & 3 \end{bmatrix}$$

$$[\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}]^{-1} = \frac{5}{3}$$

$$[\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}]^{-1} = \begin{bmatrix} 3 & -1 \\ -1 & \frac{2}{3} \end{bmatrix}.$$

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:

$$E[X_1|X_2] = \mu_{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_{(2)})$$

$$= -1 + \left[-\frac{6}{5}, \frac{2}{5}\right] \begin{bmatrix} x_2 - 0 \\ x_3 + 2 \end{bmatrix} = -1 - \frac{6}{5}x_2 + \frac{2}{5}(x_3 + 2) \qquad (i)$$

and

$$\text{Cov}(X_1|X_2) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \frac{3}{5}; \qquad (ii)$$

$$E[X_2|X_1] = \mu_{(2)} + \Sigma_{12} \Sigma_{22}^{-1} (X_1 - \mu_{(1)})$$

$$= \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \begin{bmatrix} -2/3 \\ 0 \end{bmatrix} (x_1 + 1) = \begin{bmatrix} -\frac{2}{3}(x_1 + 1) \\ -2 \end{bmatrix} \qquad (iii)$$

and

$$\text{Cov}(X_2|X_1) = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \begin{bmatrix} 2/3 & 1 \\ 1 & 3 \end{bmatrix}. \qquad (iv)$$

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

$$f_1(x_1) = \frac{1}{\sqrt{(2\pi)}\sqrt{3}} e^{-\frac{1}{6}(x_1+1)^2}, \quad -\infty < x_1 < \infty,$$

$$f_2(X_2) = \frac{1}{(2\pi)\sqrt{5}} e^{-\frac{1}{2}Q_1}, \quad Q_1 = \frac{1}{5}[3(x_2)^2 - 2(x_2)(x_3 + 2) + 2(x_3 + 2)^2]$$

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 $\tilde{x}_1, \ldots, \tilde{x}_p$ 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:

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}, \ \tilde{\mu} = \begin{bmatrix} \tilde{\mu}_{(1)} \\ \tilde{\mu}_{(2)} \end{bmatrix}, \ \tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}, \ \tilde{U} = \begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \end{bmatrix}$$

where $\tilde{X}_1, \tilde{\mu}_{(1)}, \tilde{U}_1$ are $r \times 1$ and Σ^{11} is $r \times r$. Consider

$$\begin{aligned} \tilde{U}^*\Sigma^{-1}\tilde{U} &= [\tilde{U}_1^*, \tilde{U}_2^*] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \tilde{U}_1 \\ \tilde{U}_2 \end{bmatrix} \\ &= \tilde{U}_1^*\Sigma^{11}\tilde{U}_1 + \tilde{U}_2^*\Sigma^{22}\tilde{U}_2 + \tilde{U}_1^*\Sigma^{12}\tilde{U}_2 + \tilde{U}_2^*\Sigma^{21}\tilde{U}_1 \quad (i) \end{aligned}$$

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

$$\begin{aligned} (\tilde{U}_2 + C)^*\Sigma^{22}(\tilde{U}_2 + C) &= \tilde{U}_2^*\Sigma^{22}\tilde{U}_2 + \tilde{U}_2^*\Sigma^{22}C \\ &\quad + C^*\Sigma^{22}\tilde{U}_2 + C^*\Sigma^{22}C, \quad (ii) \end{aligned}$$

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

$$\tilde{U}_1^*(\Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21})\tilde{U}_1 + (\tilde{U}_2 + C)^*\Sigma^{22}(\tilde{U}_2 + C), \ C = (\Sigma^{22})^{-1}\Sigma^{21}\tilde{U}_1.$$

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

$$\tilde{f}_1(\tilde{X}_1) = \frac{1}{\pi^r|\det(\Sigma_{11})|}e^{-(\tilde{X}_1-\tilde{\mu}_{(1)})^*\Sigma_{11}^{-1}(\tilde{X}_1-\tilde{\mu}_{(1)})}, \quad \Sigma_{11} > O. \tag{3.3a.1}$$

It is an r-variate complex Gaussian density. Similarly $\tilde{X}_2$ has the $(p-r)$-variate complex Gaussian density

$$\tilde{f}_2(\tilde{X}_2) = \frac{1}{\pi^{p-r}|\det(\Sigma_{22})|}e^{-(\tilde{X}_2-\tilde{\mu}_{(2)})^*\Sigma_{22}^{-1}(\tilde{X}_2-\tilde{\mu}_{(2)})}, \quad \Sigma_{22} > O. \tag{3.3a.2}$$

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 $\tilde{U} = \tilde{X} - \tilde{\mu}$, $\tilde{U}_1 = \tilde{X}_1 - \tilde{\mu}_{(1)}$, $\tilde{U}_2 = \tilde{X}_2 - \tilde{\mu}_{(2)}$:

$$\begin{aligned}
(\tilde{X} - \tilde{\mu})^*&\Sigma^{-1}(\tilde{X} - \tilde{\mu}) - (\tilde{X}_2 - \tilde{\mu}_{(2)})^*\Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \\
&= \tilde{U}_1^*\Sigma^{11}\tilde{U}_1 + \tilde{U}_2^*\Sigma^{22}\tilde{U}_2 + \tilde{U}_1^*\Sigma^{12}\tilde{U}_2 \\
&\quad + \tilde{U}_2^*\Sigma^{21}\tilde{U}_1 - \tilde{U}_2^*(\Sigma^{22} - \Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12})\tilde{U}_2 \\
&= \tilde{U}_1^*\Sigma^{11}\tilde{U}_1 + \tilde{U}_2^*\Sigma^{21}(\Sigma^{11})^{-1}\Sigma^{12}\tilde{U}_2 + 2\tilde{U}_1^*\Sigma^{12}\tilde{U}_2 \\
&= [\tilde{U}_1 + (\Sigma^{11})^{-1}\Sigma^{12}\tilde{U}_2]^*\Sigma^{11}[\tilde{U}_1 + (\Sigma^{11})^{-1}\Sigma^{12}\tilde{U}_2]. \tag{3.3a.3}
\end{aligned}$$

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

$$\tilde{g}_1(\tilde{X}_1|\tilde{X}_2) = \frac{1}{\pi^r|\det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})|}e^{-(\tilde{X}_1-\tilde{\mu}_{(1)}+C)^*\Sigma^{11}(\tilde{X}_1-\tilde{\mu}_{(1)}+C)},$$

$$C = -(\Sigma^{11})^{-1}\Sigma^{12}(\tilde{X}_2 - \tilde{\mu}_{(2)}). \qquad (3.3a.4)$$

The conditional expectation of $\tilde{X}_1$ given $\tilde{X}_2$ is then

$$\begin{aligned} E[\tilde{X}_1|\tilde{X}_2] &= \tilde{\mu}_{(1)} - (\Sigma^{11})^{-1}\Sigma^{12}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \\ &= \tilde{\mu}_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) \text{ (linear in } \tilde{X}_2) \end{aligned} \qquad (3.3a.5)$$

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

$$\text{Cov}(\tilde{X}_1|\tilde{X}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \text{ (free of } \tilde{X}_2).$$

From symmetry, the conditional density of $\tilde{X}_2$ given $\tilde{X}_1$ is given by

$$\begin{aligned} \tilde{g}_2(\tilde{X}_2|\tilde{X}_1) &= \frac{1}{\pi^{p-r}|\det(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})|} \\ &\times e^{-(\tilde{X}_2-\tilde{\mu}_{(2)}+C_1)^*\Sigma^{22}(\tilde{X}_2-\tilde{\mu}_{(2)}+C_1)}, \\ C_1 &= -(\Sigma^{22})^{-1}\Sigma^{21}(\tilde{X}_1 - \tilde{\mu}_{(1)}), \ \Sigma^{22} > O. \end{aligned} \qquad (3.3a.6)$$

Then the conditional expectation and the conditional covariance of $\tilde{X}_2$ given $\tilde{X}_1$ are the following:

$$\begin{aligned} E[\tilde{X}_2|\tilde{X}_1] &= \tilde{\mu}_{(2)} - (\Sigma^{22})^{-1}\Sigma^{21}(\tilde{X}_1 - \tilde{\mu}_{(1)}) \\ &= \tilde{\mu}_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\tilde{X}_1 - \tilde{\mu}_{(1)}) \text{ (linear in } \tilde{X}_1) \end{aligned} \qquad (3.3a.7)$$

$$\text{Cov}(\tilde{X}_2|\tilde{X}_1) = (\Sigma^{22})^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \text{ (free of } \tilde{X}_1),$$

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

$$\Sigma_{22}^{-1} = \frac{1}{3}, \ \Sigma_{11}^{-1} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}^{-1} = \frac{1}{4}\begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}; \tag{i}$$

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} - \begin{bmatrix} 0 \\ i \end{bmatrix}\left(\frac{1}{3}\right)[0, -i]$$

$$= \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1+i \\ 1-i & \frac{5}{3} \end{bmatrix} \Rightarrow \tag{ii}$$

$$[\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]^{-1} = \begin{bmatrix} 3+1+i & \\ 1-i & \frac{5}{3} \end{bmatrix}^{-1} = \frac{1}{3}\begin{bmatrix} \frac{5}{3} & -(1+i) \\ -(1-i) & 3 \end{bmatrix}; \tag{iii}$$

$$\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 3 - [0, -i]\left(\frac{1}{4}\begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}\right)\begin{bmatrix} 0 \\ i \end{bmatrix}$$

$$= 3 - \frac{3}{4} = \frac{9}{4} \Rightarrow \tag{iv}$$

$$[\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]^{-1} = \frac{4}{9}. \tag{v}$$

As well,

$$\Sigma_{12}\Sigma_{22}^{-1} = \frac{1}{3}\begin{bmatrix} 0 \\ i \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{i}{3} \end{bmatrix}; \tag{vi}$$

$$\Sigma_{21}\Sigma_{11}^{-1} = [0, -i]\left(\frac{1}{4}\begin{bmatrix} 2 & -(1+i) \\ -(1-i) & 3 \end{bmatrix}\right) = \frac{1}{4}[1+i, -3i]. \tag{vii}$$

With these computations, all the questions can be answered. We have

$$\tilde{X}_1 \sim \tilde{N}_2(\tilde{\mu}_{(1)}, \Sigma_{11}), \ \tilde{\mu}_{(1)} = \begin{bmatrix} 1+i \\ 2-i \end{bmatrix}, \ \Sigma_{11} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix}$$

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

$$\tilde{f}_2(\tilde{x}_3) = \frac{1}{(\pi)(3)}e^{-\frac{1}{3}(\tilde{x}_3-3i)^*(\tilde{x}_3-3i)};$$

$$\tilde{f}_1(\tilde{X}_1) = \frac{1}{(\pi^2)(4)}e^{-Q_1},$$

$$Q_1 = \frac{1}{4}[2(\tilde{x}_1 - (1+i))^*(\tilde{x}_1 - (1+i))$$
$$- (1+i)(\tilde{x}_1 - (1+i))^*(\tilde{x}_2 - (2-i)) - (1-i)(\tilde{x}_2$$
$$- (2-i))^*(\tilde{x}_1 - (1+i)) + 3(\tilde{x}_2 - (2-i))^*(\tilde{x}_2 - (2-i)).$$

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:

$$\tilde{g}_1(\tilde{X}_1|\tilde{X}_2) = \frac{1}{(\pi^2)(3)}e^{-Q_2},$$

$$Q_2 = \frac{1}{3}\left[\frac{5}{3}(\tilde{x}_1 - (1+i))^*(\tilde{x}_1 - (1+i))\right.$$

$$- (1+i)(\tilde{x}_1 - (1+i))^*(\tilde{x}_2 - (2-i) - \frac{i}{3}(\tilde{x}_3 - 3i))$$

$$- (1-i)(\tilde{x}_2 - (2-i) - \frac{i}{3}(\tilde{x}_3 - 3i))^*(\tilde{x}_1 - (1+i))$$

$$\left. + 3(\tilde{x}_2 - (2-i) - \frac{i}{3}(\tilde{x}_3 - 3i))^*(\tilde{x}_2 - (2-i) - \frac{i}{3}(\tilde{x}_3 - 3i))\right];$$

$$\tilde{g}_2(\tilde{X}_2|\tilde{X}_1) = \frac{1}{(\pi)(9/4)}e^{-Q_3},$$

$$Q_3 = \frac{4}{9}[(\tilde{x}_3 - M_3)^*(\tilde{x}_3 - M_3) \text{ where}$$

$$M_3 = 3i + \frac{1}{4}\{(1+i)[\tilde{x}_1 - (1+i)] - 3i[\tilde{x}_2 - (2-i)]\}$$

$$= 3i + \frac{1}{4}\{(1+i)\tilde{x}_1 - 3i\tilde{x}_2 + 3 + 4i\}.$$

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 $p = 2$,

$$E[\tilde{X}_1|\tilde{X}_2] = \tilde{\mu}_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\tilde{X}_2 - \tilde{\mu}_{(2)}) = \tilde{\mu}_1 + \frac{\sigma_{12}}{\sigma_2^2}(\tilde{x}_2 - \tilde{\mu}_2)$$

$$= \tilde{\mu}_1 + \frac{\sigma_1\sigma_2\rho}{\sigma_2^2}(\tilde{x}_2 - \tilde{\mu}_2) = \tilde{\mu}_1 + \frac{\sigma_1}{\sigma_2}\rho\,(\tilde{x}_2 - \tilde{\mu}_2) = E[\tilde{x}_1|\tilde{x}_2] \quad \text{(linear in } \tilde{x}_2\text{)}.$$

$$(3.3a.8)$$

Similarly,

$$E[\tilde{x}_2|\tilde{x}_1] = \tilde{\mu}_2 + \frac{\sigma_2}{\sigma_1}\rho\,(\tilde{x}_1 - \tilde{\mu}_1) \quad \text{(linear in } \tilde{x}_1\text{)}. \qquad (3.3a.9)$$

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, $E[e^{\Re(t_1\tilde{x}_1)}|\tilde{x}_2, \tilde{x}_3]$.

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 $P'AP = \text{diag}(\lambda_1, \ldots, \lambda_p)$ where $\lambda_1, \ldots, \lambda_p$ 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, $y_j^2 \overset{iid}{\sim} \chi_1^2$, $j = 1, \ldots, p$, or the y_j^2's are independently distributed chisquares, each having one degree of freedom. Note that

$$u = X'AX = Y'P'APY = \lambda_1 y_1^2 + \cdots + \lambda_p y_p^2. \tag{3.4.1}$$

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:

$$M_{y_j^2}(t) = (1 - 2t)^{-\frac{1}{2}}, \; M_{\lambda_j y_j^2}(t) = (1 - 2\lambda_j t)^{-\frac{1}{2}}, \; M_u(t) = \prod_{j=1}^{p}(1 - 2\lambda_j t)^{-\frac{1}{2}}$$

for $1 - \lambda_j t > 0$, $j = 1, \ldots, p$. Hence, we have the following identity:

$$(1 - 2t)^{-\frac{r}{2}} = \prod_{j=1}^{p}(1 - 2\lambda_j t)^{-\frac{1}{2}}, 1 - 2t > 0, 1 - 2\lambda_j t > 0, \; j = 1, \ldots, p. \quad (3.4.2)$$

Taking natural logarithm on both sides of (3.4.2), expanding and then comparing the coefficients of $2t$, $\frac{(2t)^2}{2}, \ldots$, we have

$$r = \sum_{j=1}^{p}\lambda_j = \sum_{j=1}^{p}\lambda_j^2 = \sum_{j=1}^{p}\lambda_j^3 = \cdots \quad (3.4.3)$$

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

$$u = Z'\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}}Z, \; \Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}} = (\Sigma^{\frac{1}{2}}A\Sigma^{\frac{1}{2}})',$$

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

$$\lambda_1(u_1 + b_1)^2 + \cdots + \lambda_p(u_p + b_p)^2 \equiv \lambda_1 w_1^2 + \cdots + \lambda_p w_p^2 \qquad (3.4.4)$$

where $U = (u_1, \ldots, u_p)' \sim N_p(O, I)$, the λ_j's, $j = 1, \ldots, p$, 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, $j = 1, \ldots, p$. 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 noncentality 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

$$\text{Var}(y_1) = \text{Var}(x_1) + \text{Var}(x_2) + \text{Var}(x_3) + 2[\text{Cov}(x_1, x_2) + \text{Cov}(x_1, x_3) + \text{Cov}(x_2, x_3)]$$
$$= \frac{1}{3}[2 + 2 + 3 + 0 - 2 - 2] = \frac{3}{3} = 1.$$

Hence, $y_1 = x_1 + x_2 + x_3$ has $E[u_1] = 0$ and $\text{Var}(u_1) = 1$, 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

$$AB = O \Rightarrow PABP = O \Rightarrow P'APP'BP = D_1 D_2 = O,$$
$$D_1 = \text{diag}(\lambda_1, \dots, \lambda_p), \ D_2 = \text{diag}(\nu_1, \dots, \nu_p), \tag{3.4.5}$$

where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of A and $\nu_1, \dots, \nu_p$ are the eigenvalues of B. Let $Y = P'X$, then the canonical representations of u_1 and u_2 are the following:

$$u_1 = \lambda_1 y_1^2 + \cdots + \lambda_p y_p^2 \tag{3.4.6}$$
$$u_2 = \nu_1 y_1^2 + \cdots + \nu_p y_p^2 \tag{3.4.7}$$

where y_j's are real and independently distributed. But $D_1 D_2 = O$ means that whenever a $\lambda_j \neq 0$ then the corresponding $\nu_j = 0$ 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 $u_1 = X'AX$ and $u_2 = X'BX$ where

$$
X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \Sigma = \frac{1}{3} \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},
$$

$$
B = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad \mu = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.
$$

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 $A = JJ'$ and $B = I - \frac{1}{3}JJ'$. Further, $J'J = 3$, $J'\Sigma = J'$ and hence $A\Sigma = JJ'\Sigma = JJ'$. Then $A\Sigma B = JJ'[I - \frac{1}{3}JJ'] = JJ' - JJ' = O$. 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 BX, namely $I - \frac{1}{3}JJ'$. The first component of BX 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 $Q^*AQ = \text{diag}(\lambda_1, \ldots, \lambda_p)$ where $\lambda_1, \ldots, \lambda_p$ 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

$$\tilde{u} = \tilde{X}^*A\tilde{X} = \tilde{Y}^*Q^*AQ\tilde{Y} = \lambda_1|\tilde{y}_1|^2 + \cdots + \lambda_p|\tilde{y}_p|^2 \qquad (3.4a.1)$$

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

$$\tilde{u} = |\tilde{y}_1|^2 + \cdots + |\tilde{y}_r|^2 \sim \tilde{\chi}_r^2,$$

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 $t, \frac{t^2}{2}, \ldots$ to obtain

$$r = \sum_{j=1}^{p}\lambda_j = \sum_{j=1}^{p}\lambda_j^2 = \cdots \tag{3.4a.3}$$

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

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \quad \Sigma = \frac{1}{3}\begin{bmatrix} 3 & -(1+i) & -(1-i) \\ -(1-i) & 3 & -(1+i) \\ -(1+i) & -(1-i) & 3 \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \tilde{\mu} = \begin{bmatrix} 2+i \\ -i \\ 2i \end{bmatrix}.$$

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

$$
\begin{aligned}
\mathrm{Var}(\tilde{u}_1) &= \mathrm{Var}(\tilde{x}_1) + \mathrm{Var}(\tilde{x}_2) + \mathrm{Var}(\tilde{x}_3) + [\mathrm{Cov}(\tilde{x}_1, \tilde{x}_2) + \mathrm{Cov}(\tilde{x}_2, \tilde{x}_1)] \\
&\quad + [\mathrm{Cov}(\tilde{x}_1, \tilde{x}_3) + \mathrm{Cov}(\tilde{x}_3, \tilde{x}_1)] + [\mathrm{Cov}(\tilde{x}_2, \tilde{x}_3) + \mathrm{Cov}(\tilde{x}_3, \tilde{x}_2)] \\
&= \frac{1}{3}\{3 + 3 + 3 + [-(1+i) - (1-i)] + [-(1-i) - (1+i)] \\
&\quad + [-(1+i) - (1-i)]\} = \frac{1}{3}[9 - 6] = 1.
\end{aligned}
$$

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 $\Sigma^{-1} = \frac{1}{|\Sigma|}[\mathrm{Cof}(\Sigma)]'$, which also holds for the complex case. Earlier, the determinant was found to be equal to $\frac{13}{3^3}$ and

$$
\frac{1}{|\Sigma|}[\mathrm{Cof}(\Sigma)] = \frac{3^3}{13}\begin{bmatrix} 7 & 3-i & 3+i \\ 3+i & 7 & 3-i \\ 3-i & 3+i & 7 \end{bmatrix}; \text{ then}
$$

$$
\Sigma^{-1} = \frac{1}{|\Sigma|}[\mathrm{Cof}(\Sigma)]' = \frac{3^3}{13}\begin{bmatrix} 7 & 3+i & 3-i \\ 3-i & 7 & 3+i \\ 3+i & 3-i & 7 \end{bmatrix}
$$

and

$$\lambda = \tilde{\mu}^* \Sigma^{-1} \tilde{\mu} = \frac{3^3}{13} [2 - i, i, -2i] \begin{bmatrix} 7 & 3+i & 3-i \\ 3-i & 7 & 3+i \\ 3+i & 3-i & 7 \end{bmatrix} \begin{bmatrix} 2+i \\ -i \\ 2i \end{bmatrix}$$

$$= \frac{(76)(3^3)}{13} = \frac{2052}{13} \approx 157.85.$$

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

$$AB = O = O^* = (AB)^* = B^* A^* = BA. \tag{3.4a.4}$$

This means that there exists a unitary matrix Q, $QQ^* = I$, $Q^*Q = I$, that will diagonalize both A and B. That is, $Q^* A Q = \text{diag}(\lambda_1, \ldots, \lambda_p) = D_1$, $Q^* B Q = \text{diag}(\nu_1, \ldots, \nu_p) = D_2$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of A and $\nu_1, \ldots, \nu_p$ are the eigenvalues of B. But $AB = O$ implies that $D_1 D_2 = O$. As well,

$$\tilde{u}_1 = \tilde{X}^* A \tilde{X} = \tilde{Y}^* Q^* A Q \tilde{Y} = \lambda_1 |\tilde{y}_1|^2 + \cdots + \lambda_p |\tilde{y}_p|^2, \tag{3.4a.5}$$

$$\tilde{u}_2 = \tilde{X}^* B \tilde{X} = \tilde{Y}^* Q^* B Q \tilde{Y} = \nu_1 |\tilde{y}_1|^2 + \cdots + \nu_p |\tilde{y}_p|^2. \tag{3.4a.6}$$

Since $D_1 D_2 = O$, whenever a $\lambda_j \neq 0$, the corresponding $\nu_j = 0$ 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

$$M_{\tilde{y}}(t) = E[e^{t\tilde{y}}] = \frac{1}{\pi^p |\det(\Sigma)|} \int_{\tilde{X}} e^{t(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})-(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})} d\tilde{X}$$

$$= \frac{1}{\pi^p |\det(\Sigma)|} \int_{\tilde{X}} e^{-(1-t)(\tilde{X}-\tilde{\mu})^* \Sigma^{-1}(\tilde{X}-\tilde{\mu})} d\tilde{X}$$

$$= (1-t)^{-p} \text{ for } 1-t > 0. \tag{3.4a.7}$$

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

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}, \ \tilde{\mu} = \begin{bmatrix} 2-i \\ 3+2i \\ 1-i \end{bmatrix}, \ \Sigma = \frac{1}{3} \begin{bmatrix} 3 & -(1+i) & -(1-i) \\ -(1-i) & 3 & -(1+i) \\ -(1+i) & -(1-i) & 3 \end{bmatrix},$$

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

(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 $A = J J', J' = [1, 1, 1], J'J = 3, J'\Sigma = \frac{1}{3}J', J'B = O$. Hence $A\Sigma = J J'\Sigma = J(J'\Sigma) = \frac{1}{3}J J' \Rightarrow A\Sigma B = \frac{1}{3}J J'B = \frac{1}{3}J(J'B) = O$. 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 $(\tilde{U}, \tilde{V}^*)$, $(\tilde{U}, \tilde{V}\tilde{V}^*)$, $(\tilde{U}, \tilde{V}^*\tilde{V})$, ..., $(\tilde{U}\tilde{U}^*, \tilde{V}\tilde{V}^*)$, 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) $\lambda_1 z_1^2 + \cdots + \lambda_4 z_4^2$ 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 $X_1, \ldots, X_n$ be iid as $N_p(\mu, \Sigma), \; \Sigma > O$. Then, the collection $X_1, \ldots, X_n$ is called a simple random sample of size n from this $N_p(\mu, \Sigma), \; \Sigma > O$. Then the joint density of $X_1, \ldots, X_n$ is the following:

$$L = \prod_{j=1}^{n} f(X_j) = \prod_{j=1}^{n} \frac{e^{-\frac{1}{2}(X_j - \mu)' \Sigma^{-1}(X_j - \mu)}}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}}$$

$$= [(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}]^{-1} e^{-\frac{1}{2} \sum_{j=1}^{n} (X_j - \mu)' \Sigma^{-1}(X_j - \mu)}. \tag{3.5.1}$$

This L at an observed set of $X_1, \ldots, X_n$ 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

$$\mathbf{X} = [X_1, \ldots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{bmatrix}, \; X_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{pk} \end{bmatrix}, \; k = 1, \ldots, n. \tag{i}$$

Let the sample average be denoted by $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then $\bar{X}$ will be of the following form:

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix}, \; \bar{x}_i = \frac{1}{n} \sum_{k=1}^{n} x_{ik} = \text{ average on the } i\text{-th component of any } X_j. \tag{ii}$$

Let the bold-faced $\bar{\mathbf{X}}$ be defined as follows:

$$\bar{\mathbf{X}} = [\bar{X}, \ldots, \bar{X}] = \begin{bmatrix} \bar{x}_1 & \bar{x}_1 & \ldots & \bar{x}_1 \\ \bar{x}_2 & \bar{x}_2 & \ldots & \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \bar{x}_p & \bar{x}_p & \ldots & \bar{x}_p \end{bmatrix}.$$

Then,

$$\mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \ldots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \ldots & x_{2n} - \bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} - \bar{x}_p & x_{p2} - \bar{x}_p & \ldots & x_{pn} - \bar{x}_p \end{bmatrix},$$

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:

$$\sum_{j=1}^{n} (X_j - \mu)' \Sigma^{-1}(X_j - \mu) = \text{tr}\left[\sum_{j=1}^{n} (X_j - \mu)' \Sigma^{-1}(X_j - \mu)\right]$$

$$= \text{tr}[\Sigma^{-1} \sum_{j=1}^{n} (X_j - \mu)(X_j - \mu)']$$

$$= \text{tr}\left[\Sigma^{-1} \sum_{j=1}^{n} (X_j - \bar{X} + \bar{X} - \mu)(X_j - \bar{X} + \bar{X} - \mu)'\right]$$

$$= \text{tr}[\Sigma^{-1}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'] + n\text{tr}[\Sigma^{-1}(\bar{X} - \mu)(\bar{X} - \mu)']$$

$$= \text{tr}(\Sigma^{-1} S) + n(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu)$$

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

$$L = \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1} S) - \frac{n}{2}(\bar{X} - \mu)' \Sigma^{-1}(\bar{X} - \mu)}. \tag{3.5.2}$$

If we wish to estimate the parameters μ and Σ from a set of observation vectors corresponding to $X_1, \ldots, X_n$, 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:

$$\tilde{f}(\tilde{X}_j) = \frac{e^{-(\tilde{X}_j-\tilde{\mu})^*\tilde{\Sigma}^{-1}(\tilde{X}_j-\tilde{\mu})}}{\pi^p|\det(\Sigma)|}, \ \tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \tilde{\Sigma}), \ \tilde{\Sigma} = \tilde{\Sigma}^* > O.$$

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ 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 $\tilde{\mathbf{X}} = [\tilde{X}_1, \ldots, \tilde{X}_n]$, $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \cdots + \tilde{X}_n)$, $\bar{\tilde{\mathbf{X}}} = (\bar{\tilde{X}}, \ldots, \bar{\tilde{X}})$, 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 $\tilde{X}_k$, $k = 1, \ldots, n$, $\frac{1}{n}\tilde{s}_{ii}$ being the sample variance on the component $\tilde{x}_i$. The joint density of $\tilde{X}_1, \ldots, \tilde{X}_n$, denoted by $\tilde{L}$, is given by

$$\tilde{L} = \prod_{j=1}^{n} \frac{e^{-(\tilde{X}_j-\mu)^*\tilde{\Sigma}^{-1}(\tilde{X}_j-\mu)}}{\pi^p|\det(\tilde{\Sigma})|} = \frac{e^{-\sum_{j=1}^{n}(\tilde{X}_j-\mu)^*\tilde{\Sigma}^{-1}(\tilde{X}_j-\mu)}}{\pi^{np}|\det(\tilde{\Sigma})|^n}, \tag{3.5a.1}$$

which can be simplified to the following expression by making use of steps parallel to those utilized in the real case:

$$L = \frac{e^{-\text{tr}(\tilde{\Sigma}^{-1}\tilde{S})-n(\bar{\tilde{X}}-\mu)^*\tilde{\Sigma}^{-1}(\bar{\tilde{X}}-\mu)}}{\pi^{np}|\det(\tilde{\Sigma})|^n}. \tag{3.5a.2}$$

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 $\tilde{X}_j$, $j = 1, 2, 3, 4$:

$$\tilde{X}_1 = \begin{bmatrix} 1+i \\ 2-i \\ 1-i \end{bmatrix}, \ \tilde{X}_2 = \begin{bmatrix} -1+2i \\ 3i \\ -1+i \end{bmatrix}, \ \tilde{X}_3 = \begin{bmatrix} -2+2i \\ 3+i \\ 4+2i \end{bmatrix}, \ \tilde{X}_4 = \begin{bmatrix} -2+3i \\ 3+i \\ -4+2i \end{bmatrix}.$$

Solution 3.5a.1. The sample matrix $\tilde{X}$ and the sample average $\bar{\tilde{X}}$ are

$$\tilde{\mathbf{X}} = \begin{bmatrix} 1+i & -1+2i & -2+2i & -2+3i \\ 2-i & 3i & 3+i & 3+i \\ 1-i & -1+i & 4+2i & -4+2i \end{bmatrix}, \ \bar{\tilde{X}} = \begin{bmatrix} -1+2i \\ 2+i \\ i \end{bmatrix}.$$

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 $\tilde{S}$ is

$$\tilde{S} = [\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}][\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}]^*$$

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

$$= \begin{bmatrix} 8 & 5i & 5+i \\ -5i & 14 & 6-6i \\ 5-i & 6+6i & 40 \end{bmatrix}.$$

The maximum likelihood estimates are as follows:

$$\hat{\tilde{\mu}} = \bar{\tilde{X}} = \begin{bmatrix} -1+2i \\ 2+i \\ i \end{bmatrix}, \ \hat{\tilde{\Sigma}} = \frac{1}{4}\tilde{S}$$

where $\tilde{S}$ 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

$$\mathbf{X} = [X_1, \ldots, X_n] = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{bmatrix}, \ X_k = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \vdots \\ x_{pk} \end{bmatrix}$$

where the rows are iid variables on the components of the p-vector X_1. For example, $(x_{11}, x_{12}, \ldots, x_{1n})$ are iid variables distributed as the first component of X_1. Let

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{k=1}^n x_{1k} \\ \vdots \\ \frac{1}{n} \sum_{k=1}^n x_{pk} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}(x_{11}, \ldots, x_{1n})J \\ \vdots \\ \frac{1}{n}(x_{p1}, \ldots, x_{pn})J \end{bmatrix} = \frac{1}{n}\mathbf{X}J, \ J = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ n \times 1.$$

Consider the matrix

$$\bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X}) = \frac{1}{n}\mathbf{X}JJ' = \mathbf{X}B, B = \frac{1}{n}JJ'.$$

Then,

$$\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}A, \ A = I - B = I - \frac{1}{n}JJ'.$$

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 $E[e^{\,tr(T_1'\mathbf{X}A)+\text{tr}(T_2'\mathbf{X}B)}]$. Let us evaluate the joint mgf for $X_j \sim N_p(O, I)$,

$$E[e^{\text{tr}(T_1'\mathbf{X}A)+\text{tr}(T_2'\mathbf{X}B)}] = \int_{\mathbf{X}} \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{\text{tr}(T_1'\mathbf{X}A)+\text{tr}(T_2'\mathbf{X}B)-\frac{1}{2}\text{tr}(\mathbf{XX}')}d\mathbf{X}.$$

Let us simplify the exponent,

$$-\frac{1}{2}\{\text{tr}(\mathbf{XX}') - 2\text{tr}[\mathbf{X}(AT_1' + BT_2')]\}. \tag{i}$$

If we expand $\text{tr}[(\mathbf{X} - C)(\mathbf{X} - C)']$ for some C, we have

$$\text{tr}(\mathbf{XX}') - \text{tr}(C\mathbf{X}') - \text{tr}(\mathbf{X}C') + \text{tr}(CC')$$
$$= \text{tr}(\mathbf{XX}') - 2\text{tr}(\mathbf{X}C') + \text{tr}(CC') \tag{ii}$$

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

$$\text{tr}(CC') = \text{tr}[(T_1A' + T_2B')(AT_1' + BT_2')]$$
$$= \text{tr}(T_1A'AT_1') + \text{tr}(T_2B'BT_2') + \text{tr}(T_1A'BT_2') + \text{tr}(T_2B'AT_1'). \tag{iii}$$

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

$$M_{\mathbf{X}A}(T_1) = e^{\frac{1}{2}\text{tr}(T_1 A' A T_1')} \text{ and } M_{\mathbf{X}B}(T_2) = e^{\frac{1}{2}\text{tr}(T_2 B' B T_2')}.$$

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 $X_1, \ldots, X_n$ are iid as $X_j \sim N_p(O, I)$, let the $p \times n$ matrix $\mathbf{X} = (X_1, \ldots, X_n)$ and $\bar{X} = \frac{1}{n}\mathbf{X}J, \ J' = (1, 1, .., 1)$. 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 $\bar{X}$ 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 $\bar{X}$ 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 $\bar{X}$ 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, $j = 1, \ldots, n$, be iid as $X_j \sim N_p(\mu, \Sigma), \ \Sigma > O$. Let us consider a linear function $a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ where $a_1, \ldots, a_n$ are real scalar constants. Then the mgf's of $X_j, \ a_j X_j, \ U = \sum_{j=1}^{n} a_j X_j$ are obtained as follows:

$$M_{X_j}(T) = E[e^{T'X_j}] = e^{T'\mu + \frac{1}{2}T'\Sigma T}, \quad M_{a_j X_j}(T) = e^{T'(a_j\mu) + \frac{1}{2}a_j^2 T'\Sigma T}$$

$$M_{\sum_{j=1}^n a_j X_j}(T) = \prod_{j=1}^n M_{a_j X_j}(T) = e^{T'\mu(\sum_{j=1}^n a_j) + \frac{1}{2}(\sum_{j=1}^n a_j^2)T'\Sigma T},$$

which implies that $U = \sum_{j=1}^n a_j X_j$ 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 $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \ldots, n$, and $U = a_1 X_1 + \cdots + a_n X_n$ be a linear function of the X_j's, $j = 1, \ldots, n$, where $a_1, \ldots, a_n$ 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, $a_j = \frac{1}{n}$, $j = 1, \ldots, n$, then $\sum_j^n a_j = \sum_{j=1}^n \frac{1}{n} = 1$ and $\sum_{j=1}^n a_j^2 = \sum_{j=1}^n (\frac{1}{n})^2 = \frac{1}{n}$. However, when $a_j = \frac{1}{n}$, $j = 1, \ldots, n$, $U = \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Hence we have the following corollary.

Corollary 3.5.1. *Let the X_j's be $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \ldots, n$. Then, the sample mean $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ 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, $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, the sample average, and the $p \times n$ matrix $\bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X})$, $\mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}(I - \frac{1}{n}JJ') = \mathbf{X}A$, $J' = (1, \ldots, 1)$. Since A is idempotent of rank $n - 1$, there exists an orthonormal matrix P, $PP' = I$, $P'P = I$, such that $P'AP = \text{diag}(1, \ldots, 1, 0) \equiv D$, $A = PDP'$ 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:

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = \mathbf{X}AA'\mathbf{X} = \mathbf{X}A\mathbf{X}' = \mathbf{Z}DD'\mathbf{Z}' = \mathbf{Z}_{n-1}\mathbf{Z}_{n-1}' \quad (3.5.3)$$

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 $N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \ldots, n$, $X_j - \bar{X} = (X_j - \mu) - (\bar{X} - \mu)$ 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 $\tilde{X}_1, \ldots, \tilde{X}_n$ be iid as $\tilde{X}_j \sim \tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$ or $\tilde{\mathbf{X}} = [\tilde{X}_1, \ldots, \tilde{X}_n]$ 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 $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \cdots + \tilde{X}_n)$ and the matrix of sample means be the bold-faced $p \times n$ matrix $\bar{\tilde{\mathbf{X}}}$. Let $\tilde{S} = (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})(\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})^*$. Then $\bar{\tilde{\mathbf{X}}} = \frac{1}{n}\tilde{\mathbf{X}}JJ' = \tilde{\mathbf{X}}B$, $\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}} = \tilde{\mathbf{X}}(I - \frac{1}{n}JJ') = \tilde{\mathbf{X}}A$. Then, $A = A^2$, $A = A' = A^*$, $B = B' = B^*$, $B = B^2$, $AB = O$, $BA = O$. 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 $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ where $\tilde{X}_1, \ldots, \tilde{X}_n$ are iid as $\tilde{N}_p(\tilde{\mu}, \Sigma)$, $\Sigma > O$. Let $\tilde{\mathbf{X}}$, $\bar{\tilde{\mathbf{X}}}$, $\tilde{S}$, $\tilde{\mathbf{X}}A$, $\tilde{\mathbf{X}}B$ be as defined above. Then, $\tilde{\mathbf{X}}A$ and $\tilde{\mathbf{X}}B$ 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, $j = 1, \ldots, n$, are iid as $\tilde{N}_p(\tilde{\mu}, \tilde{\Sigma})$, $\tilde{\Sigma} > O$. Let the linear function $\tilde{U} = a_1\tilde{X}_1 + \cdots + a_n\tilde{X}_n$ where $a_1, \ldots, a_n$ 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 $\sum_{j=1}^n a_j a_j^* = |\tilde{a}_1|^2 + \cdots + |\tilde{a}_n|^2$. For example, if $a_j = \frac{1}{n}$, $j = 1, \ldots, n$, 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, $j = 1, \ldots, n$, being iid as this p-variate complex Gaussian. Let $a_1, \ldots, a_n$ be scalar constants, real or complex. Consider the linear function $\tilde{U} = a_1 \tilde{X}_1 + \cdots + a_n \tilde{X}_n$. 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 $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \cdots + \tilde{X}_n)$ 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

$$\tilde{S} = \tilde{\mathbf{Z}}_{n-1}\tilde{\mathbf{Z}}_{n-1}^* \tag{3.5a.3}$$

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 $X_1, \ldots, X_n$, which are $p \times 1$ iid Gaussian vectors constituting a simple random sample of size n, we have

$$L = \prod_{j=1}^n \frac{e^{-\frac{1}{2}(X_j-\mu)'\Sigma^{-1}(X_j-\mu)}}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} = \frac{e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)-\frac{1}{2}n(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu)}}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} \tag{3.5.4}$$

where, as previously denoted, $\mathbf{X}$ is the $p \times n$ matrix

$$\mathbf{X} = (X_1, \ldots, X_n), \ \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n), \ \bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X}),$$

$$S = (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = (s_{ij}), \ s_{ij} = \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j).$$

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 $\frac{\partial}{\partial Y} = (\frac{\partial}{\partial y_{ij}})$ 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, $|Y| = y_{i1}C_{i1} + \cdots + y_{ip}C_{ip}$ for each $i = 1, \ldots, p$, 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

$$\frac{\partial}{\partial Y}|Y| = \begin{bmatrix} C_{11} & 2C_{12} & \cdots & 2C_{1p} \\ 2C_{21} & C_{22} & \cdots & 2C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 2C_{p1} & 2C_{p2} & \cdots & C_{pp} \end{bmatrix}$$

$$= |Y|[2Y^{-1} - \text{diag}(Y^{-1})].$$

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 $\frac{\partial}{\partial Y}[\text{tr}(Y)] = A'$ 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,

$$\frac{\partial}{\partial \mu}\ln L = O \Rightarrow 0 - \frac{n}{2}\frac{\partial}{\partial \mu}(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) = O$$

$$\Rightarrow n\,\Sigma^{-1}(\bar{X} - \mu) = O \Rightarrow \bar{X} - \mu = O$$

$$\Rightarrow \mu = \bar{X},$$

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 $B = B'$,

$$\frac{\partial}{\partial B}\ln L(\hat{\mu}, B) = O \Rightarrow \frac{n}{2}[2B^{-1} - \text{diag}(B^{-1})] - \frac{1}{2}[2S - \text{diag}(S)] = O$$

$$\Rightarrow n[2\Sigma - \text{diag}(\Sigma)] = 2S - \text{diag}(S)$$

$$\Rightarrow \hat{\sigma}_{jj} = \frac{1}{n}s_{jj}, \ \hat{\sigma}_{ij} = \frac{1}{n}s_{ij}, i \neq j$$

$$\Rightarrow (\hat{\mu} = \bar{X}, \hat{\Sigma} = \frac{1}{n}S).$$

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 $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of B, observe that $\lambda_j > 0, \ j = 1, \ldots, p$, 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 $\lambda_2, \ldots, \lambda_p$ 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 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 $\tilde X_1, \ldots, \tilde X_n$ be iid as $\tilde N_p(\tilde\mu, \Sigma)$, $\Sigma > O$. and let the joint density of the $\tilde X_j$'s, $j = 1, \ldots, n$, be denoted by $\tilde L$. Then

$$\tilde L = \prod_{j=1}^n \frac{e^{-(\tilde X_j - \tilde\mu)^* \Sigma^{-1}(\tilde X_j - \tilde\mu)}}{\pi^p |\det(\Sigma)|} = \frac{e^{-\sum_{j=1}^n (\tilde X_j - \tilde\mu)^* \Sigma^{-1}(\tilde X_j - \tilde\mu)}}{\pi^{np} |\det(\Sigma)|^n}$$

$$= \frac{e^{-\operatorname{tr}(\Sigma^{-1}\tilde S) - n(\bar{\tilde X} - \tilde\mu)^* \Sigma^{-1}(\bar{\tilde X} - \tilde\mu)}}{\pi^{np} |\det(\Sigma)|^n}$$

where $|\det(\Sigma)|$ denotes the absolute value of the determinant of Σ,

$$\tilde S = (\tilde{\mathbf X} - \bar{\tilde{\mathbf X}})(\tilde{\mathbf X} - \bar{\tilde{\mathbf X}})^* = (\tilde s_{ij}), \ \tilde s_{ij} = \sum_{k=1}^n (\tilde x_{ik} - \bar{\tilde x}_i)(\tilde x_{jk} - \bar{\tilde x}_j)^*,$$

$$\tilde{\mathbf X} = [\tilde X_1, \ldots, \tilde X_n], \ \bar{\tilde X} = \frac{1}{n}(\tilde X_1 + \cdots + \tilde X_n), \ \bar{\tilde{\mathbf X}} = [\bar{\tilde X}, \ldots, \bar{\tilde X}],$$

where $\tilde{\mathbf X}$ and $\bar{\tilde{\mathbf X}}$ are $p \times n$. Hence,

$$\ln \tilde L = -np \ln \pi - n \ln |\det(\Sigma)| - \operatorname{tr}(\Sigma^{-1}\tilde S) - n(\bar{\tilde X} - \tilde\mu)^* \Sigma^{-1}(\bar{\tilde X} - \tilde\mu). \qquad (3.5a.4)$$

3.5a.4. Matrix derivatives in the complex domain

Consider $\operatorname{tr}(\tilde B \tilde S^*)$, $\tilde B = \tilde B^* > O$, $\tilde S = \tilde S^* > O$. Let $\tilde B = B_1 + i B_2$, $\tilde S = S_1 + i S_2$, $i = \sqrt{(-1)}$. 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 $\frac{\partial}{\partial \tilde B}[\operatorname{tr}(\tilde B \tilde S^*)]$? Consider

$$\tilde B \tilde S^* = (B_1 + i B_2)(S_1' - i S_2') = B_1 S_1' + B_2 S_2' + i(B_2 S_1' - B_1 S_2'),$$

$$\operatorname{tr}(\tilde B \tilde S^*) = \operatorname{tr}(B_1 S_1' + B_2 S_2') + i[\operatorname{tr}(B_2 S_1') - \operatorname{tr}(B_1 S_2')].$$

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 $\operatorname{tr}(B_2 S_1') = 0$, $\operatorname{tr}(B_1 S_2') = 0$. 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, $\mathrm{tr}(AB) = 0$.*

Proof: $\mathrm{tr}(AB) = \mathrm{tr}(AB)' = \mathrm{tr}(B'A') = -\mathrm{tr}(BA) = -\mathrm{tr}(AB)$, which implies that $\mathrm{tr}(AB) = 0$.

Thus, $\mathrm{tr}(\tilde{B}\tilde{S}^*) = \mathrm{tr}(B_1 S_1' + B_2 S_2')$. The diagonal elements of S_1 in $\mathrm{tr}(B_1 S_1')$ 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,

$$\frac{\partial}{\partial B_1}\mathrm{tr}(B_1 S_1') = 2S_1 - \mathrm{diag}(S_1).$$

In B_2 and S_2, the diagonal elements are zeros and hence

$$\frac{\partial}{\partial B_2}\mathrm{tr}(B_2 S_2') = 2S_2.$$

Therefore

$$\left(\frac{\partial}{\partial B_1} + i\frac{\partial}{\partial B_2}\right)\mathrm{tr}(B_1 S_1' + B_2 S_2') = 2(S_1 + i S_2) - \mathrm{diag}(S_1) = 2\tilde{S} - \mathrm{diag}(\tilde{S}).$$

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 $\tilde{B} = B_1 + i B_2$ and $\tilde{S} = S_1 + i S_2$ 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*

$$\frac{\partial}{\partial \tilde{B}}\mathrm{tr}(\tilde{B}\tilde{S}^*) = 2\tilde{S} - \mathrm{diag}(\tilde{S}).$$

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,*

$$\frac{\partial}{\partial \tilde{\Sigma}}\ln|\det(\tilde{\Sigma})| = 2\tilde{\Sigma}^{-1} - \mathrm{diag}(\tilde{\Sigma}^{-1}).$$

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,*

$$\frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{y}) = 0, \ \frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{y}^*) = 0, \ \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{y}) = 2\tilde{y}, \ \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{y}^*) = 2\tilde{y}^*,$$

$$\frac{\partial}{\partial \tilde{x}}(\tilde{x}\tilde{x}^*) = \frac{\partial}{\partial \tilde{x}}(\tilde{x}^*\tilde{x}) = 2\tilde{x}, \ \frac{\partial}{\partial \tilde{x}^*}(\tilde{x}^*\tilde{x}) = \frac{\partial}{\partial \tilde{x}^*}(\tilde{x}\tilde{x}^*) = 2\tilde{x}^*$$

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,

$$\det(X) = x_{11}C_{11} + x_{12}C_{12} + \cdots + x_{1p}C_{1p} \tag{1}$$
$$= x_{21}C_{21} + x_{22}C_{22} + \cdots + x_{2p}C_{2p} \tag{2}$$
$$\vdots$$
$$= x_{p1}C_{p1} + x_{p2}C_{p2} + \cdots + x_{pp}C_{pp}. \tag{p}$$

When $\tilde{X} = \tilde{X}^*$, the diagonal elements x_{jj}'s are all real. From Lemma 3.5a.2 and equation (1), we have

$$\frac{\partial}{\partial x_{11}}(x_{11}C_{11}) = C_{11}, \ \frac{\partial}{\partial x_{1j}}(x_{1j}C_{1j}) = 0, \ j = 2, \ldots, p.$$

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

$$\frac{\partial}{\partial x_{12}}(x_{12}^*C_{12}^*) = C_{12}^*, \ \frac{\partial}{\partial x_{22}}(x_{22}^*C_{22}^*) = C_{22}^*, \ \frac{\partial}{\partial x_{2j}}(x_{2j}C_{2j}) = 0, \ j = 3, \ldots, p,$$

observing that $x_{22}^* = x_{22}$ and $C_{22}^* = C_{22}$. Now, continuing the process with Eqs. (3), (4), ..., (p), we have the following result:

$$\frac{\partial}{\partial x_{ij}}[\det(\tilde{X})] = \begin{cases} C_{jj}^*, \ j = 1, \ldots, p \\ 2C_{ij}^* \text{ for all } i \neq j. \end{cases}$$

Observe that for $\tilde{\Sigma}^{-1} = \tilde{B} = \tilde{B}^*$,

$$\frac{\partial}{\partial\tilde{B}}[\ln(\det(\tilde{B}))] = \frac{1}{\det(\tilde{B})}\begin{bmatrix} B_{11}^* & 2B_{12}^* & \cdots & B_{1p}^* \\ 2B_{21}^* & B_{22}^* & \cdots & 2B_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ 2B_{p1}^* & 2B_{p2}^* & \cdots & B_{pp}^* \end{bmatrix}$$

$$= 2\tilde{B}^{-1} - \operatorname{diag}(\tilde{B}^{-1}) = 2\tilde{\Sigma} - \operatorname{diag}(\tilde{\Sigma})$$

where B_{rs} is the cofactor of $\tilde{b}_{rs}$, $\tilde{B} = (\tilde{b}_{rs})$. Therefore, at $\hat{\mu} = \bar{\tilde{X}}$, for $\tilde{\Sigma}^{-1} = \tilde{B}$, and from Theorems 3.5a.5 and 3.5a.6, we have

$$\frac{\partial}{\partial\tilde{B}}[\ln\tilde{L}] = O \Rightarrow n[\tilde{\Sigma} - \operatorname{diag}(\tilde{\Sigma})] - [\tilde{S} - \operatorname{diag}(\tilde{S})] = O$$

$$\Rightarrow \tilde{\Sigma} = \frac{1}{n}\tilde{S} \Rightarrow \hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S} \text{ for } n \geq p,$$

where a hat denotes the estimate/estimator.

Again, from Lemma 3.5a.2 we have the following:

$$\frac{\partial}{\partial\tilde{\mu}}[(\bar{\tilde{X}} - \tilde{\mu})^*\tilde{\Sigma}^{-1}(\bar{\tilde{X}} - \tilde{\mu})] = \frac{\partial}{\partial\tilde{\mu}}\{\bar{\tilde{X}}^*\tilde{\Sigma}^{-1}\bar{\tilde{X}} + \tilde{\mu}^*\tilde{\Sigma}^{-1}\tilde{\mu} - \bar{\tilde{X}}^*\tilde{\Sigma}^{-1}\tilde{\mu} + \tilde{\mu}^*\tilde{\Sigma}^{-1}\bar{\tilde{X}}\} = O$$

$$\Rightarrow O + 2\tilde{\Sigma}^{-1}\tilde{\mu} - O - 2\tilde{\Sigma}^{-1}\bar{\tilde{X}} = O$$

$$\Rightarrow \hat{\tilde{\mu}} = \bar{\tilde{X}}.$$

Thus, the MLE of $\tilde{\mu}$ and $\tilde{\Sigma}$ are respectively $\hat{\tilde{\mu}} = \bar{\tilde{X}}$ and $\hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S}$ for $n \geq p$. It is not difficult to show that the only critical point $(\hat{\tilde{\mu}}, \hat{\tilde{\Sigma}}) = (\bar{\tilde{X}}, \frac{1}{n}\tilde{S})$ corresponds to a local maximum for $\tilde{L}$. Consider $\ln\tilde{L}$ at $\hat{\tilde{\mu}} = \bar{\tilde{X}}$. Let $\lambda_1, \ldots, \lambda_p$ 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, $\bar{\tilde{X}}$ 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 $\hat{\mu} = \bar{X}$, $\hat{\Sigma} = \frac{1}{n}S$, $\hat{\tilde{\mu}} = \bar{\tilde{X}}$, $\hat{\tilde{\Sigma}} = \frac{1}{n}\tilde{S}$ 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 $X_1, \ldots, X_n$ 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 $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \cdots + \tilde{X}_n)$ and let S and $\tilde{S}$ be the sample sum of products matrices in the real and complex cases, respectively. Then $E[\bar{X}] = \mu$, $E[\bar{\tilde{X}}] = \tilde{\mu}$, $E[\hat{\Sigma}] = E[\frac{1}{n}S] = \frac{n-1}{n}\Sigma \to \Sigma$ as $n \to \infty$ and $E[\bar{\tilde{X}}] = \tilde{\mu}$, $E[\hat{\tilde{\Sigma}}] = E[\frac{1}{n}\tilde{S}] = \frac{n-1}{n}\tilde{\Sigma} \to \tilde{\Sigma}$ as $n \to \infty$.*

Proof: $E[\bar{X}] = \frac{1}{n}\{E[X_1] + \cdots + E[X_n]\} = \frac{1}{n}\{\mu + \cdots + \mu\} = \mu$. Similarly, $E[\bar{\tilde{X}}] = \tilde{\mu}$. Let $\mathbf{M} = (\mu, \mu, \ldots, \mu)$, that is, M is a $p \times n$ matrix wherein every column is the $p \times 1$ vector μ. Let $\bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X})$, that is, $\bar{\mathbf{X}}$ is a $p \times n$ matrix wherein every column is $\bar{X}$. Now, consider

$$E[(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})'] = E[\sum_{j=1}^{n}(X_j - \mu)(X_j - \mu)'] = \sum_{j=1}^{n}\{\Sigma + \cdots + \Sigma\} = n\Sigma.$$

As well,

$$\begin{aligned}(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})' &= (\mathbf{X} - \bar{\mathbf{X}} + \bar{\mathbf{X}} - \mathbf{M})(\mathbf{X} - \bar{\mathbf{X}} + \bar{\mathbf{X}} - \mathbf{M})' \\ &= (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' + (\mathbf{X} - \bar{\mathbf{X}})(\bar{\mathbf{X}} - \mathbf{M})' \\ &\quad + (\bar{\mathbf{X}} - \mathbf{M})(\mathbf{X} - \bar{\mathbf{X}})' + (\bar{\mathbf{X}} - \mathbf{M})(\bar{\mathbf{X}} - \mathbf{M})' \Rightarrow \end{aligned}$$

$$(\mathbf{X} - \mathbf{M})(\mathbf{X} - \mathbf{M})' = S + \sum_{j=1}^{n}(X_j - \bar{X})(\bar{X} - \mu)' + \sum_{j=1}^{n}(\bar{X} - \mu)(X_j - \bar{X})'$$

$$+ \sum_{j=1}^{n}(\bar{X} - \mu)(\bar{X} - \mu)'$$

$$= S + O + O + n(\bar{X} - \mu)(\bar{X} - \mu)' \Rightarrow$$

$$n\Sigma = E[S] + O + O + n\text{Cov}(\bar{X}) = E[S] + n\left[\frac{1}{n}\Sigma\right] = E[S] + \Sigma \Rightarrow$$

$$E[S] = (n-1)\Sigma \Rightarrow E[\hat{\Sigma}] = E\left[\frac{1}{n}S\right] = \frac{n-1}{n}\Sigma \to \Sigma \text{ as } n \to \infty.$$

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 $x_1, \ldots, x_n$ constitute a simple random sample of size n from a univariate population. Let $T(x_1, \ldots, x_n)$ be an observable function of the sample values $x_1, \ldots, x_n$. 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 $x_1, \ldots, x_n$ if the population is univariate or of the form $X_1, \ldots, X_n$ 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,

$$L = L(x_1, \ldots, x_n, \theta) = \prod_{j=1}^{n} f(x_j, \theta) \Rightarrow \ln L = \sum_{j=1}^{n} \ln f(x_j, \theta).$$

Since the total probability is 1, we have the following, taking for example the variable to be continuous and a scalar parameter θ:

$$\int_X L \, dX = 1 \Rightarrow \frac{\partial}{\partial\theta} \int_X L \, dX = 0, \ X' = (x_1, \ldots, x_n).$$

We are going to assume that the support of x is free of theta 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

$$E\left[\frac{\partial}{\partial \theta} \ln L \right] = 0 \Rightarrow E\left[\sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_j, \theta) \right] = 0. \tag{3.5.6}$$

Let $\hat{\theta}$ be the MLE of θ. Then

$$\frac{\partial}{\partial \theta} L|_{\theta = \hat{\theta}} = 0 \Rightarrow \frac{\partial}{\partial \theta} \ln L|_{\theta = \hat{\theta}} = 0$$

$$\Rightarrow E\left[\sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta = \hat{\theta}} \right] = 0. \tag{3.5.7}$$

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),

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta = \hat{\theta}} \to E\left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta_o) \right] \text{ as } n \to \infty \tag{3.5.8}$$

where θ_o is the true value of θ. Noting that $E[\frac{\partial}{\partial \theta} \ln f(x_j, \theta_o)] = 0$ owing to the fact that $\int_{-\infty}^{\infty} f(x) dx = 1$, we have the following results:

$$\sum_{j=1}^{n} \frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta = \hat{\theta}} = 0, \ E\left[\frac{\partial}{\partial \theta} \ln f(x_j, \theta)|_{\theta = \theta_o} \right] = 0.$$

This means that $E[\hat{\theta}] = \theta_0$ or $E[\hat{\theta}] \to \theta_o$ as $n \to \infty$, that is, $\hat{\theta}$ is asymptotically unbiased for the true value θ_o of θ. As well, $\hat{\theta} \to \theta_o$ as $n \to \infty$ 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 $Pr\{\hat{\theta} \to \theta_o\} \to 1$ as $n \to \infty$, 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 $\text{Cov}(\bar{X}) = \frac{1}{n}\Sigma \to O$ as $n \to \infty$. Therefore, $\hat{\mu} = \bar{X}$ is unbiased and consistent for μ. From Theorem 3.5.8, $E[\hat{\Sigma}] = \frac{n-1}{n}\Sigma \to \Sigma$ as $n \to \infty$ 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 $T = T(X_1, \ldots, X_n)$ be an estimator of θ, where $X_1, \ldots, X_n$ are iid as the given population. If the conditional distribution of the sample values $X_1, \ldots, X_n$, 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 $\theta_1, \ldots, \theta_k$ associated with a given population and if $T_1(X_1, \ldots, X_n), \ldots, T_r(X_1, \ldots, X_n)$ are r statistics, where r may be greater, smaller or equal to k, then if the conditional distribution of $X_1, \ldots, X_n$, given $T_1, \ldots, T_r$, is free of $\theta_1, \ldots, \theta_k$, then we say that $T_1, \ldots, T_r$ are *jointly sufficient* for $\theta_1, \ldots, \theta_k$. If there are several sets of statistics, where each set is sufficient for $\theta_1, \ldots, \theta_k$, then that set of statistics which allows for the maximal reduction of the data is called the *minimal sufficient set of statistics* for $\theta_1, \ldots, \theta_k$.

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 $X_1, \ldots, X_n$ be a simple random sample from a $N_p(\mu, \Sigma)$. Then the joint density of $X_1, \ldots, X_n$ can be written as

$$L = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}}e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)-\frac{n}{2}(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu)}, \tag{i}$$

referring to (3.5.2). Since $\bar{X}$ is a function of $X_1, \ldots, X_n$, the joint density of $X_1, \ldots, X_n$ and $\bar{X}$ is L itself. Hence, the conditional density of $X_1, \ldots, X_n$, 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

$$\frac{L}{f_1(\bar{X})} = \frac{1}{(2\pi)^{\frac{n(p-1)}{2}} n^{\frac{p}{2}} |\Sigma|^{\frac{n-1}{2}}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}S)}, \qquad (ii)$$

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 $T = T(x_1, \ldots, x_n)$ 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,

$$E[y] = E[E(y|x)] \qquad (3.5.9)$$

and

$$\text{Var}(y) = \text{Var}(E[y|x]) + E[\text{Var}(y|x)] \qquad (3.5.10)$$

whenever the expected values exist. From (3.5.9),

$$g(\theta) = E[u] = E[E(u|T)] = E[h(T)] \Rightarrow E[h(T)] = g(\theta). \qquad (3.5.11)$$

Then,

$$\text{Var}(u) = E[u - g(\theta)]^2 = \text{Var}(E[u|T]) + E[\text{Var}(E[u|T])] = \text{Var}(h(T)) + \delta, \ \delta \geq 0$$
$$\Rightarrow \text{Var}(u) \geq \text{Var}(h(T)), \tag{3.5.12}$$

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 $E[\frac{\partial}{\partial \theta} \ln L] = \int_X (\frac{\partial}{\partial \theta} \ln L) L \, \mathrm{d}X = 0$. Differentiating once again with respect to θ, we have

$$0 = \int_X \frac{\partial}{\partial \theta} \Big[\Big(\frac{\partial}{\partial \theta} \ln L \Big) L \Big] \mathrm{d}X = 0$$

$$\Rightarrow \int_X \Big\{ \Big(\frac{\partial^2}{\partial \theta^2} \ln L \Big) L + \Big(\frac{\partial}{\partial \theta} \ln L \Big)^2 \Big\} \mathrm{d}X = 0$$

$$\Rightarrow \int_X \Big(\frac{\partial}{\partial \theta} \ln L \Big)^2 L \, \mathrm{d}X = - \int_X \Big(\frac{\partial^2}{\partial \theta^2} \ln L \Big) L \, \mathrm{d}X,$$

so that

$$\text{Var}\Big(\frac{\partial}{\partial \theta} \ln L \Big) = E\Big[\frac{\partial}{\partial \theta} \ln L \Big]^2 = -E\Big[\frac{\partial^2}{\partial \theta^2} \ln L \Big]$$

$$= nE\Big[\frac{\partial}{\partial \theta} \ln f(x_j, \theta) \Big]^2 = -nE\Big[\frac{\partial^2}{\partial \theta^2} \ln f(x_j, \theta) \Big]. \tag{3.5.13}$$

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,

$$\int_X T L \, \mathrm{d}X = \theta + b(\theta) \Rightarrow$$

$$1 + b'(\theta) = \int_X T \frac{\partial}{\partial \theta} L \, \mathrm{d}X, \quad b'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} b(\theta)$$

$$\Rightarrow E[T(\frac{\partial}{\partial \theta} \ln L)] = 1 + b'(\theta)$$

$$= \text{Cov}(T, \frac{\partial}{\partial \theta} \ln L)$$

because $E[\frac{\partial}{\partial\theta}\ln L] = 0$. Hence,

$$[\text{Cov}(T, \frac{\partial}{\partial\theta}\ln L)]^2 = [1 + b'(\theta)]^2 \le \text{Var}(T)\text{Var}\left(\frac{\partial}{\partial\theta}\ln L\right) \Rightarrow$$

$$\text{Var}(T) \ge \frac{[1 + b'(\theta)]^2}{\text{Var}(\frac{\partial}{\partial\theta}\ln L)} = \frac{[1 + b'(\theta)]^2}{n\text{Var}(\frac{\partial}{\partial\theta}\ln f(x_j, \theta))}$$

$$= \frac{[1 + b'(\theta)]^2}{E[\frac{\partial}{\partial\theta}\ln L]^2} = \frac{[1 + b'(\theta)]^2}{nE[\frac{\partial}{\partial\theta}\ln f(x_j, \theta)]^2} , \qquad (3.5.14)$$

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

$$\text{Var}(T) \ge \frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)} \qquad (3.5.15)$$

where

$$I_n(\theta) = \text{Var}\left(\frac{\partial}{\partial\theta}\ln L\right) = E\left[\frac{\partial}{\partial\theta}\ln L\right]^2 = nE\left[\frac{\partial}{\partial\theta}\ln f(x_j, \theta)\right]^2$$

$$= -E\left[\frac{\partial^2}{\partial\theta^2}\ln L\right] = -nE\left[\frac{\partial^2}{\partial\theta^2}\ln f(x_j, \theta)\right] = nI_1(\theta) \qquad (3.5.16)$$

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

$$0 = \frac{\partial}{\partial\theta}\ln L(X, \theta)|_{\theta=\hat{\theta}} , \qquad (i)$$

which is the likelihood equation giving rise to the MLE. Let us expand *(i)* in a neighborhood of the true parameter value θ_o :

$$0 = \frac{\partial}{\partial\theta}\ln L(X, \theta)|_{\theta=\theta_o} + (\hat{\theta} - \theta_o)\frac{\partial^2}{\partial\theta^2}\ln L(X, \theta)|_{\theta=\theta_o}$$

$$+ \frac{(\hat{\theta} - \theta_o)^2}{2}\frac{\partial^3}{\partial\theta^3}\ln L(X, \theta)|_{\theta=\theta_1} \qquad (ii)$$

where $|\hat{\theta} - \theta_1| < |\hat{\theta} - \theta_o|$. Multiplying both sides by $\sqrt{n}$ and rearranging terms, we have the following:

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ln L(X,\theta)|_{\theta=\theta_o}}{\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln(X,\theta)|_{\theta=\theta_o} + \frac{1}{n}\frac{(\hat{\theta}-\theta_o)}{2}\frac{\partial^3}{\partial\theta^3}\ln L(X,\theta)|_{\theta=\theta_1}}. \tag{iii}$$

The second term in the denominator of *(iii)* goes to zero because $\hat{\theta} \to \theta_o$ as $n \to \infty$, and the third derivative is assumed to be bounded. Then the first term in the denominator is such that

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta)|_{\theta=\theta_o} = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_j,\theta)|_{\theta=\theta_o}$$

$$\to E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x_j,\theta)\right] = -I_1(\theta)|_{\theta_o}$$

$$= -\text{Var}\left[\frac{\partial}{\partial\theta}\ln f(x_j,\theta)\right]\Big|_{\theta=\theta_o},$$

$$I_1(\theta)|_{\theta=\theta_o} = \text{Var}\left[\frac{\partial}{\partial\theta}\ln f(x_j,\theta)\right]\Big|_{\theta=\theta_o},$$

which is the information bound $I_1(\theta_o)$. Thus,

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta)|_{\theta=\theta_o} \to -I_1(\theta_o), \tag{iv}$$

and we may write *(iii)* as follows:

$$\sqrt{I_1(\theta_o)}\sqrt{n}(\hat{\theta} - \theta_o) \approx \frac{\sqrt{n}}{\sqrt{I_1(\theta_o)}}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_j,\theta)|_{\theta=\theta_o}, \tag{v}$$

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, $f(x_j,\theta)$, $j = 1,\ldots,n$ are iid variables. Hence, by the central limit theorem which is stated in Sect. 2.6,

$$\frac{\sqrt{n}}{\sqrt{I(\theta_o)}}\frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_j,\theta) \to N_1(0,1) \text{ as } n \to \infty. \tag{3.5.17}$$

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:

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_j,\theta)|_{\theta=\theta_o} \to N_1(0,I_1(\theta_o))$$

or

$$\sqrt{I_1(\theta_o)}\sqrt{n}(\hat{\theta} - \theta_o) \to N_1(0, 1) \text{ as } n \to \infty. \tag{3.5.18}$$

Since $I_1(\theta_o)$ is free of n, this result can also be written as

$$\sqrt{n}(\hat{\theta} - \theta_o) \to N_1\left(0, \frac{1}{I_1(\theta_o)}\right). \tag{3.5.19}$$

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,

$$f(x_j, \theta) = \frac{1}{\theta}e^{-\frac{x_j}{\theta}}, \ 0 \leq x_j < \infty, \ \theta > 0,$$

$$L = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{j=1}^{n} x_j}.$$

In the exponential population, $E[x_j] = \theta$, $\text{Var}(x_j) = \theta^2$, $j = 1, \ldots, n$, the MLE of θ is $\hat{\theta} = \bar{x}$, $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ and $\text{Var}(\hat{\theta}) = \frac{\theta^2}{n} \to 0$ as $n \to \infty$. Thus, $E[\hat{\theta}] = \theta$ and $\text{Var}(\hat{\theta}) \to 0$ as $n \to \infty$. 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

$$\frac{\bar{x} - \theta_o}{\sqrt{\text{Var}(\bar{x})}} = \frac{\sqrt{n}(\hat{\theta} - \theta_o)}{\theta_o} \to N_1(0, 1) \text{ as } n \to \infty,$$

and hence the asymptotic normality is also verified. Is $\hat{\theta}$ sufficient for θ? Let us consider the statistic $u = x_1 + \cdots + x_n$, the sample sum. If u is sufficient, then $\bar{x} = \hat{\theta}$ is also sufficient. The mgf of u is given by

$$M_u(t) = \prod_{j=1}^{n}(1 - \theta t)^{-1} = (1 - \theta t)^{-n}, \ 1 - \theta t > 0 \Rightarrow u \sim \text{gamma}(\alpha = n, \beta = \theta)$$

whose density is $f_1(u) = \frac{u^{n-1}}{\theta^n \Gamma(n)} e^{-\frac{u}{\theta}}$, $u = x_1 + \cdots + x_n$. However, the joint density of $x_1, \ldots, x_n$ is $L = \frac{1}{\theta^n} e^{-\frac{1}{\theta}(x_1 + \cdots + x_n)}$. Accordingly, the conditional density of $x_1, \ldots, x_n$ 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 $X_1, \ldots, X_n$ be iid with common mean value vector μ and the common covariance matrix $\Sigma > O$, that is, $E(X_j) = \mu$ and $\text{Cov}(X_j) = \Sigma > O$, $j = 1, \ldots, n$. 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 $\text{Cov}(Y_j) = I$, $j = 1, \ldots, n$, and letting $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, $\bar{Y} = \Sigma^{-\frac{1}{2}} \bar{X}$, $E(\bar{X}) = \mu$ and $E(\bar{Y}) = \Sigma^{-\frac{1}{2}} \mu$. If we let

$$U = \sqrt{n} \, \Sigma^{-\frac{1}{2}} (\bar{X} - \mu), \tag{3.5.20}$$

the following result holds:

Theorem 3.5.10. *Let the $p \times 1$ vector U be as defined in (3.5.20). Then, as $n \to \infty$, $U \to N_p(O, I)$.*

Proof: Let $L' = (a_1, \ldots, a_p)$ be an arbitrary constant vector such that $L'L = 1$. Then, $L'X_j$, $j = 1, \ldots, n$, 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 $\text{Var}(u_j) = L'\Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} L = L'L = 1$, $j = 1, \ldots, n$. Note that $\bar{u} = \frac{1}{n}(u_1 + \cdots + u_n) = L'\bar{Y} = L'\Sigma^{-\frac{1}{2}} \bar{X}$ 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 $\sqrt{n} L'\Sigma^{-\frac{1}{2}} (\bar{X} - \mu) \to N_1(0, 1)$ as $n \to \infty$. 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,

$$U = \sqrt{n} \Sigma^{-\frac{1}{2}} (\bar{X} - \mu) \to N_p(O, I) \text{ as } n \to \infty, \tag{3.5.21}$$

which completes the proof.

A parallel result also holds in the complex domain. Let $\tilde{X}_j$, $j = 1, \ldots, n$, 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 $\bar{\tilde{X}} = \frac{1}{n}(\tilde{X}_1 + \cdots + \tilde{X}_n)$, we have

$$\sqrt{n}\,\tilde{\Sigma}^{-\frac{1}{2}}(\bar{\tilde{X}} - \tilde{\mu}) \to N_p(O, I) \text{ as } n \to \infty. \tag{3.5a.5}$$

Exercises 3.5

3.5.1. By making use of the mgf or otherwise, show that the sample mean $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ 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 $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \ldots, n$ and iid. Let $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $\mathbf{M} = (\mu, \ldots, \mu)$ 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 $j = 1, \ldots, n$ and iid. Let $\mathbf{X} = (X_1, \ldots, X_n)$ the $p \times n$ sample matrix. Derive the density of $\text{tr}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})'$ where $\bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X})$ 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, $X_1, \ldots, X_n$, from this population where X_j is 2×1, $j = 1, \ldots, n$. 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 $u_1 = X_1'A_1X_1, u_2 = X_1'A_2X_1$. 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 $u_j = X_1'A_jX_1$, $A_j = A_j^2$, $j = 1, \ldots, k$, $A_1 + \cdots + A_k = I$. 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 $X_j \sim N_p(\mu, \Sigma)$, $\Sigma > O$, $j = 1, \ldots, n$ and iid. Let $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. 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 $\mathrm{tr}(\bar{X}'\Sigma^{-1}\bar{X})$.

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 $x_1, \ldots, x_p$ 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

$$\int_X |A|^{\frac{1}{2}}g((X - B)'A(X - B))dX = 1, \tag{3.6.1}$$

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),

$$\int_Y g(Y'Y)dY = 1 \tag{3.6.2}$$

where

$$Y'Y = y_1^2 + y_2^2 + \cdots + y_p^2, \ Y' = (y_1, \ldots, y_p).$$

We can further simplify (3.6.2) via a general polar coordinate transformation:

$$y_1 = r \ \sin \theta_1$$
$$y_2 = r \ \cos \theta_1 \sin \theta_2$$
$$y_3 = r \ \cos \theta_1 \cos \theta_2 \sin \theta_3$$
$$\vdots$$
$$y_{p-2} = r \ \cos \theta_1 \cdots \cos \theta_{p-3} \sin \theta_{p-2}$$
$$y_{p-1} = r \ \cos \theta_1 \cdots \cos \theta_{p-2} \sin \theta_{p-1}$$
$$y_p = r \ \cos \theta_1 \cdots \cos \theta_{p-1} \tag{3.6.3}$$

for $-\frac{\pi}{2} < \theta_j \le \frac{\pi}{2}$, $j = 1, \ldots, p-2$, $-\pi < \theta_{p-1} \le \pi$, $0 \le r < \infty$. It then follows that

$$dy_1 \wedge \ldots \wedge dy_p = r^{p-1}(\cos \theta_1)^{p-1} \cdots (\cos \theta_{p-1}) \, dr \wedge d\theta_1 \wedge \ldots \wedge d\theta_{p-1}. \tag{3.6.4}$$

Thus,

$$y_1^2 + \cdots + y_p^2 = r^2.$$

Given (3.6.3) and (3.6.4), observe that $r, \theta_1, \ldots, \theta_{p-1}$ are mutually independently distributed. Separating the factors containing θ_i from (3.6.4) and then, normalizing it, we have

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_i)^{p-i-1} d\theta_i = 1 \ \Rightarrow \ 2 \int_0^{\frac{\pi}{2}} (\cos \theta_i)^{p-i-1} d\theta_i = 1. \tag{i}$$

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

$$\int_0^1 v^{\frac{1}{2}-1}(1 - v)^{\frac{p-i}{2}-1} dv = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{p-i}{2})}{\Gamma(\frac{p-i+1}{2})}. \tag{ii}$$

Thus, the density of θ_j, denoted by $f_j(\theta_j)$, is

$$f_j(\theta_j) = \frac{\Gamma(\frac{p-j+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{p-j}{2})}(\cos \theta_j)^{p-j-1}, \ -\frac{\pi}{2} < \theta_j \le \frac{\pi}{2}, \tag{3.6.5}$$

and zero, elsewhere, for $j = 1, \ldots, p-2$, and

$$f_{p-1}(\theta_{p-1}) = \frac{1}{2\pi}, \ -\pi < \theta_{p-1} \le \pi, \tag{iii}$$

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

$$\left\{\prod_{j=1}^{p-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos\theta_j)^{p-j-1}\mathrm{d}\theta_j\right\} \int_{-\pi}^{\pi} \mathrm{d}\theta_{p-1} = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}. \tag{3.6.6}$$

The expression in (3.6.6), excluding 2, can also be obtained by making the transformation $s = Y'Y$ and then writing $\mathrm{d}s$ in terms of $\mathrm{d}Y$ by appealing to Theorem 4.2.3.

3.6.2. The density of $u = r^2$

From (3.6.2) and (3.6.3),

$$\frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_0^\infty r^{p-1} g(r^2)\mathrm{d}r = 1, \tag{3.6.7}$$

that is,

$$2\int_{r=0}^\infty r^{p-1} g(r^2)\mathrm{d}r = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}. \tag{iv}$$

Letting $u = r^2$, we have

$$\int_0^\infty u^{\frac{p}{2}-1} g(u)\mathrm{d}u = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}, \tag{v}$$

and the density of r, denoted by $f_r(r)$, is available from (3.6.7) as

$$f_r(r) = \frac{2\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} r^{p-1} g(r^2), \ 0 \le r < \infty, \tag{3.6.8}$$

and zero, elsewhere. The density of $u = r^2$ is then

$$f_u(u) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} u^{\frac{p}{2}-1} g(u), \ 0 \le u < \infty, \tag{3.6.9}$$

and zero, elsewhere. Considering the density of Y given in (3.6.2), we may observe that $y_1, \ldots, y_p$ are identically distributed.

Theorem 3.6.1. *If $y_j = r u_j$, $j = 1, \ldots, p$, in the transformation in (3.6.3), then $E[u_j^2] = \frac{1}{p}$, $j = 1, \ldots, p$, and $u_1, \ldots, u_p$ are uniformly distributed.*

Proof: From (3.6.3), $y_j = r u_j$, $j = 1, \ldots, p$. We may observe that $u_1^2 + \cdots + u_p^2 = 1$ and that $u_1, \ldots, u_p$ are identically distributed. Hence $E[u_1^2] + E[u_2^2] + \cdots + E[u_p^2] = 1 \Rightarrow E[u_j^2] = \frac{1}{p}$.

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),

$$\int_0^\infty u^{\frac{p}{2}-1} g(u) \mathrm{d}u = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}}. \tag{vi}$$

However,

$$E[u] = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} \int_{u=0}^\infty u^{\frac{p}{2}+1-1} g(u) \mathrm{d}u. \tag{vii}$$

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),

$$E[X] = |A|^{\frac{1}{2}} \int_X X\, g((X-B)'A(X-B)) \mathrm{d}X.$$

Noting that

$$E[X] = E[X - B + B] = B + E[X - B]$$
$$= B + |A|^{\frac{1}{2}} \int_X (X-B) g((X-B)'A(X-B)) \mathrm{d}X$$

and letting $Y = A^{\frac{1}{2}}(X - B)$, we have

$$E[X] = B + \int_Y Y g(Y'Y) \mathrm{d}Y, \quad -\infty < y_j < \infty, \ j = 1, \ldots, p.$$

But $Y'Y$ is even whereas each element in Y is linear and odd. Hence, if the integral exists, $\int_Y Y g(Y'Y) \mathrm{d}Y = O$ and so, $E[X] = B \equiv \mu$. Let $V = \mathrm{Cov}(X)$, the covariance matrix associated with X. Then

$$V = E[(X - \mu)(X - \mu)'] = A^{-\frac{1}{2}}[\int_Y (YY')g(Y'Y)dY]A^{-\frac{1}{2}},$$

where
$$Y = A^{\frac{1}{2}}(X - \mu),$$
(3.6.11)

$$YY' = \begin{bmatrix} y_1^2 & y_1y_2 & \cdots & y_1y_p \\ y_2y_1 & y_2^2 & \cdots & y_2y_p \\ \vdots & \vdots & \ddots & \vdots \\ y_py_1 & y_py_2 & \cdots & y_p^2 \end{bmatrix}.$$
(*viii*)

Since the non-diagonal elements, y_iy_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 $j = 1, \ldots, p$, 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,

$$\text{Cov}(X) = V = \frac{1}{2\pi}A^{-1} \text{ or } V = \frac{1}{p}E[r^2]A^{-1}.$$
(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

$$f(X) = |A|^{\frac{1}{2}}g((X - \mu)'A(X - \mu)), \ A > O, \ -\infty < x_j < \infty, \ -\infty < \mu_j < \infty$$
(3.6.13)

where $X' = (x_1, \ldots, x_p)$, $\mu' = (\mu_1, \ldots, \mu_p)$, $A = (a_{ij}) > O$. Consider the following partitioning of X, μ and A:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \ \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix}, \ A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

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,

$$
\begin{aligned}
(X - \mu)' A (X - \mu) &= (X_1 - \mu_{(1)})' A_{11}(X_1 - \mu_{(1)}) + 2(X_2 - \mu_{(2)})' A_{21}(X_1 - \mu_{(1)}) \\
&\quad + (X_2 - \mu_{(2)})' A_{22}(X_2 - \mu_{(2)}) \\
&= (X_1 - \mu_{(1)})'[A_{11} - A_{12} A_{22}^{-1} A_{21}](X_1 - \mu_{(1)}) \\
&\quad + (X_2 - \mu_{(2)} + C)' A_{22}(X_2 - \mu_{(2)} + C), \; C = A_{22}^{-1} A_{21}(X_1 - \mu_{(1)}).
\end{aligned}
$$

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

$$
\begin{aligned}
g_1(X_1) = |A|^{\frac{1}{2}} \int_{X_2} g((X_1 - \mu_{(1)})'[A_{11} - A_{12} A_{22}^{-1} A_{21}](X_1 - \mu_{(1)}) \\
+ (X_2 - \mu_{(2)} + C)' A_{22}(X_2 - \mu_{(2)} + C)) \mathrm{d} X_2.
\end{aligned}
$$

Letting $A_{22}^{\frac{1}{2}}(X_2 - \mu_{(2)} + C) = Y_2$, $\mathrm{d} Y_2 = |A_{22}|^{\frac{1}{2}} \mathrm{d} X_2$ and

$$
g_1(X_1) = |A|^{\frac{1}{2}} |A_{22}|^{-\frac{1}{2}} \int_{Y_2} g((X_1 - \mu_{(1)})'[A_{11} - A_{12} A_{22}^{-1} A_{21}](X_1 - \mu_{(1)}) + Y_2' Y_2) \, \mathrm{d} Y_2.
$$

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,

$$
\mathrm{d} Y_2 = \frac{\pi^{\frac{p_2}{2}}}{\Gamma(\frac{p_2}{2})} \int_{s_2 > 0} s_2^{\frac{p_2}{2} - 1} g(s_2 + u_1) \, \mathrm{d} s_2 \tag{3.6.14}
$$

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

$$
g_1(Y_{11}) = \frac{\pi^{\frac{p_2}{2}}}{\Gamma(\frac{p_2}{2})} \int_{s_2 > 0} s_2^{\frac{p_2}{2} - 1} g(s_2 + Y_{11}' Y_{11}) \, \mathrm{d} s_2. \tag{3.6.15}
$$

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:

$$g_2(X_2) = |A_{22} - A_{21}A_{11}^{-1}A_{12}|^{\frac{1}{2}} \frac{\pi^{\frac{p_1}{2}}}{\Gamma(\frac{p_1}{2})}$$

$$\times \int_{s_1>0} s_1^{\frac{p_1}{2}-1} g(s_1 + (X_2 - \mu_{(2)})'[A_{22} - A_{21}A_{11}^{-1}A_{12}](X_2 - \mu_{(2)})) \, ds_1,$$

$$g_{22}(Y_{22}) = \frac{\pi^{\frac{p_1}{2}}}{\Gamma(\frac{p_1}{2})} \int_{s_1>0} s_1^{\frac{p_1}{2}-1} g(s_1 + Y_{22}'Y_{22}) \, ds_1. \tag{3.6.16}$$

3.6.5. The characteristic function of an elliptically contoured distribution

Let T be a $p \times 1$ parameter vector, $T' = (t_1, \ldots, t_p)$, so that $T'X = t_1x_1 + \cdots + t_px_p$. Then, the characteristic function of X, denoted by $\phi_X(T)$, is $E[e^{i\,T'X}]$ where E denotes the expected value and $i = \sqrt{(-1)}$, that is,

$$\phi_X(T) = E[e^{i\,T'X}] = \int_X e^{i\,T'X}|A|^{\frac{1}{2}}g((X-\mu)'A(X-\mu))dX. \tag{3.6.17}$$

Writing X as $X - \mu + \mu$ and then making the transformation $Y = A^{\frac{1}{2}}(X - \mu)$, we have

$$\phi_X(T) = e^{i\,T'\mu} \int_Y e^{i\,T'A^{-\frac{1}{2}}Y} g(Y'Y)dY. \tag{3.6.18}$$

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,

$$\phi_X(T) = e^{i\,T'\mu} \int_Z e^{i\,T'A^{-\frac{1}{2}}P'Z} g(Z'Z)dZ \tag{3.6.19}$$

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,

$$\phi_X(T) = e^{i\,T'\mu}\psi(T'A^{-1}T) \tag{3.6.20}$$

where A^{-1} is proportional to Σ, the covariance matrix of X, and

$$\frac{\partial}{\partial T}\phi_X(T)|_{T=O} = i\mu \Rightarrow E(X) = \mu;$$

the reader may refer to Chap. 1 for vector/matrix derivatives. Now, considering $\phi_{X-\mu}(T)$, we have

$$\frac{\partial}{\partial T}\phi_{X-\mu}(T) = \frac{\partial}{\partial T}\psi(T'A^{-1}T) = \psi'(TA^{-1}T)2A^{-1}T$$

$$\Rightarrow \frac{\partial}{\partial T'}\psi(T'A^{-1}T) = \psi'(T'A^{-1}T)2T'A^{-1}$$

$$\Rightarrow \frac{\partial}{\partial T}\frac{\partial}{\partial T'}\psi(T'A^{-1}T) = \psi''(T'A^{-1}T)(2A^{-1}T)(2T'A^{-1}) + \psi'(TA^{-1}T)2A^{-1}$$

$$\Rightarrow \frac{\partial}{\partial T}\frac{\partial}{\partial T'}\psi(T'A^{-1}T)|_{T=O} = 2A^{-1},$$

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 $x_1, \ldots, x_k$ be independently distributed real scalar random variables with density functions $f_j(x_j)$, $j = 1, \ldots, k$. If the joint density of $x_1, \ldots, x_k$ 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 $x_1, \ldots, x_k$ are identically distributed as Gaussian random variables.

3.6.2. Letting the real scalar random variables $x_1, \ldots, x_k$ have a joint density such that $f(x_1, \ldots, x_k) = c$ for $x_1^2 + \cdots + x_k^2 \leq r^2$, $r > 0$, show that (1) $(x_1, \ldots, x_k)$ is uniformly distributed over the volume of the k-dimensional sphere; (2) $E[x_j] = 0$, $\text{Cov}(x_i, x_j) = 0$, $i \neq j = 1, \ldots, k$; (3) $x_1, \ldots, x_k$ 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 $g(u) = c_1(1 - a\,u)^\rho$, $a > 0$, $1 - a\,u > 0$ and c_1 is an appropriate constant. If $|A|^{\frac{1}{2}} g(u)$ 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 $g(u) = c_2(1 + a\,u)^{-\rho}$, where c_2 is an appropriate constant.

3.6.5. Solve Exercise 3.6.3 for $g(u) = c_3\,u^{\gamma-1}(1 - a\,u)^{\beta-1}$, $a > 0$, $1 - a\,u > 0$ and c_3 is an appropriate constant.

3.6.6. Solve Exercise 3.6.3 for $g(u) = c_4\,u^{\gamma-1}\mathrm{e}^{-a\,u}$, $a > 0$ where c_4 is an appropriate constant.

3.6.7. Solve Exercise 3.6.3 for $g(u) = c_5\,u^{\gamma-1}(1 + a\,u)^{-(\rho+\gamma)}$, $a > 0$ 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 $s = y_1^2 + \cdots + y_p^2$ where y_j, $j = 1, \ldots, p$, are real scalar random variables. Let $\mathrm{d}Y = \mathrm{d}y_1 \wedge \ldots \wedge \mathrm{d}y_p$ and let $\mathrm{d}s$ be the differential in s. Then, it can be shown that $\mathrm{d}Y = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})} s^{\frac{p}{2}-1}\mathrm{d}s$. 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) $g(u) = (a - c\,u)^\alpha$, $a > 0$, $c > 0$, $a - c\,u > 0$; (2) $g(u) = (a + c\,u)^{-\beta}$, $a > 0$, $c > 0$, 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.