

The exact and near-exact distributions of the main likelihood ratio test statistics used in the complex multivariate normal setting

Carlos A. Coelho · Barry C. Arnold ·
Filipe J. Marques

Received: 21 January 2014 / Accepted: 7 November 2014 / Published online: 6 December 2014
© Sociedad de Estadística e Investigación Operativa 2014

Abstract In this paper the authors show how it is possible to establish a common structure for the exact distribution of the main likelihood ratio test (LRT) statistics used in the complex multivariate normal setting. In contrast to what happens when dealing with real random variables, for complex random variables it is shown that it is possible to obtain closed-form expressions for the exact distributions of the LRT statistics to test independence, equality of mean vectors and the equality of an expected value matrix to a given matrix. For the LRT statistics to test sphericity and the equality of covariance matrices, cases where the exact distribution has a non-manageable expression, easy to implement and very accurate near-exact distributions are developed. Numerical studies show how these near-exact distributions outperform by far any other available approximations. As an example of application of the results obtained, the authors develop a near-exact approximation for the distribution of the LRT statistic to test the equality of several complex normal distributions.

Electronic supplementary material The online version of this article (doi:[10.1007/s11749-014-0418-y](https://doi.org/10.1007/s11749-014-0418-y)) contains supplementary material, which is available to authorized users.

C. A. Coelho · F. J. Marques
Departamento de Matemática, Faculdade de Ciências e Tecnologia,
Centro de Matemática e Aplicações (CMA-FCT/UNL),
Universidade Nova de Lisboa, Caparica, Portugal
e-mail: cmac@fct.unl.pt

F. J. Marques
e-mail: fjm@fct.unl.pt

B. C. Arnold (✉)
Statistics Department, University of California, Riverside, USA
e-mail: barry.arnold@ucr.edu

Keywords Covariance matrix · Equality of covariance matrices · Equality of mean vectors · Fourier transforms · Generalized integer gamma (GIG) distribution · Generalized near-integer gamma (GNIG) distribution · Independence · Mixtures · Expected value matrix · Sphericity · Statistical distributions (distribution functions)

Mathematics Subject Classification 62H05 · 62H10 · 62E15 · 62H15 · 62E20

1 Introduction

The implementation of many important likelihood ratio tests (LRTs) used in multivariate analysis is often hindered because of difficulties in handling the exact distribution, which is most of the time too complicated to be used in practice. On the other hand, the commonly available asymptotic distributions do not deliver the necessary accuracy, particularly for small samples and/or large numbers of variables involved.

In this paper the authors show how, in the complex multivariate normal setting, it is possible to obtain simple closed form expressions for the exact distribution of the LRT statistics to test (i) independence of sets of variables, (ii) equality of mean vectors and (iii) the equality of an expected value matrix to a given matrix, and very sharp near-exact distributions for the LRT statistics to test (iv) sphericity and (v) equality of covariance matrices.

These results are obtained, based on the fact that the LRT statistics in (i)–(v) have, under the null hypothesis, a common structure for their exact distributions, which for some $p \in \mathbb{N}$ and $u \in \mathbb{N}_0$, may be written as

$$\Lambda \stackrel{st}{\sim} \left(\prod_{j=1}^{p-1} e^{-Z_j} \right) \times \left(\prod_{k=1}^u Y_k \right). \tag{1}$$

In (1), Λ represents the LRT statistic, ‘ $\stackrel{st}{\sim}$ ’ is to be read as ‘is stochastically equivalent to’, and, for $r_j \in \mathbb{N}$ ($j = 1, \dots, p - 1$) and $a_k, b_k > 0$ ($k = 1, \dots, u$),

$$Z_j \sim \Gamma \left(r_j, \frac{n - 1 - j}{n} \right) \text{ and } Y_k \sim \text{Beta}(a_k, b_k), \tag{2}$$

are independent r.v.’s (random variables). In (2), $\Gamma(r_j, \lambda_j)$ denotes a gamma distribution with shape parameter r_j and rate parameter λ_j and n is the sample size. In (1), p represents the number of variables involved, for the tests in (i), (iv) and (v), or the sum of the number of variables plus the number of vectors minus 1, for the test in (ii), or the sum of the number of variables involved plus the number of columns in the matrix μ in (22), for the test in (iii). Also, in (1), we have $u = 0$ for the LRT statistics in (i), (ii) and (iii).

Krishnaiah et al. (1976) used the first four moments of the LRT statistics to test independence, equality of covariance matrices and sphericity to approximate the distributions of certain powers of these LRT statistics by Pearson type I distributions, while Fang et al. (1982) use infinite mixtures of beta distributions to asymptotically

approximate the distributions of the first two of those LRT statistics. [Khatri \(1965\)](#) obtained the exact distribution of the LRT statistic to test if an expected value matrix is null also in terms of an infinite mixture of beta distributions, while [Gupta \(1971\)](#) obtained closed form expressions for the distribution of the same test statistic for the cases of two and three variables, stating the possibility of extending such representations to a general number of variables, although without obtaining explicit expressions for the parameters in the distribution. [Pillai and Nagarsenker \(1971\)](#) and [Nagarsenker and Das \(1975\)](#) address the exact distribution of the LRT statistic for sphericity through contour integration, obtaining representations in terms of Meijer G functions and infinite series. Authors such as [Nagarsenker and Nagarsenker \(1981\)](#) and [Nagar et al. \(1985\)](#) studied a block sphericity test, which has as a particular case the common sphericity test, obtaining the exact distribution for the corresponding LRT statistic in the form of series representation, which in the case of the first authors has the form of an infinite mixture of beta distributions for the LRT statistic itself, or the form of an infinite mixture of chi-square distributions for its logarithm, while the second authors start with representations based on Meijer G and hypergeometric functions, to end-up with highly complicated series representations. [Gupta \(1976\)](#) and [Gupta and Rathie \(1983\)](#) address the non-null distribution of Wilks Λ statistic for MANOVA, obtaining in the first reference explicit expressions for the probability density function (pdf) of Λ only for $p = 2$ and $p = 3$ and only the general form of the expression for $p > 3$ and in the second reference the distribution for general p , in the form of Meijer G functions and what the authors claim to be computable series forms, which anyway in the null case remain too intricate. Somehow similar representations were obtained by [Pillai and Jouris \(1971\)](#) also for the LRT statistic for MANOVA and the LRT statistics to test equality of two covariance matrices and independence of two groups of variables. [Mathai \(1973\)](#) refers the non-computability of the representations involving Meijer G functions, which is still a reality, and obtains series representations for the distribution of the LRT for the equality of two covariance matrices and for sphericity, which anyway have very elaborate expressions.

In this paper the authors obtain

- very simple expressions for the exact pdf and cumulative distribution function (cdf) of the statistics in (i), (ii) and (iii), avoiding the need for the approximations in [Krishnaiah et al. \(1976\)](#), [Fang et al. \(1982\)](#), [Khatri \(1965\)](#) and the complex expressions in [Khatri \(1965\)](#), [Gupta \(1971\)](#), [Tang and Gupta \(1986\)](#), [Gupta \(1976\)](#), [Gupta and Rathie \(1983\)](#), [Pillai and Jouris \(1971\)](#), [Mathai \(1973\)](#) even for the null case,
- very accurate manageable near-exact distributions for the statistics in (iv) and (v), which are far more precise and manageable than the approximations and expressions in [Krishnaiah et al. \(1976\)](#), [Fang et al. \(1982\)](#), [Tang and Gupta \(1986\)](#), [Pillai and Jouris \(1971\)](#), [Mathai \(1973\)](#), [Pillai and Nagarsenker \(1971\)](#), [Nagar et al. \(1985\)](#), [Nagarsenker and Das \(1975\)](#), [Nagarsenker and Nagarsenker \(1981\)](#).

In Sect. 2 the exact distributions for the LRT statistics in (i), (ii) and (iii) are derived and it is shown that they have simple closed form expressions, which do not involve any infinite sums. In Sects. 3 and 4 the exact distribution of the LRT statistics in (iv) and (v) is obtained in the form of (1). Then, in Sect. 5, near-exact distributions for these statistics are developed, and in Sect. 6 some numerical studies are carried

out, confirming the very good performance of these near-exact distributions. Based on the results obtained, near-exact distributions for the LRT statistic to test equality of several complex normal distributions are developed in Sect. 7, and an example of application of these near-exact distributions is provided in the supplementary material in the Online Resource.

Near-exact distributions are asymptotic distributions developed under a new concept. Usually based on factorizations of the characteristic function (c.f.) of the test statistic under study, or of its logarithm, they are built by leaving unchanged one part of this c.f. and replacing asymptotically the remaining part, in such a way that the resulting c.f. corresponds to a manageable distribution. These distributions, when correctly built for statistics used in Multivariate Analysis, are not only asymptotic for increasing sample sizes but also have a marked asymptotic behavior for increasing number of variables and number of populations involved. All this amounts to obtaining asymptotic distributions that lie much closer to the exact distribution than common asymptotic distributions (Marques et al. 2011).

In this paper the definition of the complex multivariate normal distribution that will be used is that found in Wooding (1956), Goodman (1963a), Brillinger (2001, sec. 4.2) and Anderson (2003, prob. 2.64): the random vector \underline{X} ($p \times 1$) has a complex multivariate normal distribution, with expected value $\underline{\mu}$ and Hermitian variance-covariance matrix Σ if the pdf of \underline{X} is

$$f_{\underline{X}}(\underline{x}) = \pi^{-p} |\Sigma|^{-1} e^{-\overline{(\underline{x}-\underline{\mu})}' \Sigma^{-1} (\underline{x}-\underline{\mu})},$$

where $\overline{(\underline{x}-\underline{\mu})}$ denotes the complex conjugate of $(\underline{x}-\underline{\mu})$ and the prime denotes the transpose, with

$$\Sigma = 2\Theta + 2i\Phi,$$

where $\Theta = \text{Cov}(\text{Re}(\underline{X})) = \text{Cov}(\text{Im}(\underline{X}))$ and $\Phi = \text{Cov}(\text{Re}(\underline{X}), \text{Im}(\underline{X}))$ are, respectively, a positive-definite and a skew-symmetric matrix. In this case we will write

$$\underline{X} \sim CN_p(\underline{\mu}, \Sigma). \tag{3}$$

The complex multivariate normal distribution has applications in a wide range of areas, from crystallography (Pannu et al. 2003) to spectral analysis in time series (Brillinger and Krishnaiah 1982; Brillinger 2001; Shumway and Stoffer 2006; Goodman 1963a; Turin 1960; Krishnaiah et al. 1976) and to studies on satellite images, signal processing, communications and information theory (Conradsen 2012; Conradsen et al. 2003; Lehmann et al. 2007).

2 The exact distribution of the LRT statistics to test independence, the equality of mean vectors and the equality of an expected value matrix to a given matrix

In this section it is shown how, in contrast to what happens when dealing with real r.v.'s (Marques et al. 2011), in the complex case it is possible to obtain the exact

distribution for the negative logarithm of the LRT statistics to test independence, the equality of mean vectors or the equality of an expected value matrix to a given matrix as generalized integer gamma (GIG) distributions (Coelho 1998), for any number of variables or mean vectors involved, this way obtaining the exact distribution of the LRT test statistics in the form in (1), with $u = 0$. The simplicity of the expressions obtained for the exact pdf's and cdf's is quite striking when compared with the expressions obtained by other authors (Fang et al. 1982; Gupta and Rathie 1983; Khatri 1965; Krishnaiah et al. 1976; Pillai and Jouris 1971), and although bearing some resemblance with the expressions in Tang and Gupta (1986), they have, in contrast to these, the advantage of involving only finite sums. Even when compared with the quite close representations in Gupta (1971), Mathai (1973), Gupta (1976), they have the advantage of being more general and with coefficients which have much simpler forms. This is especially the case with respect to the representation in Gupta (1976), while in Gupta (1971) the author did not obtain explicit expressions for such coefficients.

2.1 The LRT statistic to test independence among sets of variables

Suppose that the random vector \underline{X} in (3) is split into m subvectors \underline{X}_k ($k = 1, \dots, m$) and that we intend to test the independence of these m subvectors. The partition of the vector \underline{X} will induce the following structure for Σ :

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1k} & \cdots & \Sigma_{1m} \\ \vdots & & \ddots & & \vdots \\ \Sigma_{k1} & \cdots & \Sigma_{kk} & \cdots & \Sigma_{km} \\ \vdots & & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \cdots & \Sigma_{mk} & \cdots & \Sigma_{mm} \end{bmatrix},$$

where $\Sigma_{kk} = \text{Var}(\underline{X}_k)$ ($k = 1, \dots, m$) and $\Sigma_{ij} = \text{Cov}(\underline{X}_i, \underline{X}_j)$ ($i, j \in \{1, \dots, m\}$). The hypothesis of mutual independence of the random subvectors \underline{X}_k , may thus be written as

$$H_0 : \Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{kk}, \dots, \Sigma_{mm}) \equiv \Sigma_{ij} = 0, \text{ for all } i \neq j. \tag{4}$$

Let us further suppose that, for each k , \underline{X}_k has p_k variables, with $p = \sum_{k=1}^m p_k$. Then, for a sample of size n , the LRT statistic to test H_0 in (4) is

$$\Lambda_1 = \left(\frac{|A|}{\prod_{k=1}^m |A_{kk}|} \right)^n, \tag{5}$$

where A is the maximum likelihood estimator (MLE) of Σ and A_{kk} its k -th diagonal block ($k = 1, \dots, m$), with

$$A = \frac{1}{n} \left(\overline{X - \frac{1}{n} E_{nn} X} \right)' \left(X - \frac{1}{n} E_{nn} X \right), \tag{6}$$

where once again the bar denotes the complex conjugate, X is the $n \times p$ sample matrix and E_{np} is an $n \times p$ matrix of ones (see Goodman 1963a; Anderson 2003, problem

3.11 for references concerning the MLE of Σ for the multivariate complex normal distribution).

Following a similar procedure to that used in the real case it is possible to show that, under H_0 in (4), for $h > p/n - 1$,

$$E \left(\Lambda_1^h \right) = \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} \frac{\Gamma(n-j) \Gamma(n-q_k-j+nh)}{\Gamma(n-q_k-j) \Gamma(n-j+nh)} = \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} E \left(Y_{jk}^{nh} \right), \quad (7)$$

where $Y_{jk} \sim \text{Beta}(n - q_k - j, q_k)$ are a set of $p = \sum_{k=1}^m p_k$ independent r.v.'s (for a proof, see Appendix B, section B.1 in the Online Resource, or section B.5.1 for an alternative proof based on the results in Jensen (1988, Thm. 5)). Note that (7) matches the expressions for the moments of Λ_1 in Krishnaiah et al. (1976) and Fang et al. (1982).

But then, since $0 < \Lambda_1 < 1$, its distribution is well defined by the set of its moments and consequently

$$\Lambda_1 \stackrel{st}{\sim} \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} (Y_{jk})^n.$$

Therefore, the c.f. of

$$W_1 = -\log \Lambda_1, \quad (8)$$

may be written as

$$\Phi_{W_1}(t) = E \left(e^{itW_1} \right) = E \left(\Lambda_1^{-it} \right) = \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} \frac{\Gamma(n-j)}{\Gamma(n-q_k-j)} \frac{\Gamma(n-q_k-j-nit)}{\Gamma(n-j-nit)}, \quad (9)$$

giving rise to Theorem 1.

Theorem 1 *The exact distribution of W_1 in (8) is a GIG distribution of depth $p - 1$ with pdf (see Marques et al. (2011, App. B) for the notation on the GIG pdf and cdf)*

$$f_{W_1}(w) = f^{\text{GIG}} \left(w \mid r_1, \dots, r_{p-1}; \frac{n-2}{n}, \dots, \frac{n-p-2}{n}; p-1 \right)$$

and cdf

$$F_{W_1}(w) = F^{\text{GIG}} \left(w \mid r_1, \dots, r_{p-1}; \frac{n-2}{n}, \dots, \frac{n-p-2}{n}; p-1 \right),$$

where $w > 0$ and

$$r_j = \begin{cases} h_j & j = 1 \\ h_j + r_{j-1} & j = 2, \dots, p - 1 \end{cases} \quad (10)$$

with, $h_j = (\text{number of } p_k \text{ greater or equal to } j) - 1$, for $j = 1, \dots, p - 1$.

Proof From (9), using, for any complex number z and positive integer n , the relation

$$\frac{\Gamma(z+n)}{\Gamma(z)} = \prod_{v=0}^{n-1} (z+v), \tag{11}$$

it follows that

$$\Phi_{W_1}(t) = \prod_{k=1}^{m-1} \prod_{j=1}^{p_k} \prod_{v=0}^{q_k-1} \frac{n-q_k-j+v}{n-q_k-j+v-nit} = \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n}\right)^{r_j} \left(\frac{n-1-j}{n} - it\right)^{-r_j},$$

which is the c.f. of the sum of $p-1$ independent gamma r.v.'s with shape parameters r_j , given by (10) in the body of the Theorem, and rate parameters $\frac{n-1-j}{n}$ ($j = 1, \dots, p-1$), which is a GIG distribution of depth $p-1$, with shape parameters r_j and rate parameters $\frac{n-1-j}{n}$.

Then Corollary 1 gives the exact pdf and cdf of $\Lambda_1 = e^{-W_1}$.

Corollary 1 *The exact pdf and cdf of the statistic $\Lambda_1 = e^{-W_1}$ in (5) are, for $0 < \ell \leq 1$,*

$$f_{\Lambda_1}(\ell) = f^{\text{GIG}}\left(-\log \ell \mid r_1, \dots, r_{p-1}; \frac{n-2}{n}, \dots, \frac{n-p-2}{n}; p-1\right) \frac{1}{\ell}$$

and

$$F_{\Lambda_1}(\ell) = 1 - F^{\text{GIG}}\left(-\log \ell \mid r_1, \dots, r_{p-1}; \frac{n-2}{n}, \dots, \frac{n-p-2}{n}; p-1\right),$$

for r_j given by (10).

The exact distribution of Λ_1 is thus of the form (1), with $u = 0$.

2.2 The LRT statistic to test the equality of several mean vectors

Let us suppose that

$$\underline{X}_j \sim CN_p(\underline{\mu}_j, \Sigma_j), \quad j = 1, \dots, q$$

and that, assuming $\Sigma_1 = \dots = \Sigma_q (= \Sigma)$, the null hypothesis to be tested is

$$H_0 : \underline{\mu}_1 = \dots = \underline{\mu}_q, \tag{12}$$

based on q independent samples, the j -th of which is from \underline{X}_j , with size n_j . Also, suppose that the j -th sample is stored in the $n_j \times p$ matrix X_j .

Then, the LRT statistic used to test H_0 in (12) is given by

$$\Lambda_2 = \left(\frac{|A|}{|A + B|} \right)^n \tag{13}$$

where $n = \sum_{j=1}^q n_j$,

$$A = \sum_{j=1}^q \left(\overline{X_j - E_{n_j} \tilde{X}_j} \right)' \left(X_j - E_{n_j} \tilde{X}_j \right) \quad \text{and} \quad B = \sum_{j=1}^q n_j \left(\tilde{X}_j - \tilde{X} \right) \left(\overline{\tilde{X}_j - \tilde{X}} \right)', \tag{14}$$

where once again the bar denotes the complex conjugate and where

$$\tilde{X}_j = \frac{1}{n_j} X_j' E_{n_j} 1,$$

is the vector of sample means from the j -th sample, and

$$\tilde{X} = \frac{1}{n} \sum_{j=1}^q n_j \tilde{X}_j.$$

The $p \times p$ matrix A has what is called a complex Wishart distribution [Goodman \(1963a, b\)](#) with $n - q$ degrees of freedom and parameter matrix Σ . This fact is denoted by $A \sim CW_p(n - q, \Sigma)$. Under H_0 in (12), $B \sim CW_p(q - 1, \Sigma)$, and then, given the independence, for normal r.v.'s, of the MLE's of the mean and variance, the matrices A and B are independent and thus, under H_0 in (12),

$$A + B \sim CW_p(n - 1, \Sigma).$$

It then follows that (see [Goodman 1963b](#)),

$$2^p |A| \stackrel{st}{\sim} |\Sigma| \prod_{j=1}^p W_j \quad \text{and} \quad 2^p |A + B| \stackrel{st}{\sim} |\Sigma| \prod_{j=1}^p Z_j,$$

where W_j ($j = 1, \dots, p$) and Z_j ($j = 1, \dots, p$) are two independent sets of p independent r.v.'s, with

$$W_j \sim \chi_{2(n-q-j+1)}^2 \quad \text{and} \quad Z_j = W_j + W_j^* \sim \chi_{2(n-1-j+1)}^2,$$

where each

$$W_j^* \sim \chi_{2(q-1)}^2, \quad j = 1, \dots, p,$$

is independent of W_j ($j = 1, \dots, p$).

Thus, under H_0 in (12),

$$\Lambda_2 \stackrel{st}{\sim} \prod_{j=1}^p (Y_j)^n \quad \text{where } Y_j \sim \text{Beta}(n - q - j + 1, q - 1), \quad (15)$$

are p independent r.v.'s.

But then, for $h > (p + q)/n - 1$,

$$E(\Lambda_2^h) = \prod_{j=1}^p \frac{\Gamma(n - j)}{\Gamma(n - q - j + 1)} \frac{\Gamma(n - q - j + 1 + nh)}{\Gamma(n - j + nh)}. \quad (16)$$

This result may also be obtained in an alternative way as in Appendix B.2 in the Online Resource.

From (16), the c.f. of

$$W_2 = -\log \Lambda_2 \quad (17)$$

may be written as

$$\Phi_{W_2}(t) = E(e^{itW_2}) = E(\Lambda_2^{-it}) = \prod_{j=1}^p \frac{\Gamma(n - j)}{\Gamma(n - q - j + 1)} \frac{\Gamma(n - q - j + 1 - nit)}{\Gamma(n - j - nit)}, \quad (18)$$

from which Theorem 2 may be established.

Theorem 2 *The exact distribution of W_2 in (17) is a GIG distribution of depth $p + q - 2$ with pdf*

$$f_{W_2}(w) = f^{\text{GIG}}\left(w \mid r_1, \dots, r_{p+q-2}; \frac{n-2}{n}, \dots, \frac{n-p-q+1}{n}; p+q-2\right)$$

and cdf

$$F_{W_2}(w) = F^{\text{GIG}}\left(w \mid r_1, \dots, r_{p+q-2}; \frac{n-2}{n}, \dots, \frac{n-p-q+1}{n}; p+q-2\right),$$

where $w > 0$ and

$$r_j = \begin{cases} h_j & j = 1 \\ h_j + r_{j-1} & j = 2, \dots, p + q - 2 \end{cases} \quad (19)$$

with $h_j = (\# \text{ of elements in } \{p, q - 1\} \text{ greater or equal to } j) - 1$, for $j = 1, \dots, p + q - 2$.

Proof From (18), using the relation in (11), the c.f. of W_2 may be written as

$$\Phi_{W_2}(t) = \prod_{j=1}^p \prod_{\ell=0}^{q-2} \frac{n - q - j + 1 + \ell}{n - q - j + 1 + \ell - nit} = \prod_{j=1}^{p+q-2} \left(\frac{n-1-j}{n} \right)^{r_j} \left(\frac{n-1-j}{n} - it \right)^{-r_j} \tag{20}$$

with r_j given by (19) in the body of the theorem. From (20) it is possible to conclude that the distribution of W_2 is indeed a GIG distribution of depth $p + q - 2$, with shape parameters r_j and rate parameters $\frac{n-1-j}{n}$ ($j = 1, \dots, p + q - 2$).

Then Corollary 2 gives the exact pdf and cdf of $\Lambda_2 = e^{-W_2}$.

Corollary 2 *The exact pdf and cdf of the statistic $\Lambda_2 = e^{-W_2}$ in (13) are, for $0 < \ell \leq 1$,*

$$f_{\Lambda_2}(\ell) = f^{\text{GIG}} \left(-\log \ell \mid r_1, \dots, r_{p+q-2}; \frac{n-2}{n}, \dots, \frac{n-p-q+1}{n}; p+q-2 \right) \frac{1}{\ell}$$

and

$$F_{\Lambda_2}(\ell) = 1 - F^{\text{GIG}} \left(-\log \ell \mid r_1, \dots, r_{p+q-2}; \frac{n-2}{n}, \dots, \frac{n-p-q+1}{n}; p+q-2 \right),$$

for r_j given by (19).

Once again, the distribution of Λ_2 is of the form (1) with $u = 0$.

2.3 The LRT statistic to test if an expected value matrix is equal to a given matrix

A generalization of the test for an expected value matrix developed in [Khatri \(1965, Subsection 3.2\)](#) will be addressed in this subsection. Let Z ($p \times n$) be a matrix with a complex multivariate normal distribution with expected value μM , where μ is a $p \times q$ complex matrix and M is $q \times n$ of rank q ($\leq n$), and variance $I_n \otimes \Sigma$, that is, with $\text{var}(\text{vec}(Z)) = I_n \otimes \Sigma$. This fact is denoted by

$$Z_{p \times n} \sim CN_{p \times n}(\mu M, I_n \otimes \Sigma). \tag{21}$$

Let us then suppose that the hypothesis to be tested, for a given matrix \mathcal{E} , is

$$H_0 : \mu_{(p \times q)} = \mathcal{E}_{(p \times q)}. \tag{22}$$

Then, according to [Khatri \(1965\)](#), the LRT statistic to test H_0 is

$$\Lambda_3 = \left(\frac{|\Psi|}{\left| \Psi + \frac{1}{n}(\beta - \mathcal{E})(M\overline{M}')(\beta - \mathcal{E})' \right|} \right)^n, \tag{23}$$

where

$$\Psi = \frac{1}{n}Z \left(I_n - \bar{M}'(M\bar{M}')^{-1}M \right) \bar{Z}' \quad \text{and} \quad \beta = Z\bar{M}'(M\bar{M}')^{-1} \tag{24}$$

are, respectively, the MLE's of Σ and μ , and as such independent.

But then, since $(I_n - \bar{M}'(M\bar{M}')^{-1}M)$ is the projector on the null space of the columns of M , it is idempotent with

$$\text{rank} \left(I_n - \bar{M}'(M\bar{M}')^{-1}M \right) = \text{tr} \left(I_n - \bar{M}'(M\bar{M}')^{-1}M \right) = n - q.$$

Then, given the distribution of Z in (21),

$$\Psi \sim CW_p \left(n - q, \frac{1}{n}\Sigma \right). \tag{25}$$

From (21),

$$\beta = Z\bar{M}'(M\bar{M}')^{-1} \sim CN_{p \times q} \left(\mu, (M\bar{M}')^{-1} \otimes \Sigma \right),$$

so that

$$(\beta - \mathcal{E})(M\bar{M}')^{1/2} \sim CN_{p \times q} \left((\mu - \mathcal{E})(M\bar{M}')^{1/2}, I_q \otimes \Sigma \right),$$

where, under H_0 in (22), $(\mu - \mathcal{E})(M\bar{M}')^{1/2} = 0$. Consequently under H_0 in (22),

$$(\beta - \mathcal{E})(M\bar{M}')(\overline{\beta - \mathcal{E}})' \sim CW_p(q, \Sigma),$$

independent of Ψ , so that, under H_0 in (22),

$$\Psi + \frac{1}{n}(\beta - \mathcal{E})(M\bar{M}')(\overline{\beta - \mathcal{E}})' \sim CW_p \left(n, \frac{1}{n}\Sigma \right).$$

Thus, following similar steps to those in Sect. 2.2 (see also Appendix B.2 in the Online Resource), under H_0 in (22) for $n > q + p - 1$,

$$\Lambda_3 \stackrel{st}{\sim} \prod_{j=1}^p (Y_j)^n \quad \text{where} \quad \begin{array}{l} Y_j \sim \text{Beta}(n - q - j + 1, q) \\ \text{are } p \text{ independent r.v.'s} \end{array} \tag{26}$$

so that, for $h > (q + p - 1)/n - 1$,

$$E \left(\Lambda_3^h \right) = \prod_{j=1}^p \frac{\Gamma(n + 1 - j)}{\Gamma(n + 1 - q - j)} \frac{\Gamma(n + 1 - q - j + nh)}{\Gamma(n + 1 - j + nh)}.$$

Thus, the c.f. of $W_3 = -\log \Lambda_3$ is

$$\begin{aligned} \Phi_{W_3}(t) &= E \left(e^{itW_3} \right) = E \left(\Lambda_3^{-it} \right) \\ &= \prod_{j=1}^p \frac{\Gamma(n + 1 - j) \Gamma(n + 1 - q - j - nit)}{\Gamma(n + 1 - q - j) \Gamma(n + 1 - j - nit)} = \prod_{j=1}^p \prod_{\ell=0}^{q-1} \frac{n + 1 - q - j + \ell}{n + 1 - q - j + \ell - nit} \end{aligned} \tag{27}$$

$$= \prod_{j=1}^{p+q-1} (n - j)^{r_j} (n - j - nit)^{-r_j} = \prod_{j=1}^{p+q-1} \left(\frac{n - j}{n} \right)^{r_j} \left(\frac{n - j}{n} - it \right)^{-r_j} \tag{28}$$

for r_j given by (19) in Theorem 2, with q replaced by $q + 1$.

Comparing (26) with (5.2.2) and (5.2.3) in Khatri (1965) and the first expression in (27) with (5.3.1) and (5.3.3) in the same reference, it can be seen that there is a small mistake in Khatri’s paper, where q has to be subtracted from the first argument of the beta r.v.’s in (5.2.2) and (5.2.3) and also from the arguments of all gamma functions in (5.3.1) of Khatri (1965).

From (28) Theorem 3 and Corollary 3 may be established.

Theorem 3 *The exact distribution of $W_3 = -\log \Lambda_3$ is a GIG distribution of depth $p + q - 1$ with pdf*

$$f_{W_3}(w) = f^{\text{GIG}} \left(w \mid r_1, \dots, r_{p+q-1}; \frac{n - 1}{n}, \dots, \frac{n - p - q + 1}{n}; p + q - 1 \right)$$

and cdf

$$F_{W_3}(w) = F^{\text{GIG}} \left(w \mid r_1, \dots, r_{p+q-1}; \frac{n - 1}{n}, \dots, \frac{n - p - q + 1}{n}; p + q - 1 \right),$$

where $w > 0$ and r_j ($j = 1, \dots, p$) are given by (19) in Theorem 2, with q replaced by $q + 1$.

Corollary 3 *The exact pdf and cdf of the statistic $\Lambda_3 = e^{-W_3}$ in (23) are, for $0 < \ell \leq 1$,*

$$f_{\Lambda_3}(\ell) = f^{\text{GIG}} \left(-\log \ell \mid r_1, \dots, r_{p+q-1}; \frac{n - 1}{n}, \dots, \frac{n - p - q + 1}{n}; p + q - 1 \right) \frac{1}{\ell}$$

and

$$F_{\Lambda_3}(\ell) = 1 - F^{\text{GIG}}\left(-\log \ell \mid r_1, \dots, r_{p+q-1}; \frac{n-1}{n}, \dots, \frac{n-p-q+1}{n}; p+q-1\right),$$

for r_j given by (19), with q replaced by $q + 1$.

This test may be easily extended to test hypothesis of the type

$$H_0 : \mu D = 0, \tag{29}$$

where D is a $q \times q$ non-random matrix.

By the invariance property of the MLEs, which entails the invariance of the LRT statistics, the LRT statistic to test H_0 in (29) may be obtained just by replacing $\beta = \hat{\mu}$ in the expression of Λ_3 by $\beta^* = \widehat{\mu D} = \hat{\mu} D$. Since in many cases D will not be full-rank, the distribution of the matrix B^* , where

$$B^* = \frac{1}{n} \beta^* (M \overline{M}') \overline{\beta^{*'}} \quad \text{with} \quad \beta^* = Z \overline{M}' (M \overline{M}')^{-1} D \tag{30}$$

is now a complex Wishart distribution with a number of degrees of freedom equal to

$$\text{rank}(\overline{M}' \overline{D}' (M \overline{M}')^{-1} D M)$$

which, since $M \overline{M}'$ is full-rank, will be equal to $\text{rank}(D)$.

The LRT statistic to test H_0 in (29) is thus

$$\Lambda_3^* = \left(\frac{|\Psi|}{|\Psi + B^*|} \right)^n, \tag{31}$$

with $\Psi \sim CW_p(n - q, \frac{1}{n} \Sigma)$, as in (25), and B^* as in (30). If $\text{rank}(D) = q^* (\leq q)$, then

$$B^* \sim CW_p\left(q^*, \frac{1}{n} \Sigma\right),$$

independent of Ψ , so that $\Psi + B^* \sim CW_p(n - q + q^*, \frac{1}{n} \Sigma)$ and as such (see Appendix B.2 in the Online Resource)

$$\Lambda_3^* \stackrel{st}{\sim} \prod_{j=1}^p (Y_j^*)^n \quad \text{where} \quad Y_j^* \sim \text{Beta}(n - q - j + 1, q^*)$$

are p independent r.v.'s.

Then, with an appropriate choice of the matrices Z , M and D , this test may be used to implement the test of equality of mean vectors in Sect. 2.2 Indeed, if Z is the sample matrix, with

$$Z = \begin{bmatrix} Z_{111} \dots Z_{11n_1} & Z_{211} \dots Z_{21n_2} & \dots & Z_{k11} \dots Z_{k1n_k} & \dots & Z_{q11} \dots Z_{q1n_q} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{1j1} \dots Z_{1jn_1} & Z_{2j1} \dots Z_{2jn_2} & \dots & Z_{kj1} \dots Z_{kjn_k} & \dots & Z_{qj1} \dots Z_{qjn_q} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{1p1} \dots Z_{1pn_1} & Z_{2p1} \dots Z_{2pn_2} & \dots & Z_{kp1} \dots Z_{kpn_k} & \dots & Z_{qp1} \dots Z_{qpn_q} \end{bmatrix},$$

where $Z_{kj\ell}$ represents the ℓ -th observation for the j -th variable in the k -th population, for $k = 1, \dots, q, j = 1, \dots, p$ and $\ell = 1, \dots, n_k$, the matrix M the design matrix, with

$$M = \begin{bmatrix} \overbrace{1 \ 1 \ \dots \ 1}^{n_1} & \overbrace{0 \ 0 \ \dots \ 0}^{n_2} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_k} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_q} \\ \overbrace{0 \ 0 \ \dots \ 0}^{n_1} & \overbrace{1 \ 1 \ \dots \ 1}^{n_2} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_k} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_q} \\ \vdots & \vdots & & \vdots & & \vdots \\ \overbrace{0 \ 0 \ \dots \ 0}^{n_1} & \overbrace{0 \ 0 \ \dots \ 0}^{n_2} & \dots & \overbrace{1 \ 1 \ \dots \ 1}^{n_k} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_q} \\ \vdots & \vdots & & \vdots & & \vdots \\ \overbrace{0 \ 0 \ \dots \ 0}^{n_1} & \overbrace{0 \ 0 \ \dots \ 0}^{n_2} & \dots & \overbrace{0 \ 0 \ \dots \ 0}^{n_k} & \dots & \overbrace{1 \ 1 \ \dots \ 1}^{n_q} \end{bmatrix}$$

and

$$D_{(q \times q)} = I_q - \frac{1}{n}MM'E_{qq} = \begin{bmatrix} 1 - \frac{n_1}{n} & -\frac{n_1}{n} & \dots & -\frac{n_1}{n} \\ -\frac{n_2}{n} & 1 - \frac{n_2}{n} & \dots & -\frac{n_2}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_q}{n} & -\frac{n_q}{n} & \dots & 1 - \frac{n_q}{n} \end{bmatrix},$$

clearly with $\text{rank}(D) = q - 1$, then the matrix B^* in (30) is exactly the same as matrix B in (14), while the matrix Ψ in (24) is the same as matrix A in (14) and the distribution of the LRT statistic Λ_3^* in (31) the same as that of Λ_2 in Sect. 2.2.

3 The exact distribution of the LRT statistic to test sphericity of the covariance matrix

Suppose that $\underline{X} \sim CN_p(\underline{\mu}, \Sigma)$, and that the hypothesis to be tested is

$$H_0 : \Sigma = \sigma^2 I_p \quad (\text{for some unspecified } \sigma^2 > 0). \tag{32}$$

Then, for a sample of size n , the LRT statistic to test H_0 in (32), is,

$$\Lambda_4 = \left(\frac{|A|}{\left(\text{tr} \frac{1}{p} A\right)^p} \right)^n, \tag{33}$$

where A is the MLE of Σ [see (6) and the note after this expression for references on the MLE of Σ]. Under H_0 in (32), the h -th moment of Λ_4 is

$$E \left(\Lambda_4^h \right) = \prod_{j=1}^{p-1} \frac{\Gamma \left(n - 1 + \frac{j}{p} \right)}{\Gamma(n - j - 1)} \frac{\Gamma(n - j - 1 + nh)}{\Gamma \left(n - 1 + \frac{j}{p} + nh \right)} = \prod_{j=1}^{p-1} E \left(Y_j \right)^{nh}, \quad (34)$$

for $h > \frac{p}{n} - 1$, and where $Y_j \sim \text{Beta} \left(n - j - 1, \frac{j}{p} + j \right)$ are $p - 1$ independent r.v.'s. See sections B.3 and B.5.2 of Appendix B in the Online Resource for details on the derivation of the expression for $E \left(\Lambda_4^h \right)$.

Since the whole set of moments of Λ_4 defines its distribution,

$$\Lambda_4 \stackrel{st}{\sim} \prod_{j=1}^{p-1} \left(Y_j \right)^n .$$

Then, for $W_4 = -\log \Lambda_4$, since the expression in (34) is well defined for h in a neighborhood of zero, it follows that (see Appendix C in the Online Resource for details),

$$\begin{aligned} \Phi_{W_4}(t) &= E \left(e^{itW_4} \right) = E \left(\Lambda_4^{-it} \right) = \prod_{j=1}^{p-1} \frac{\Gamma \left(n - 1 + \frac{j}{p} \right)}{\Gamma(n - j - 1)} \frac{\Gamma(n - j - 1 - nit)}{\Gamma \left(n - 1 + \frac{j}{p} - nit \right)} \\ &= \underbrace{\left\{ \prod_{j=1}^{p-1} \left(\frac{n - j - 1}{n} \right)^{p-j} \left(\frac{n - j - 1}{n} - it \right)^{-(p-j)} \right\}}_{\Phi_{1,W_4}(t)} \underbrace{\left\{ \prod_{j=1}^{p-1} \frac{\Gamma \left(n - 1 + \frac{j}{p} \right) \Gamma(n - 1 - nit)}{\Gamma(n - 1) \Gamma \left(n - 1 + \frac{j}{p} - nit \right)} \right\}}_{\Phi_{2,W_4}(t)} \end{aligned} \quad (35)$$

which shows that the exact distribution of W_4 is the same as the distribution of the sum of a GIG distributed r.v. with depth $p - 1$, with shape parameters $r_j = p - j$ and rate parameters $\frac{n-j-1}{n}$ ($j = 1, \dots, p - 1$), with $p - 1$ independent Logbeta $\left(n - 1, \frac{j}{p} \right)$ r.v.'s ($j = 1, \dots, p - 1$). This shows that the distribution of Λ_4 is thus of the form (1), with $u = p - 1$. The possibility of expressing the exact distribution of Λ_4 in this form will enable us to develop a very accurate near-exact distribution for Λ_4 in Sect. 5.1.

4 The exact distribution of the LRT statistic to test equality of several covariance matrices

In this section the authors will only address the case of equal sample sizes. The case of unequal sample sizes, given its complexity, will only be addressed in the next section.

Hence, suppose that $X_k \sim CN_p(\mu_k, \Sigma_k)$, ($k = 1, \dots, q$), and that, based on q independent samples, each of size n , the hypothesis to be tested is

$$H_0 : \Sigma_1 = \dots = \Sigma_q. \quad (36)$$

Then, the LRT statistic is

$$\Lambda_5 = \left(q^{pq} \frac{\prod_{k=1}^q |A_k|}{|A|^q} \right)^n, \quad (37)$$

where A_k is n times the MLE of Σ_k ($k = 1, \dots, q$), and $A = A_1 + \dots + A_q$ [see (6) and the note after for references on the MLE's of Σ_k]. Under H_0 in (36), the h -th moment of Λ_5 is

$$E\left(\Lambda_5^h\right) = \prod_{j=1}^p \prod_{k=1}^q \frac{\Gamma(n+n h-j)}{\Gamma(n-j)} \frac{\Gamma\left(n-1+\frac{k-j}{q}\right)}{\Gamma\left(n-1+\frac{k-j}{q}+n h\right)} = \prod_{j=1}^p \prod_{k=1}^q E\left(Y_{j k}^{n h}\right) \tag{38}$$

which may be obtained from expression (70) in Appendix B.4 in the Online Resource (see also subsection B.5.3 in Appendix B in the Online Resource for an alternative proof based on the results in Jensen (1988, Thm. 5)), and where

$$Y_{j k} \sim \text{Beta}\left(n-j, j-1+\frac{k-j}{q}\right)$$

are $p q - 1$ independent r.v.'s. Since the distribution of Λ_5 is determined by the set of its moments,

$$\Lambda_5 \stackrel{st}{\sim} \prod_{j=1}^p \prod_{k=1}^q (Y_{j k})^n .$$

(except for $j=k=1$)

Then, taking $W_5 = -\log \Lambda_5$, we have (see Appendix D in the Online Resource for details)

$$\begin{aligned} \Phi_{W_5}(t) &= E\left(e^{-i t W_5}\right) = E\left(\Lambda_5^{-i t}\right) = \prod_{j=1}^p \prod_{k=1}^q \frac{\Gamma\left(n-1+\frac{k-j}{q}\right) \Gamma(n-j-n i t)}{\Gamma\left(n-1+\frac{k-j}{q}-n i t\right) \Gamma(n-j)} \\ &= \underbrace{\left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n}\right)^{r_j} \left(\frac{n-1-j-i t}{n}\right)^{-r_j} \right\}}_{\Phi_{1, W_5}(t)} \\ &\quad \times \underbrace{\left\{ \prod_{j=1}^p \prod_{k=1}^q \frac{\Gamma\left(n-1+\frac{k-j}{q}\right) \Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor-n i t\right)}{\Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor\right) \Gamma\left(n-1+\frac{k-j}{q}-n i t\right)} \right\}}_{\Phi_{2, W_5}(t)} \end{aligned} \tag{39}$$

for

$$r_j = \begin{cases} q(q-1)\left(j-\frac{1}{2}\right) & j=1, \dots, \left\lceil \frac{p-1}{q} \right\rceil-1 \\ \frac{1}{2}\left(p-p^2+2 j p q+q-3 j q-q^2(j-1)^2\right) & j=\left\lceil \frac{p-1}{q} \right\rceil \\ q(p-j) & j=\left\lceil \frac{p-1}{q} \right\rceil+1, \dots, p-1, \end{cases} \tag{40}$$

which shows that the exact distribution of Λ_5 is of the form (1), with $u = pq$. As a consequence, as was the case for the statistic in the previous section, very accurate near-exact distributions can be developed for Λ_5 . See Sect. 5.2.

5 Near-exact distributions

Given the complexity of the exact distributions of the statistics in Sects. 3 and 4, or rather, of the expressions that might be obtained for their exact pdf's and cdf's and the concomitant issues related with their manageability, the development of near-exact distributions for such statistics arises as a sensible goal.

Let then, as a general notation, W stand for the negative logarithm of the LRT statistics in Sects. 3 and 4. The near-exact distributions developed in this section, for W , assume the form of mixtures of generalized near-integer gamma (GNIG) distributions (see Marques et al. 2011, App. B for the complete notation of the pdf and cdf of these distributions), which, for some $m^* \in \mathbb{N}$, will have pdf's and cdf's, respectively, of the form

$$f(w) = \sum_{v=0}^{m^*} \pi_v f^{\text{GNIG}} \left(w \mid r_1, \dots, r_{p-1}, r + v; \frac{n-2}{n}, \dots, \frac{n-p}{n}, \lambda; p \right) \quad (41)$$

and

$$F(w) = \sum_{v=0}^{m^*} \pi_v F^{\text{GNIG}} \left(w \mid r_1, \dots, r_{p-1}, r + v; \frac{n-2}{n}, \dots, \frac{n-p}{n}, \lambda; p \right), \quad (42)$$

where $w > 0$ represents a possible value of W and where, for the statistic in Sect. 3,

$$r = \frac{p-1}{2}, \quad \lambda = \frac{n-1}{n} \quad \text{and} \quad r_j = p-j, \quad (43)$$

for $j = 1, \dots, p-1$, while for the statistic in Sect. 4, r_j are given by (40),

$$r = p \frac{q-1}{2} \quad \text{and} \quad \lambda = \lambda^*,$$

where λ^* is the rate parameter in

$$\Phi_{2,W_5}^{**}(t) = p_1(\lambda^*)^{s_1}(\lambda^* - it)^{-s_1} + (1 - p_1)(\lambda^*)^{s_2}(\lambda^* - it)^{-s_2}, \quad (44)$$

which is determined together with s_1, s_2 and p_1 in such a way that the first 4 derivatives of $\Phi_{2,W_5}^{**}(t)$ and $\Phi_{2,W_5}(t)$ in (39) at $t = 0$ are the same.

These near-exact distributions are built by leaving $\Phi_{1,W_4}(t)$ and $\Phi_{1,W_5}(t)$, respectively, in (35) and (39), unchanged and then asymptotically approximating the c.f.'s $\Phi_{2,W_4}(t)$ and $\Phi_{2,W_5}(t)$ in the same expressions by c.f.'s of finite mixtures of gamma

distributions. This replacement is done based on the results in Sects. 5 and 6 of [Tricomi and Erdélyi \(1951\)](#), which show that any Logbeta(a, b) distribution may, for non-integer b , be asymptotically approximated by an infinite mixture of $\Gamma(b + \nu, a)$ ($\nu = 0, 1, \dots$) distributions.

This yields as near-exact distributions for W , finite mixtures of sums of p independent gamma r.v.'s, $p - 1$ of which have integer shape parameters r_j while the p th one has shape parameter r . For non-integer r , these are finite mixtures of GNIG distributions of depth p , but since for odd p , $r = \frac{p-1}{2}$ and for odd q , $r = p\frac{q-1}{2}$ are integers, for Λ_4 with odd p and for Λ_5 with odd q , the components of the mixtures are indeed GIG distributions (see [Marques et al. 2011](#), App. B for a complete notation of the pdf and cdf of these distributions), although for generality of notation the components of the mixtures in (41) and (42) are denoted as GNIG distributions.

The near-exact distributions built in this way are asymptotic not only for increasing sample sizes but also for increasing number of variables and populations involved, as is shown by the numerical studies carried out in the next section. These studies also show the extreme closeness of these near-exact distributions to the corresponding exact distributions even for very small sample sizes. Their parameters being very simple to determine, their implementation is very easy with the help of adequate software.

Further details on the construction of these near-exact distributions are given in the subsections ahead.

5.1 Near-exact distribution for the LRT statistic in Sect. 3

Using the results in Sects. 5 and 6 of [Tricomi and Erdélyi \(1951\)](#), as a first step, $\Phi_{2,W_4}(t)$ in (35), which is the c.f. of the sum of $p - 1$ independent Logbeta($n - 1, \frac{j}{p}$) r.v.'s ($j = 1, \dots, p - 1$), multiplied by n , is replaced by the c.f. of the sum of $p - 1$ independent infinite mixtures of $\Gamma(\frac{j}{p} + \nu, n - 1)$ distributions ($\nu = 0, 1, \dots$), multiplied by n , which is the c.f. of the sum of $p - 1$ independent infinite mixtures of $\Gamma(\frac{j}{p} + \nu, \frac{n-1}{n})$ distributions ($\nu = 0, 1, \dots$), and which in turn, and given that the rate parameters of the gamma distributions are not functions of either j or ν , is an infinite mixture of $\Gamma(\left(\sum_{j=1}^{p-1} \frac{j}{p}\right) + \nu, \frac{n-1}{n})$ distributions, where $\sum_{j=1}^{p-1} \frac{j}{p} = \frac{p-1}{2}$. This way we obtain a representation of the distribution which bears some resemblance with the representations in [Nagar et al. \(1985\)](#), [Nagarsenker and Das \(1975\)](#), [Tang and Gupta \(1986\)](#), but with coefficients which are much simpler to compute. This representation has the advantage of allowing for the easy development of very accurate near-exact approximations. Moreover, our representation does not have any parameters that are not well-defined, in contrast to the representation in [Tang and Gupta \(1986\)](#).

Therefore, to obtain a near-exact distribution for W_4 , $\Phi_{2,W_4}(t)$ in (35) will be replaced by the c.f. of a finite mixture of $\Gamma(\frac{p-1}{2} + \nu, \frac{n-1}{n})$ distributions, for $\nu = 0, \dots, m^*$,

$$\Phi_{2,W_4}^*(t) = \sum_{\nu=0}^{m^*} \pi_\nu \left(\frac{n-1}{n}\right)^{\frac{p-1}{2}+\nu} \left(\frac{n-1}{n} - it\right)^{-\left(\frac{p-1}{2}+\nu\right)}, \tag{45}$$

where the weights $\pi_0, \dots, \pi_{m^*-1}$ are determined as the numerical solution of the system of m^* equations

$$\frac{\partial^h}{\partial t^h} \Phi_{2,W_4}^*(t) \Big|_{t=0} = \frac{\partial^h}{\partial t^h} \Phi_{2,W_4}(t) \Big|_{t=0}, \quad h = 1, \dots, m^* \tag{46}$$

with $\pi_{m^*} = 1 - \sum_{v=0}^{m^*-1} \pi_v$. In this way these near-exact distributions have, by construction, the first m^* moments equal to the first m^* exact moments of W_4 . Hence,

$$\Phi_{W_4}^*(t) = \Phi_{1,W_4}(t) \Phi_{2,W_4}^*(t),$$

with $\Phi_{1,W_4}(t)$ given by (35) and $\Phi_{2,W_4}^*(t)$ given by (45), will be used as a near-exact c.f. for W_4 , to which corresponds the pdf and the cdf in (41) and (42), with r and λ given by (43).

Then, Theorem 4 and Corollary 4 may be stated.

Theorem 4 *Let $W_4 = -\log \Lambda_4$, where Λ_4 is the LRT statistic in (33). Then, for some $m^* \in \mathbb{N}$, distributions with pdf and cdf given by (41) and (42), with r, λ and r_j given by (43) and $\pi_0, \dots, \pi_{m^*-1}$ determined from (46), with $\Phi_{2,W_4}^*(t)$ given by (45) and $\pi_{m^*} = 1 - \sum_{v=0}^{m^*-1} \pi_v$, are near-exact distributions for W_4 .*

From Theorem 4, Corollary 4 is then readily obtained.

Corollary 4 *Let Λ_4 be the LRT statistic in (33). Then, distributions with pdf and cdf, respectively, given by $f(-\log \ell)^{\frac{1}{\nu}}$ and $1 - F(-\log \ell)$ where $f(\cdot)$ and $F(\cdot)$ are given by (41) and (42), with $0 < \ell \leq 1$ representing a possible value for Λ_4 and r_j ($j = 1, \dots, p - 1$), r, λ and π_v ($v = 0, \dots, m^*$) defined as in the previous theorem, are near-exact distributions for Λ_4 .*

5.2 Near-exact distribution for the LRT statistic in Sect. 4

5.2.1 The case of equal sample sizes

For the statistic Λ_5 in Sect. 4, in the case where all q samples have size n , we may asymptotically replace $\Phi_{2,W_5}(t)$ in (39), which is the c.f. of the sum of $pq - \min(p, q)$ independent Logbeta $\left(n - 1 + \left\lfloor \frac{k-j}{q} \right\rfloor, \frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor \right)$ r.v.'s ($j = 1, \dots, p; k = 1, \dots, q; j \neq k$), multiplied by n , by the c.f. of the sum of $pq - \min(p, q)$ independent infinite mixtures of $\Gamma \left(\frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor + \nu, \frac{1}{n} \left(n - 1 + \left\lfloor \frac{k-j}{q} \right\rfloor \right) \right)$ distributions ($\nu = 0, 1, \dots$). If it were the case that the rate parameters in these gamma distributions were functions of neither j nor k , as happened in the previous subsection, then this sum of mixtures would yield a simple mixture of gamma distributions. However, since now the rate parameters of the gamma distributions in the sum of mixtures are functions of both j and k , it renders it difficult to add the different mixtures of gamma distributions. For this reason it is used as an

asymptotic replacement for $\Phi_{2,W_5}(t)$ in (39) the c.f. of a finite mixture of $m^* + 1$ $\Gamma\left(\sum_{j=1}^p \sum_{k=1}^q \frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor + \nu, \lambda^*\right)$ distributions ($\nu = 0, \dots, m^*$), which is

$$\Phi_{2,W_5}^*(t) = \sum_{\nu=0}^{m^*} \pi_\nu (\lambda^*)^{r+\nu} (\lambda^* - it)^{-(r+\nu)}, \tag{47}$$

with

$$r = \sum_{j=1}^p \sum_{k=1}^q \frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor = p \frac{q-1}{2}, \tag{48}$$

and where λ^* is the common rate parameter of a mixture of two gamma distributions whose first four moments match the first four moments of the sum of independent logbeta r.v.'s whose c.f. is $\Phi_{2,W_5}(t)$ in (39), that is, where λ^* is the rate parameter in $\Phi_{2,W_5}^{**}(t)$ in (44), which is determined together with s_1, s_2 and p_1 in such a way that,

$$\frac{\partial^h}{\partial t^h} \Phi_{2,W_5}^{**}(t) \Big|_{t=0} = \frac{\partial^h}{\partial t^h} \Phi_{2,W_5}(t) \Big|_{t=0}, \quad h = 1, \dots, 4, \tag{49}$$

and where the weights π_ν ($\nu = 0, \dots, m^* - 1$), are then determined in such a way that

$$\frac{\partial^h}{\partial t^h} \Phi_{2,W_5}^*(t) \Big|_{t=0} = \frac{\partial^h}{\partial t^h} \Phi_{2,W_5}(t) \Big|_{t=0}, \quad h = 1, \dots, m^*, \tag{50}$$

with $\pi_{m^*} = 1 - \sum_{\nu=0}^{m^*-1} \pi_\nu$.

Therefore,

$$\Phi_{W_5}^*(t) = \Phi_{1,W_5}(t) \Phi_{2,W_5}^*(t),$$

with $\Phi_{1,W_5}(t)$ given by (39) and $\Phi_{2,W_5}^*(t)$ given by (47), will be used as a near-exact c.f. for W_5 , and as such one may thus enunciate the results summarized in Theorem 5 and Corollary 5.

Theorem 5 *Let $W_5 = -\log \Lambda_5$, where Λ_5 is the LRT statistic in (37). Then, for some $m^* \in \mathbb{N}$, distributions with pdf and cdf given by (41) and (42), with r_j given by (40), r given by (48) and $\lambda = \lambda^*$ obtained by solving (49) in order to λ^*, s_1, s_2 and p_1 , and $\pi_0, \dots, \pi_{m^*-1}$ determined from (50), with $\Phi_{2,W_5}^*(t)$ given by (47) and $\pi_{m^*} = 1 - \sum_{\nu=0}^{m^*-1} \pi_\nu$, are near-exact distributions for W_5 .*

Corollary 5 *Let Λ_5 be the LRT statistic in (37). Then distributions with pdf and cdf, respectively, given by $f(-\log \ell) \frac{1}{\ell}$ and $1 - F(-\log \ell)$ where $f(\cdot)$ and $F(\cdot)$ are given by (41) and (42), with $0 < \ell \leq 1$ representing the running value for Λ_5 and r_j ($j = 1, \dots, p - 1$), r, λ and π_ν ($\nu = 0, \dots, m^*$) defined as in the previous theorem, are near-exact distributions for Λ_5 .*

5.2.2 The case of unequal sample sizes

When the samples have different sizes, with the k -th sample having size n_k , the LRT statistic Λ_5 is now

$$\Lambda_5 = \frac{N^{Np}}{\prod_{k=1}^q n_k^{n_k p}} \frac{\prod_{k=1}^q |A_k|^{n_k}}{|A|^{Np}}, \tag{51}$$

where $N = \sum_{k=1}^q n_k$, and A_k ($k = 1, \dots, q$) is equal to n_k times the MLE of Σ_k and $A = A_1 + \dots + A_q$.

The c.f. of $W_5 = -\log \Lambda_5$ may be obtained from (69) in the supplementary material in the Online Resource, replacing h by $-it$, as

$$\begin{aligned} \Phi_{W_5}(t) &= E\left(\Lambda_5^{-it}\right) \\ &= \frac{N^{-Npit}}{\prod_{k=1}^q n_k^{-n_k pit}} \prod_{j=1}^p \left\{ \frac{\Gamma(N - q + 1 - j)}{\Gamma(N - q + 1 - j - Nit)} \prod_{k=1}^q \frac{\Gamma(n_k - j - n_k it)}{\Gamma(n_k - j)} \right\}. \end{aligned}$$

It happens that the exact distribution of Λ_5 in (51) is quite complicated and it is not even possible to give it a structure of the form (1).

However, using a procedure similar to the one used in Coelho and Marques (2012), the c.f. of W_5 may be written as

$$\Phi_{W_5}(t) = \Phi_{1, W_5}(t) \frac{\Phi_{W_5}(t)}{\Phi_{1, W_5}(t)},$$

where $\Phi_{1, W_5}(t)$ is given by (39), now with $n = N/q$. Then, to build a near-exact approximation for W_5 , $\Phi_{1, W_5}(t)$ will be left unchanged and $\frac{\Phi_{W_5}(t)}{\Phi_{1, W_5}(t)}$ will be replaced by $\Phi_{2, W_5}^*(t)$, given by (47), once again with λ^* defined in a similar manner to the one used in the previous subsection, with r either equal to s_1 in (44) or given by (48). As it will be shown in the next section, while the first choice for r will yield near-exact distributions that are asymptotic for increasing sample sizes as well as for increasing values of p and q , the second choice may give slightly better approximations for small values of p and even better approximations for large sample sizes. However, these latter near-exact distributions will no longer be asymptotic for increasing values of p or q , but only for increasing sample sizes.

As such, in terms of near-exact distributions, for the case of unequal sample sizes, similar results to the ones in Theorem 5 and Corollary 5 still hold, with the due changes in the parameters, namely with all the occurrences of n changed to N/q .

6 Numerical studies

In order to assess the proximity of the near-exact distributions developed in Sect. 5 to the exact distribution and their performance in different situations, we will use the measure

$$\Delta = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right| dt,$$

which yields

$$\max_{0 < \ell \leq 1} |F_\Lambda(\ell) - F_\Lambda^*(\ell)| = \max_{w > 0} |F_W(w) - F_W^*(w)| \leq \Delta,$$

where W represents generally either W_4 or W_5 , $\Lambda = e^{-W}$, $\Phi_W(t)$ represents the exact c.f. of W and $\Phi_W^*(t)$ its approximate c.f., usually a near-exact c.f., but occasionally an asymptotic or other c.f., $F_W(\cdot)$ and $F_\Lambda(\cdot)$ the exact cdf's of W and Λ , corresponding to $\Phi_W(t)$, and $F_W^*(\cdot)$ and $F_\Lambda^*(\cdot)$ the cdf's corresponding to $\Phi_W^*(t)$. Some further details on Δ and its relation with the Berry–Esseen bound are discussed in [Coelho and Marques \(2012\)](#). It should be noted that, quite clearly, a given value of Δ for any approximation to the exact distribution of either W_4 or W_5 will be the same as the corresponding value of the measure for the corresponding approximation to the distribution of, respectively, either $\Lambda_4 = e^{-W_4}$ or $\Lambda_5 = e^{-W_5}$.

All computations are done with Mathematica[®], version 7.0.0.

6.1 Numerical studies on the approximations for Λ_4

In [Table 1](#) may be analyzed the values of the measure Δ for the statistic Λ_4 , for values of p , the number of variables involved, ranging from 3 to 50 and sample sizes n which exceed p by 2, 12, 50 and 100. In order to compare the performance of the near-exact distributions with other available approximations in the literature, the Box type asymptotic distribution ([Box 1949](#)), used in [Nagarsenker and Nagarsenker \(1981\)](#) and the Pearson type I distribution used in [Krishnaiah et al. \(1976\)](#) have also been considered. The Pearson type I distribution was fitted by matching the first four exact moments of $\Lambda_4^{1/nb}$. In each case it is used for b the positive integer value which would give a better fit, this way obtaining indeed much better approximations than the ones obtained in [Krishnaiah et al. \(1976\)](#). These values of b are specified in [Table 1](#), inside square brackets, right after the value of Δ for the Pearson type I approximation. It may be noted that this Pearson type I approximation, with the slight change introduced, has a much better performance than truncations, even with a rather large number of terms, of any of the expansions in [Nagarsenker and Das \(1975\)](#).

The results in [Table 1](#) show that although the Pearson type I distribution, with the improvement introduced of finding the integer value of b which gives the best fit to the exact distribution, has a good performance, with an asymptotic behavior not only for increasing sample sizes but also for increasing values of p , it still is no match even for the near-exact distribution which matches only 4 moments. Besides, finding the integer value of b which gives the best fit for the Pearson type I distribution is not an easy task, and it was not possible to fit any such distribution for $p = 50$ and $n = 52$. The Box type asymptotic distribution has the poorest behavior of all the approximations. Its performance being worse for larger values of p , with values of Δ greater than 1, which shows that in these cases the asymptotic distribution yielded by the approximation is

Table 1 Values of the measure Δ for the near-exact and asymptotic distributions for the LRT statistic to test sphericity

p	n	Near-exact distribution # of exact moments matched				Box	Pearson type I
		4	6	10	15		
3	5	1.09×10^{-8}	2.74×10^{-11}	3.57×10^{-13}	9.28×10^{-16}	5.75×10^{-2}	2.17×10^{-3} [1]
	15	6.27×10^{-11}	1.21×10^{-13}	1.39×10^{-18}	4.27×10^{-23}	9.12×10^{-3}	2.78×10^{-5} [1]
	53	2.21×10^{-11}	3.42×10^{-14}	1.19×10^{-18}	7.73×10^{-23}	2.88×10^{-3}	6.85×10^{-7} [1]
	103	1.08×10^{-12}	6.43×10^{-16}	1.16×10^{-21}	2.71×10^{-27}	1.52×10^{-3}	9.58×10^{-8} [1]
10	12	3.54×10^{-16}	2.18×10^{-20}	2.61×10^{-28}	2.39×10^{-37}	8.07×10^{-1}	3.56×10^{-4} [3]
	22	2.19×10^{-15}	7.50×10^{-20}	2.26×10^{-27}	1.64×10^{-35}	3.52×10^{-1}	2.17×10^{-6} [5]
	60	1.21×10^{-15}	3.02×10^{-20}	7.29×10^{-28}	2.04×10^{-36}	1.16×10^{-1}	7.88×10^{-8} [5]
	110	9.24×10^{-17}	7.55×10^{-22}	3.06×10^{-30}	2.87×10^{-40}	6.17×10^{-2}	1.19×10^{-8} [5]
50	52	6.91×10^{-18}	3.35×10^{-25}	8.87×10^{-35}	8.46×10^{-47}	6.48×10^0	—
	62	2.57×10^{-16}	6.21×10^{-23}	1.88×10^{-32}	4.66×10^{-43}	4.85×10^0	1.27×10^{-6} [16]
	100	1.56×10^{-16}	3.48×10^{-23}	3.92×10^{-33}	8.04×10^{-46}	2.48×10^0	2.91×10^{-8} [24]
	150	8.77×10^{-18}	6.80×10^{-25}	5.13×10^{-36}	3.71×10^{-48}	1.57×10^0	3.97×10^{-9} [27]

not a true distribution. Moreover, it produces values of Δ quite close to 1 for a number of other cases.

The near-exact distributions always show a very good performance, with very low values of Δ . They exhibit a very good performance even for the smaller sample sizes, always showing an asymptotic behavior both for increasing sample sizes as well as for increasing values of p , and, of course, with the near-exact distributions which match a larger number of exact moments showing an increasingly better performance.

6.2 Numerical studies on the approximations for Λ_5

For the case of equal sample sizes, are shown in Table 2 the values of the measure Δ for the statistic Λ_5 , for different values of p , q and n , respectively, the number of variables involved, the number of covariance matrices involved and the common sample size of the q independent samples. In order to compare the performance of the near-exact distributions with other available approximations in the literature, the mixture of two beta distributions in Fang et al. (1982) and the Pearson type I distribution used by Krishnaiah et al. (1976) are used. These two distributions were fitted by matching, respectively, the first two and four exact moments of $\Lambda_5^{1/nb}$, for the choice of $b \in \mathbb{N}$ which gives the lowest value of Δ . These values of b are specified in Table 2, inside square brackets, right after the value of Δ for each of these approximations. The mixture of two beta distributions used is a mixture of two beta distributions with the same first parameter and a second parameter given by expressions (4.7) and (4.9) in Fang et al. (1982). Then the first weight in the mixture and the first parameter

Table 2 Values of the measure Δ for the near-exact and asymptotic distributions for the LRT statistic to test the equality of q covariance matrices for equal sample sizes

p	q	n	Near-exact distribution # of exact moments matched					Mixture of beta dists.	Pearson type I
			4	6	10	15	15		
5	3	7	1.42×10^{-8}	7.99×10^{-11}	5.30×10^{-15}	8.89×10^{-20}	9.63×10^{-5}	[6]	1.36×10^{-5}
		50	1.75×10^{-13}	1.43×10^{-17}	1.41×10^{-25}	2.15×10^{-34}	7.67×10^{-6}	[4]	4.78×10^{-9}
		100	4.25×10^{-15}	8.67×10^{-20}	4.90×10^{-29}	2.29×10^{-39}	2.68×10^{-7}	[8]	5.54×10^{-10}
10	10	7	3.19×10^{-10}	1.97×10^{-13}	1.08×10^{-19}	1.91×10^{-27}	1.29×10^{-4}	[26]	3.24×10^{-6}
		50	1.85×10^{-14}	2.77×10^{-19}	8.62×10^{-29}	8.70×10^{-39}	6.12×10^{-8}	[30]	1.56×10^{-9}
		100	5.28×10^{-16}	1.98×10^{-21}	3.93×10^{-32}	9.28×10^{-44}	2.18×10^{-8}	[30]	1.50×10^{-10}
15	15	7	2.25×10^{-10}	1.22×10^{-13}	5.54×10^{-20}	1.28×10^{-27}	6.80×10^{-6}	[37]	1.60×10^{-6}
		50	1.47×10^{-14}	1.88×10^{-19}	4.49×10^{-29}	1.48×10^{-40}	5.16×10^{-8}	[46]	7.25×10^{-10}
		100	4.26×10^{-16}	1.35×10^{-21}	1.97×10^{-32}	5.84×10^{-46}	6.09×10^{-9}	[45]	8.85×10^{-11}
15	3	17	5.70×10^{-11}	2.47×10^{-14}	8.54×10^{-21}	1.84×10^{-28}	4.02×10^{-4}	[16]	1.67×10^{-5}
		50	7.30×10^{-13}	6.13×10^{-17}	7.45×10^{-25}	3.16×10^{-34}	3.07×10^{-6}	[23]	1.90×10^{-8}
		100	1.22×10^{-14}	2.54×10^{-19}	1.88×10^{-28}	4.57×10^{-39}	4.69×10^{-7}	[23]	1.26×10^{-9}
10	10	17	7.78×10^{-14}	1.89×10^{-18}	1.58×10^{-27}	1.30×10^{-38}	3.99×10^{-4}	[65]	2.88×10^{-6}
		50	1.68×10^{-15}	1.18×10^{-20}	8.00×10^{-31}	3.46×10^{-43}	1.29×10^{-6}	[83]	2.78×10^{-9}
		100	3.64×10^{-17}	7.31×10^{-23}	3.86×10^{-34}	8.19×10^{-48}	8.30×10^{-8}	[86]	2.27×10^{-10}
15	15	17	5.05×10^{-15}	1.99×10^{-20}	3.35×10^{-30}	6.42×10^{-42}	6.35×10^{-5}	[95]	1.76×10^{-6}
		50	5.11×10^{-17}	5.68×10^{-22}	1.36×10^{-32}	6.31×10^{-46}	1.27×10^{-7}	[117]	1.64×10^{-9}
		100	6.63×10^{-18}	9.61×10^{-24}	1.34×10^{-35}	2.13×10^{-50}	6.96×10^{-7}	[118]	1.25×10^{-10}

in the beta distributions were determined by equating the two first exact moments of $A_5^{1/nb}$ and the two first moments of the mixture. The mixture of two beta distributions in Fang et al. (1982) is indeed not implemented exactly in this way. The decision to follow this alternative implementation was due to the facts that not only the definition of the parameter s in (4.9) and (4.10) in Fang et al. (1982) yields a conflicting definition for A_2 in the same reference, but also to the fact that by defining the parameters as suggested above, the approximation has indeed a much better performance.

From the results in Table 2 is possible to see how the Pearson type I approximation always outperforms the mixture of two beta distributions, mainly for the larger sample sizes. However, the Pearson type I distribution is always itself largely outperformed by the near-exact distribution which matches only 4 moments. The near-exact distributions once again show not only a marked asymptotic behavior for increasing sample sizes but also for increasing values of both p and q , as well as a consistent extremely good performance for very small sample sizes, avoiding the cumbersomeness of the determination of the best value for the parameter b in the Pearson type I and in the mixture of betas approximations.

In the case of different sample sizes, it was not feasible to fit the Pearson type I approximation to any case. Actually the authors in Krishnaiah et al. (1976) have only used this approximation for the case of equal sample sizes. However, it was possible to fit the mixture of two beta distributions to powers of $A_5^{1/b}$. But, when trying to find the best integer value of b many local minima were found, rendering the process of determining the best value of b a very frustrating and time consuming task. This, together with the fact that the near-exact distributions have a very good performance also for this different sample sizes case, for all sample sizes and all values of p and q , leads to the conclusion that it is indeed much preferable to use the near-exact approximations, which seem to be the only approximations with a very good and consistent performance.

The near-exact distributions which use for λ^* and r , respectively, the values of λ^* and s_1 in (44), obtained by solving the system of equations in (49) exhibit a very good performance even for the smaller sample sizes and a clear asymptotic behavior not only for increasing sample sizes but also for increasing values of p and q . The near-exact distributions which match only four exact moments always outperform the mixture of two beta distributions except for the smaller sample sizes for $p = 5$, with the measures for this mixture exhibiting a somewhat erratic behavior. Values of Δ for these near-exact distributions as well as for the mixture of two beta distributions, with the indication, inside square brackets, of the value of b used, may be analyzed in Table 3 in the supplementary material in the Online Resource.

The near-exact distributions which use the same value of λ^* as the ones above, but use for r the value given by (48), give better values of Δ for larger sample sizes and also for smaller values of p , but they lose their asymptotic character for increasing values of p , the number of variables involved. See Table 4 in the supplementary material in the Online Resource for the values of Δ for these near-exact distributions.

7 An application of the results obtained: the LRT statistic to test the equality of several multivariate complex normal distributions

In this section the authors show how the results obtained may be combined, as proposed in [Coelho and Marques \(2009\)](#), allowing a smooth path towards the development of near-exact distributions for LRT statistics for more elaborate hypotheses. As an application, the authors will show how very sharp near-exact distributions may be obtained for the LRT statistic to test the equality of several multivariate complex normal distributions, based on the results in the previous sections. Numerical results show the extreme closeness of the near-exact distribution obtained to the exact distribution of the LRT statistic.

Let us consider q independent samples of size n , the k -th of which obtained from $CN_p(\underline{\mu}_k, \Sigma_k)$ ($k = 1, \dots, q$), and that we want to test the hypothesis of equality of the q $CN_p(\underline{\mu}_k, \Sigma_k)$ distributions,

$$H_0 : \mu_1 = \dots = \underline{\mu}_k = \dots = \underline{\mu}_q, \Sigma_1 = \dots = \Sigma_k = \dots = \Sigma_q. \tag{52}$$

The null hypothesis H_0 in (52) may be written as

$$H_0 \equiv H_{02|05} \circ H_{05}$$

where ‘‘o’’ is to be read as ‘‘after’’ or ‘‘composed with’’, and where

$$H_{05} : \Sigma_1 = \dots = \Sigma_k = \dots = \Sigma_q \tag{53}$$

and

$$H_{02|05} : \underline{\mu}_1 = \dots = \underline{\mu}_k = \dots = \underline{\mu}_q, \text{ assuming } H_{05} \tag{54}$$

which are the null hypotheses in Sects. 4 and 2.2.

According to Lemma 10.3.1 in [Anderson \(2003\)](#), the LRT statistic to test H_0 in (52) will be the product of the LRT statistics to test H_{05} in (53) and $H_{02|05}$ in (54), so that this statistic will be

$$\Lambda = \underbrace{\left(q^{pq} \frac{\prod_{k=1}^q |A_k|}{|A|^q} \right)^n}_{\Lambda_5} \underbrace{\frac{|A|^{nq}}{|A+B|^{nq}}}_{\Lambda_{2|5}} = \left(q^{pq} \frac{\prod_{k=1}^q |A_k|}{|A+B|^q} \right)^n, \tag{55}$$

where Λ_5 is the LRT statistic to test H_{05} in (53) (see Sect. 4), $\Lambda_{2|5}$ is the LRT statistic to test $H_{02|05}$ (see Sect. 2.2), $A = A_1 + \dots + A_k + \dots + A_q$ and B is given by (14), with $n_j = n$, with A_k being equal to n times the MLE of Σ_k .

Given the independence of Λ_5 and $\Lambda_{2|5}$ (see [Anderson 2003](#), Lemma 10.4.1), the h -th moment of Λ may be obtained as

$$E \left(\Lambda^h \right) = E \left(\Lambda_5^h \right) E \left(\Lambda_{2|5}^h \right),$$

where $E(\Lambda_5^h)$ is the h -th moment of Λ_5 and as such, given by (38), and $E(\Lambda_{2|5}^h)$ is the h -th moment of $\Lambda_{2|5}$, and as such given by (16), using nq in place of n . In this context and assuming H_0 in (52), Λ_5 and $\Lambda_{2|5}$ are independent, given that Λ_5 is independent of $A = A_1 + \dots + A_k + \dots + A_q$, which may be shown using a similar procedure to the one in Anderson (2003, Sec. 10.4).

The c.f. of $W = -\log \Lambda$ may be written, from (39) and (20), where in this last expression n has to be replaced by nq , as

$$\begin{aligned} \Phi_W(t) = & \left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n} \right)^{s_j} \left(\frac{n-1-j}{n} - it \right)^{-s_j} \right\} \\ & \times \underbrace{\left\{ \prod_{\substack{j=1 \\ j \neq \alpha q - 1}}^{p+q-2} \left(\frac{n-(j+1)/q}{n} \right)^{r_j} \left(\frac{n-(j+1)/q}{n} - it \right)^{-r_j} \right\}}_{\phi_{1,W}(t)} \\ & \times \underbrace{\left\{ \prod_{j=1}^p \prod_{k=1}^q \frac{\Gamma\left(n-1+\frac{k-j}{q}\right) \Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor - nit\right)}{\Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor\right) \Gamma\left(n-1+\frac{k-j}{q} - nit\right)} \right\}}_{\Phi_{2,W}(t)}, \end{aligned}$$

where

$$s_j = \begin{cases} r_j^* & j = 1, \dots, p-1, j \neq \alpha q - 1 \\ r_j^* + r_{\alpha q - 1} & j = \alpha q - 1, \end{cases}$$

for $\alpha = 2, \dots, \lfloor \frac{p-1}{q} \rfloor$, with r_j ($j = 1, \dots, p+q-2$) given by (19) and r_j^* ($j = 1, \dots, p-1$) equal to r_j in (40).

Near-exact distributions for W and Λ are thus obtained by leaving $\Phi_{1,W}(t)$ unchanged and replacing $\Phi_{2,W}(t)$ by $\Phi_{2,W}^*(t)$ in (47), with r given by (48) and λ^* determined as described in Sect. 5.2.1. These near-exact distributions will have pdf's and cdf's which for Λ are given by

$$\begin{aligned} f_\Lambda(\ell) = & \sum_{v=0}^{m^*} \pi_v f^{\text{GNIG}} \left(-\log \ell \left| \underbrace{r_1, \dots, r_j, \dots, r_{p+q-2}}_{j \neq \alpha q - 1, \alpha = 2, \dots, \lfloor (p-1)/q \rfloor}, s_1, \dots, s_{p-1}, r + v; \right. \right. \\ & \underbrace{\frac{n-2/q}{n}, \dots, \frac{n-(j+1)/q}{n}, \dots, \frac{n-(p+q-2)/q}{n}}_{j \neq \alpha q - 1, \alpha = 2, \dots, \lfloor \frac{p-1}{q} \rfloor}, \frac{n-2}{n}, \dots, \frac{n-p}{n}, \lambda^*; \\ & \left. \left. 2p+q-2 - \left\lfloor \frac{p-1}{q} \right\rfloor \right) \right) \frac{1}{\ell} \end{aligned}$$

and

$$F_{\Lambda}(\ell) = 1 - \sum_{\nu=0}^{m^*} \pi_{\nu} F^{\text{GNIG}} \left(-\log \ell \left| \underbrace{r_1, \dots, r_j, \dots, r_{p+q-2}}_{j \neq \alpha q - 1, \alpha = 2, \dots, \lfloor (p-1)/q \rfloor}, s_1, \dots, s_{p-1}, r + \nu; \right. \right. \\
 \left. \left. \underbrace{\frac{n-2/q}{n}, \dots, \frac{n-(j+1)/q}{n}, \dots, \frac{n-(p+q-2)/q}{n}}_{j \neq \alpha q - 1, \alpha = 2, \dots, \lfloor \frac{p-1}{q} \rfloor}, \frac{n-2}{n}, \dots, \frac{n-p}{n}, \lambda^*; \right. \right. \\
 \left. \left. 2p + q - 2 - \left\lfloor \frac{p-1}{q} \right\rfloor \right), \tag{56}$$

where the GNIG components of the mixture may indeed be replaced by GIG components in case the shape parameter r , given by (48) is an integer.

These near-exact distributions yield very sharp approximations to the exact distribution, which are asymptotic not only for increasing sample sizes but also for increasing values of the number of variables in the distributions (p) and the number of distributions involved (q), as it may be seen from the values of the measure Δ in Table 5 in the Online Resource.

An example of application of this test is provided in the supplementary material in the Online Resource.

8 Conclusions

The authors were able to show that the main LRT statistics used in multivariate analysis in the complex multivariate normal setting all have a distribution which can be written in the form in (1). This enabled a much deeper insight into the true structure of such distributions and made possible to obtain very simple expressions for the exact distributions of the LRT statistics to test the independence of several groups of variables, the equality of several mean vectors and the equality of an expected value matrix to a given matrix and to develop very well-fitting near-exact approximations for the LRT statistics to test sphericity and the equality of covariance matrices.

Although (1) may seem to bear some resemblance to the form obtained for the real multivariate normal distribution in Marques et al. (2011), there are major differences between the complex and real cases: in the complex case (i) it is possible to obtain the exact distribution for the statistics discussed in Sect. 2 in a closed and very manageable form, thus avoiding the necessity for approximations as in Krishnaiah et al. (1976), Fang et al. (1982), Khatri (1965), or the complex expressions in Khatri (1965), Gupta (1971), Tang and Gupta (1986), Gupta (1976), Gupta and Rathie (1983), Pillai and Jouris (1971), Mathai (1973), even for the null case, and (ii) for the statistics in Sects. 3 and 4, the shape parameters of the gamma r.v.'s in the exact and near-exact distributions have much simpler expressions.

The near-exact approximations show very good performances even for very small sample sizes, displaying an asymptotic behavior not only for increasing sample sizes

but also for increasing number of variables and matrices involved, outperforming by far any other available approximations, such as those in Krishnaiah et al. (1976), Fang et al. (1982). Using one of the symbolic softwares available their implementation is very simple and their use enables us to avoid the cumbersome problems associated with the determination of the best value for the parameter b in the Pearson approximation. Moreover, the numerical determination of their parameters is well defined and poses no numerical problems. Benefiting from all these features, the near-exact approximations are, beyond any doubt, the recommended approximation for the statistics Λ_4 and Λ_5 .

The common structure of the distributions of the LRT statistics addressed, established in (1), also enables the smooth development of near-exact distributions for LRT statistics for more elaborate hypotheses. In Sect. 7, this feature is used to address the test of equality of multivariate complex normal distributions and to develop near-exact distributions for its LRT statistic. Numerical studies confirm the extreme closeness between the near-exact distributions obtained and the exact distribution.

Modules for the near-exact distributions developed in this paper are available in the web-site: <https://sites.google.com/site/nearexactdistributions/complex-normal>, and a short users-guide for these modules is made available in Appendix A.

Acknowledgments Research partially supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0297/2014 (CMA/UNL). The authors would like to express their gratitude to two anonymous Referees, the Associate Editor and also the Editor-in-Chief, whose suggestions contributed to a more solid and self-contained paper.

Appendix A: Short user's guide for the computational modules

A.1 Modules for the exact distributions of statistics in Sect. 2

Concerning the modules for the exact distributions of the three LRT statistics addressed in Sect. 2, there are five modules made available for each statistic Λ_i . These modules are CFWi, PDFWi, CDFWi, PDFLi and CDFLi, where i is to be replaced by either 1, 2 or 3. These modules give as output, respectively, the value for the c.f. of $W_i = -\log \Lambda_i$ ($i = 1, 2, 3$), the value for its pdf, the value for its cdf, the value for the pdf of Λ_i and the value for the cdf of Λ_i , computed at the value for the running variable.

In the case of the LRT statistic Λ_1 , the arguments of these five modules are, in this order: (i) a list whose components are the number of variables of each group, (ii) the sample size, (iii) the running variable for the function. For Λ_2 , the arguments are, in this order: (i) the number of variables (equal for each population), (ii) the number of mean vectors being tested, (iii) the overall sample size, (iv) the running variable for the function. In the case of Λ_3 , the arguments are (i) the number of variables, (ii) either the rank of the matrix D , in case this matrix is used, or the number of columns of the matrix μ .

The modules used to compute the pdf and cdf of $W_i = -\log \Lambda_i$ make use of the modules for the pdf and cdf of the GIG distribution, while the modules for the pdf and cdf of Λ_i make use of the corresponding modules for W_i , using the usual technique of r.v. transformation.

To obtain the computational modules and for a few examples of their use download the Mathematica[®] file `Mathematica_Modules_for_Exact_distr_stats_Sec_2.nb`, readily available from the web-page <https://sites.google.com/site/nearexactdistributions/complex-normal>.

A.2 Modules for the near-exact distributions of statistics in Sects. 3 and 4

For the statistics Λ_4 and Λ_5 , in Sects. 3 and 4 and whose near-exact distributions are developed in Sect. 5, there are also five modules made available for each statistic. These modules are: NECFW $_i$, NEPDFW $_i$, NECDFW $_i$, NEPDFL $_i$ and NECDFL $_i$, where i is to be replaced by either 4 or 5. These modules give as output, respectively, the value for the near-exact c.f. of $W_i = -\log \Lambda_i$ ($i = 4, 5$), the value for its near-exact pdf and cdf and the value for the near-exact pdf and cdf of Λ_i , computed at the value for the running variable.

In the case of the LRT statistic Λ_4 , the arguments of the five modules are, in the following order: (i) the number of variables, (ii) the sample size, (iii) the number of exact moments to be matched, (iv) the running variable for the function and (v) an optional argument which is the number of precision digits used to compute the exact moments to be matched by the near-exact distribution and which has a default value of 100. For Λ_5 , the arguments are, in the following order: (i) the number of variables, (ii) the number of covariance matrices being tested, (iii) the sample size, (iv) the number of exact moments to be matched, (v) the running variable for the function and (vi) the same optional argument as in (v) for Λ_4 .

The modules used to compute the near-exact pdf and cdf of $W_i = -\log \Lambda_i$ ($i = 4, 5$) make use of the modules for the pdf and cdf of the GIG and GNIG distributions, while the modules for the near-exact pdf and cdf of Λ_i make use of the corresponding modules for W_i , by using the usual technique of r.v. transformation.

To obtain the computational modules and for a few examples of their use download the Mathematica[®] file `Mathematica_Modules_for_Near-Exact_distr_stats_Sec_3_4.nb`, readily available from the web-page <https://sites.google.com/site/nearexactdistributions/complex-normal>.

References

- Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New York
- Box GEP (1949) A general distribution theory for a class of likelihood criteria. *Biometrika* 36:317–346
- Brillinger DR (2001) Time series: data analysis and theory. SIAM, Philadelphia
- Brillinger DR, Krishnaiah PR (eds) (1982) Handbook of statistics. Time series in frequency domain, vol 2. North Holland, Amsterdam
- Coelho CA (1998) The generalized integer gamma distribution—a basis for distributions in multivariate statistics. *J Multivar Anal* 64:86–102
- Coelho CA, Marques FJ (2009) The advantage of decomposing elaborate hypotheses on covariance matrices into conditionally independent hypotheses in building near-exact distributions for the test statistics. *Linear Algebra Appl* 430:2592–2606
- Coelho CA, Marques FJ (2012) Near-exact distributions for the likelihood ratio test statistic to test equality of several variance–covariance matrices in elliptically contoured distributions. *Comput Stat* 27:627–659

- Conradsen K (2012) Multivariate analysis of polarimetric SAR images. In: LINSTAT'12-international conference trends and perspectives in linear statistical inference and IWMS'12-21st international workshop on matrices and statistics. Abstract Book, pp 87–88
- Conradsen K, Nielsen AA, Schou J, Skiver H (2003) A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Trans Geosci Remote Sens* 41:4–19
- Fang C, Krishnaiah PR, Nagarsenker BN (1982) Asymptotic distributions of the likelihood ratio test statistics for covariance structures of the complex multivariate normal distributions. *J Multivar Anal* 12:597–611
- Goodman NR (1963a) Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann Math Stat* 34:152–177
- Goodman NR (1963b) The distribution of the determinant of a complex Wishart distributed matrix. *Ann Math Stat* 34:178–180
- Gupta AK (1971) Distribution of Wilks' likelihood-ratio criterion in the complex case. *Ann Inst Stat Math* 23:77–87
- Gupta AK (1976) Nonnull distribution of Wilks' statistic for MANOVA in the complex case. *Commun Stat Simul Comput* 5:177–188
- Gupta AK, Rathie PN (1983) Nonnull distributions of Wilks' Lambda in the complex case. *Statistica* 43:445–450
- Jensen ST (1988) Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann Stat* 16:302–322
- Khatri CG (1965) Classical statistical analysis based on a certain multivariate complex Gaussian distribution. *Ann Math Stat* 36:98–114
- Krishnaiah PR, Lee JC, Chang TC (1976) The distributions of the likelihood ratio statistics for tests of certain covariance structures of complex multivariate normal populations. *Biometrika* 63:543–549
- Lehmann NH, Fishler E, Haimovich AM, Blum RS, Chizhik D, Cimini LJ, Valenzuela RA (2007) Evaluation of transmit diversity in MIMO-radar direction finding. *IEEE Trans Signal Process* 55:2215–2225
- Marques FJ, Coelho CA, Arnold BC (2011) A general near-exact distribution theory for the most common likelihood ratio test statistics used in multivariate analysis. *Test* 20:180–203
- Mathai AM (1973) A few remarks about some recent articles on the exact distributions of multivariate test criteria: I. *Ann Inst Stat Math* 25:557–566
- Nagar DK, Jain SK, Gupta AK (1985) Distribution of LRC for testing sphericity of a complex multivariate Gaussian model. *Int J Math Math Sci* 8:555–562
- Nagarsenker BN, Das MM (1975) Exact distribution of sphericity criterion in the complex case and its percentage points. *Commun Stat* 4:363–374
- Nagarsenker BN, Nagarsenker PB (1981) Distribution of the likelihood ratio statistic for testing sphericity structure for a complex normal covariance matrix. *Sankhya Ser B* 43:352–359
- Pannu NS, Coy AJ, Read J (2003) Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Crystallogr D Biol Crystallogr* 59:1801–1808
- Pillai KCS, Jouris GM (1971) Some distribution problems in the multivariate complex Gaussian case. *Ann Math Stat* 42:517–525
- Pillai KCS, Nagarsenker BN (1971) On the distribution of the sphericity test criterion in classical and complex normal populations having unknown covariance matrices. *Ann Math Stat* 42:764–767
- Shumway RH, Stoffer DS (2006) Time series analysis and its applications: with R examples, 2nd edn. Springer, New York
- Tang J, Gupta AK (1986) Exact distribution of certain general test statistics in multivariate analysis. *Aust J Stat* 28:107–114
- Tricomi FG, Erdélyi A (1951) The asymptotic expansion of a ratio of gamma functions. *Pac J Math* 1:133–142
- Turin GL (1960) The characteristic function of Hermitian quadratic forms in complex normal variables. *Biometrika* 47:199–201
- Wooding RA (1956) The multivariate distribution of complex normal variables. *Biometrika* 43:212–215