Journal of Statistical Theory and Practice, 9:37–58, 2015

ISSN: 1559-8608 print / 1559-8616 online
DOI: 10.1080/15598608.2014.915157

Near-Exact Distributions for the Likelihood Ratio Test Statistic for Testing Multisample Independence—The Real and Complex Cases

FILIPE J. MARQUES AND CARLOS A. COELHO

Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

We consider a generalization of the well-known independence of several groups of variables test, which we designate by multisample independence test of several groups of variables. This new generalization is of great interest whenever we want to test if in different populations, which may follow a multivariate complex or real normal distribution, the Hermitian covariance matrices have the same structure and if there is independence between different groups of variables. We show that the test statistic has the distribution of the product of independent beta random variables; however, the explicit expressions for the probability density and cumulative distribution functions turn out to be very complicated and almost impossible to use in practice. Our objective is to use a breakthrough technique to develop near-exact distributions for the test statistic. These approximations are known to be highly accurate and easy to use, which facilitates and encourages their use in practice. Using a decomposition of the null hypothesis of the test into two null hypotheses we obtain, in a simple way, the likelihood ratio test statistic, the expression of its hth null moment, and the characteristic function of its logarithm. The decomposition of the null hypothesis also induces a factorization on the characteristic function of the logarithm of the test statistic, which enables the development of near-exact distributions. The numerical studies presented highlight the good properties of these approximations and show their great precision. Simulation studies conducted show the good power of the test proposed even for alternatives quite close to the null hypothesis. An example of application of the test is also provided.

AMS Subject Classification: 62H10; 62E20; 62H05; 62H15.

Keywords: Equality of covariance matrices test; Generalized near-integer gamma distribution; Independence of several groups of variables test; Mixtures; Multivariate complex normal distribution.

1. Introduction

The multisample independence test is used whenever we need to test whether, for q populations with real or complex multivariate normal distribution, we have equality of the q covariance matrices and whether they are all equal to a block diagonal matrix, indicating the independence of a given number of sets of variables. Using the results in Coelho et al.

Received 22 February 2013; accepted 8 April 2014.
Address correspondence to: Dr. Filipe J. Marques, Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, Caparica 2829-516, Portugal. Email: fjm@fct.unl.pt

(2010), Marques et al. (2011), and Coelho et al. (2012), we show that the exact distribution of the likelihood ratio test statistic to test multisample independence, for an underlying real or complex multivariate normal distribution, is the distribution of the product of independent beta distributions that has a very complicated and nonmanageable expression. To overcome this problem, we propose the use of near-exact distributions for the test statistic and for its logarithm (Coelho 2004; Coelho et al. 2010; Marques et al. 2011; Coelho and Marques 2009). These approximations are known to be highly accurate and easy to use and may be obtained as follows: First, we decompose the null hypothesis of the test into two null hypotheses, one being the null hypothesis to test the equality of several Hermitian covariance matrices and the other the null hypothesis to test the independence of several groups of variables; then, we write the characteristic function of the logarithm of the likelihood ratio test statistic as an adequate product of characteristic functions; and finally, we leave the major part of this characteristic function unchanged and we approximate the remaining part by another characteristic function in such a way that the resulting overall characteristic function corresponds to a well-known distribution.

In Marques and Coelho (2012), the authors present a short summary concerning the multisample independence test focused on the null hypothesis and on its decomposition, and they present an overview of how it is possible to develop near-exact distributions for the test statistic. In this article we extend that work by (i) explaining in detail the necessary procedures to develop near-exact distributions; (ii) analyzing what happens when the samples are extracted from real and complex multivariate normal distributions; and (iii) presenting numerical studies that show the excellent precision and good asymptotic properties of the near-exact distributions developed for both real and complex cases.

In Coelho et al. (2010) and Marques et al. (2011), the authors present a unified approach for developing near-exact distributions for the most common likelihood ratio test statistics in multivariate analysis, when samples are extracted from multivariate real normal populations; in particular, the authors develop near-exact distributions for the test statistics used to test independence of several groups of variables, equality of several covariance matrices, sphericity, and equality of means vectors. Interestingly, in Coelho et al. (2012), the authors show that is also possible to carry out a unified approach for developing near-exact distributions for the previous test statistics when the underlying populations have multivariate complex normal distributions.

In this article we address in detail the case where the samples are extracted from complex multivariate normal populations and we make a comparison with what happens in the real case. The authors address the multisample independence test by decomposing the null hypothesis for this test into the null hypotheses of the test of independence of groups of variables and the test of equality of covariance matrices. These two tests have already been addressed separately in the literature (Khatri 1965; Gupta 1971; Pillai and Jouris 1971; Mathai 1973; Krishnaiah et al. 1976; Fang et al. 1982; Tang and Gupta 1986), but to the authors' best knowledge, these tests have never been combined in order to test simultaneously the equality of covariance matrices of several populations and their block-diagonal structure. This test may even be applied in the context of testing for consistency of independence structures or statements (see Geiger and Pearl 1993).

The relevance of studying the complex case is well described in Coelho et al. (2012), where the authors report that the complex multivariate normal distribution arises in several areas of applications, such as crystallography, spectral analysis in time series, and performance of radar receivers (Goodman 1963; Hannan 1970; Conradsen et al. 2003; Pannu et al. 2003).

After obtaining the likelihood ratio test statistic, the authors develop very sharp near-exact approximations for its distribution. For simplicity we only address the case of samples with equal sizes; however, the case of samples of different sizes can be dealt using the results in Coelho et al. (2012) and Marques and Coelho (2012).

2. The Decomposition of the Null Hypothesis and the Test Statistic

The random vector $\underline{X}$ $(p \times 1)$ has a complex p-multivariate normal distribution, with expected value $\underline{\mu}$ and Hermitian variance–covariance matrix Σ, if the probability density function of $\underline{X}$ is

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

where $\overline{(\underline{x}-\underline{\mu})}$ denotes the complex conjugate of $(\underline{x}-\underline{\mu})$ and the prime denotes the transpose (for details see Coelho et al. 2012; Goodman 1963; Wooding 1956). Let us consider that $\underline{X}_k$ has a complex p-multivariate normal distribution with expected value $\underline{\mu}_k$ and variance–covariance matrix Σ_k, for $k = 1, \ldots, q$, and that we have q independent samples of dimension n, one from each of the populations $\underline{X}_k$. Let us suppose now that we are interested in testing the null hypothesis

$$H_0 : \Sigma_1 = \ldots = \Sigma_q = \begin{pmatrix} \Sigma_{11} & 0 & \ldots & 0 \\ 0 & \Sigma_{22} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \Sigma_{mm} \end{pmatrix} \tag{2}$$

where Σ_{ii} is of order $p_i \times p_i$ and $p = \sum_{i=1}^{m} p_i$. The null hypothesis in Eq. (2) can be written as

$$H_0 \equiv H_{0b|0a} \circ H_{0a} \tag{3}$$

where

$$H_{0a} : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_q (= \Sigma), \quad (\Sigma \text{ unspecified}) \tag{4}$$

is the null hypothesis to test the equality of q Hermitian covariance matrices of dimension $p \times p$, with

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \ldots & \Sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \ldots & \Sigma_{mm} \end{pmatrix},$$

and

$$\begin{gathered} H_{0b|0a} : \Sigma_{ij} = 0 \text{ for } i \neq j, \quad (i, j = 1, \ldots, m) \\ \text{assuming that } \Sigma_1 = \Sigma_2 = \ldots = \Sigma_q (= \Sigma) \end{gathered} \tag{5}$$

is the null hypothesis to test the independence of m groups of variables. The likelihood ratio test statistics to test H_{0a} and $H_{0b|0a}$ are given respectively by (see Coelho et al. 2012)

$$\Lambda_a = \left(q^{pq} \frac{\prod_{j=1}^{q} |A_j|}{|A|^q} \right)^n \quad \text{and} \quad \Lambda_{b|a} = \left(\frac{|A|}{\prod_{k=1}^{m} |A_{kk}|} \right)^{nq} \tag{6}$$

where A_k is the maximum likelihood estimator of $\Sigma_k (k = 1, \ldots, q)$, $A_1 + \ldots + A_q$ (see Goodman [1963] and problem 3.11 in Anderson [2003] for references concerning the maximum likelihood estimators of Σ_k in the complex case) and A_{kk} is the kth diagonal block of $A (k = 1, \ldots, m)$.

The use of this decomposition in Eq. (3) allows us to obtain the likelihood ratio test statistic to test H_0 in Eq. (2) as the product of the test statistics in Eq. (6), used to test H_{0a} and $H_{0b|0a}$ in Eqs. (4) and (5), respectively, and hence

$$\begin{aligned} \Lambda &= \Lambda_{b|a} \Lambda_a \\ &= \left(\frac{|A|}{\prod_{k=1}^{m} |A_{kk}|} \right)^{nq} \left(q^{pq} \frac{\prod_{j=1}^{q} |A_j|}{|A|^q} \right)^n = q^{pqn} \left(\frac{\prod_{j=1}^{q} |A_j|}{\prod_{k=1}^{m} |A_{kk}|^q} \right)^n . \end{aligned} \tag{7}$$

Given the independence of the likelihood ratio statistics Λ_a and $\Lambda_{b|a}$, under H_0, the expression of the hth null moment of Λ may be obtained from the expressions of the hth null moments of each of the test statistics Λ_a and $\Lambda_{b|a}$

$$E\left(\Lambda^h\right) = E\left(\Lambda_a^h\right) E\left(\Lambda_{b|a}^h\right),$$

and the expressions of $E(\Lambda_a^h)$ and $E(\Lambda_{b|a}^h)$ can be found in Coelho et al. (2012).

In the real case, if we consider that each of the random vectors $\underline{X}_k (p \times 1)$ has a real p-multivariate normal distribution, the likelihood ratio test statistic can be obtained in a similar way, using now the results in Marques et al. (2011), and is given by

$$\Lambda = q^{nqp/2} \left(\frac{\prod_{j=1}^{q} |A_j|}{\prod_{k=1}^{q} |A_{kk}|^q} \right)^{n/2} .$$

3. The Exact Distribution of $W = -\log \Lambda$ and Λ

The decomposition of the null hypothesis into two null hypotheses in Eq. (3) induces a factorization on the characteristic function, $\Phi_W(t)$, of $W = -\log \Lambda$ which may be written as the product of two characteristic functions, $\Phi_{Wa}(t)$, the characteristic function of $W_a = -\log \Lambda_a$, and $\Phi_{W_{b|a}}(t)$, the characteristic function of $W_{b|a} = -\log \Lambda_{b|a}$. The expressions of these characteristic functions can be easily obtained from the expressions of the hth null moments of Λ_a and $\Lambda_{b|a}$ available in Coelho et al. (2012) and Marques et al. (2011) for the complex and real cases, respectively, by using the equalities

$$\Phi_{W_a}(t) = E\left(\mathrm{e}^{W_a \mathrm{i}t}\right) = E\left(\mathrm{e}^{-\log \Lambda_a \mathrm{i}t}\right) = E\left(\Lambda_a^{-\mathrm{i}t}\right)$$

and

$$\Phi_{W_{b|a}}(t) = E\left(\mathrm{e}^{W_{b|a}\mathrm{i}t}\right) = E\left(\mathrm{e}^{-\log \Lambda_{b|a}\mathrm{i}t}\right) = E\left(\Lambda_{b|a}^{-\mathrm{i}t}\right) .$$

In Coelho et al. (2012), using the preceding relations, the authors present the characteristic function of W_a,

$$\Phi_{W_a}(t) = \prod_{j=1}^{p}\prod_{k=1}^{q} \frac{\Gamma\left(n-1+\frac{k-j}{q}\right)\Gamma\left(n-j-n\mathrm{i}t\right)}{\Gamma\left(n-1+\frac{k-j}{q}-n\mathrm{i}t\right)\Gamma\left(n-j\right)}$$

and of $W_{b|a}$ (in this case replacing the sample size by nq),

$$\Phi_{W_{b|a}}(t) = \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} \frac{\Gamma\left((n-1)q+1-j\right)}{\Gamma((n-1)q+1-q_k-j)} \frac{\Gamma((n-1)q+1-q_k-j-\mathrm{i}tnq)}{\Gamma((n-1)q+1-j-\mathrm{i}tnq)} ,$$

and therefore the characteristic function of W can be written as

$$\Phi_{W_a}(t) = \Phi_{W_a}(t) \times \Phi_{W_{b|a}}(t).$$

As such, we may observe that the exact distribution of W is the distribution of the sum of independent logbeta distributions with different parameters, some of them multiplied by n and the others by nq. This is the same as saying that the exact distribution of Λ is the distribution of the product of independent beta distributions raised to different exponents, more precisely,

$$\Lambda_a \sim \prod_{\substack{j=1 \\ \text{except for } j=k=1}}^{p}\prod_{k=1}^{q} \left(Y_{jk}\right)^n$$

with $Y_{jk} \sim Beta\,(n-j, j-1+(k-j)/q)$ and

$$\Lambda_{b|a} \sim \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} (Y_j)^{nq}$$

with $Y_j \sim Beta((n-1)q+1-q_k-j, q_k)$ where $q_k = p_{k+1} + \ldots + p_m$; as such,

$$\Lambda \sim \left\{ \prod_{\substack{j=1 \\ \text{except for } j=k=1}}^{p}\prod_{k=1}^{q} \left(Y_{jk}\right)^n \right\} \left\{ \prod_{k=1}^{m-1}\prod_{j=1}^{p_k} (Y_j)^{nq} \right\} .$$

Next, we show that is possible to obtain a different representation for the exact distribution of W and Λ that will be crucial for the development of near-exact distributions. Using the results in Coelho et al. (2012) it is possible to obtain the following factorization of the characteristic function of W_a:

$$\Phi_{W_a}(t) = \underbrace{\left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n} \right)^{r_{a,j}} \left(\frac{n-1-j}{n} - \mathrm{i}t \right)^{-r_{a,j}} \right\}}_{\Phi_{1,W_a}(t)}$$
$$\times \underbrace{\left\{ \prod_{j=1}^{p} \prod_{k=1}^{q} \frac{\Gamma\left(n-1+\frac{k-j}{q}\right) \Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor - n\mathrm{i}t\right)}{\Gamma\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor\right) \Gamma\left(n-1+\frac{k-j}{q} - n\mathrm{i}t\right)} \right\}}_{\Phi_{2,W_a}(t)} \tag{8}$$

for

$$r_{a,j} = \begin{cases} q(q-1)\left(j-\frac{1}{2}\right) & j = 1,\dots,\left\lceil \frac{p-1}{q} \right\rceil - 1 \\ \frac{1}{2}\left(p - p^2 + 2jpq + q \right. \\ \quad \left. -3jq - q^2(j-1)^2\right) & j = \left\lceil \frac{p-1}{q} \right\rceil \\ q(p-j) & j = \left\lceil \frac{p-1}{q} \right\rceil + 1,\dots,p-1\,. \end{cases} \tag{9}$$

Also from Coelho et al. (2012) we may obtain a different representation for the characteristic function of $W_{b|a}$, considering in this case the sample size equal to nq:

$$\Phi_{W_{b|a}}(t) = \prod_{j=1}^{p-1} \left(\frac{n-1-j/q}{n} \right)^{r_{b,j}} \left(\frac{n-1-j/q}{n} - \mathrm{i}t \right)^{-r_{b,j}} \tag{10}$$

with

$$r_{b,j} = \begin{cases} h_j & j = 1 \\ h_j + r_{b,j-1} & j = 2,\dots,p-1 \end{cases} \tag{11}$$

where for $j = 1,\dots,p-1$,

$$h_j = (\text{number of } p_k \text{ greater or equal to } j) - 1\,. \tag{12}$$

Using these factorizations we may establish the following theorem.

Theorem 3.1. *The characteristic function of W may be written as*

$$\Phi_W(t) = \left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n} \right)^{r^*_{a,j}} \left(\frac{n-1-j}{n} - \mathrm{i}t \right)^{-r^*_{a,j}} \right\}$$
$$\times \underbrace{\left\{ \prod_{\substack{j=1 \\ j \neq q,\dots,aq}}^{p-1} \left(\frac{n-1-j/q}{n} \right)^{r_{b,j}} \left(\frac{n-1-j/q}{n} - \mathrm{i}t \right)^{-r_{b,j}} \right\}}_{\Phi_{W_1}(t)} \tag{13}$$

$$\times \underbrace{\left\{ \prod_{j=1}^{p} \prod_{k=1}^{q} \frac{\Gamma\left(n-1+\frac{k-j}{q}\right)}{\Gamma\left(n-1+\left\lfloor\frac{k-j}{q}\right\rfloor\right)} \frac{\Gamma\left(n-1+\left\lfloor\frac{k-j}{q}\right\rfloor - nit\right)}{\Gamma\left(n-1+\frac{k-j}{q}-nit\right)} \right\}}_{\Phi_{2,W_a}(t)} \tag{14}$$

with $\alpha = \left\lfloor \frac{p-1}{q} \right\rfloor$, $r_{a,j}$ *and* $r_{b,j}$ *given respectively by expressions (9) and (11) and*

$$r^*_{a,j} = \begin{cases} r_{a,j} + r_{b,q\times j}, & j = 1, \cdots, \alpha \\ r_{a,j}, & j = \alpha+1, \cdots, p-1\,. \end{cases} \tag{15}$$

From expressions (13) and (14) we may conclude that the characteristic function of W is the characteristic function of the sum of two independent random variables, one with a generalized integer gamma distribution (Coelho, 1998), and the other with the distribution of the sum of $pq - \min(p,q)$ independent logbeta distributions, denoted by $Logbeta\left(n-1+\left\lfloor\frac{k-j}{q}\right\rfloor, \frac{k-j}{q} - \left\lfloor\frac{k-j}{q}\right\rfloor\right)$, multiplied by n. This representation enables a different look at the exact distribution of W that, as already mentioned, it will be the basis for the development of near-exact distributions.

For the real case, a similar structure can be obtained, now using the results in Marques et al. (2011); therefore, if we consider samples of the same size extracted from real normal multivariate distributions, the characteristic function of W is given by

$$\Phi_W(t) = \left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j}{n}\right)^{r^*_{a,j}} \left(\frac{n-j-1}{n} - it\right)^{-r^*_{a,j}} \right\}$$
$$\times \underbrace{\left\{ \prod_{j=1}^{p-1} \left(\frac{n-1-j/q}{n}\right)^{r_{b,j+1}} \left(\frac{n-1-j/q}{n} - it\right)^{-r_{b,j+1}} \right\}}_{\Phi_{W_1}(t)} \tag{16}$$

$$\times \left(\frac{\Gamma\left(\frac{(n-1)q}{2}\right) \Gamma\left(\frac{(n-1)q}{2} - \frac{1}{2} - \frac{(nq}{2}\mathrm{i}t\right)}{\Gamma\left(\frac{(n-1)q}{2} - \frac{1}{2}\right) \Gamma\left(\frac{(n-1)q}{2} - \frac{nq}{2}\mathrm{i}t\right)} \right)^{k^*}$$
$$\times \left\{ \prod_{j=1}^{\lfloor p/2 \rfloor} \prod_{k=1}^{q} \frac{\Gamma(a_j + b_{jk})}{\Gamma(a_j + b^*_{jk})} \frac{\Gamma(a_j + b^*_{jk} - n\mathrm{i}t)}{\Gamma(a_j + b_{jk} - n\mathrm{i}t)} \right\} \tag{17}$$
$$\times \underbrace{\left(\prod_{k=1}^{q} \frac{\Gamma(a_p + b_{pk})}{\Gamma(a_p + b^*_{pk})} \frac{\Gamma\left(a_p + b^*_{pk} - \frac{n}{2}\mathrm{i}t\right)}{\Gamma\left(a_p + b_{pk} - \frac{n}{2}\mathrm{i}t\right)} \right)^{p \perp\!\!\!\perp 2}}_{\Phi_{W_2}(t)}$$

but now with

$$r^*_{a,j} = \begin{cases} r_{a,j+1} + r_{b,q\times(j+1)}, & j = 1, \cdots, \alpha \\ r_{a,j+1}, & j = \alpha+1, \cdots, p-1\,. \end{cases} \tag{18}$$

where $r_{a,j}$ and $r_{b,j}$ are given by expressions (A.2) and (A.12) in Marques et al. (2012), and with a_j, b_{jk}, b^*_{jk}, a_p, b_{pk}, and b^*_{pk} given by expressions (A.10) and (A.11) of the same paper,

$$p \perp\!\!\!\perp 2 = \left\lfloor \frac{p+1}{2} \right\rfloor - \left\lfloor \frac{p}{2} \right\rfloor = \begin{cases} 0\,, & \text{for } p \text{ even} \\ 1\,, & \text{for } p \text{ odd} \end{cases},$$

and $k^* = \left\lfloor \frac{\ell}{2} \right\rfloor$, where ℓ denotes the number of $\underline{X}_k$'s with an odd number of variables.

4. Near-Exact Distributions of $W = -\log \Lambda$ and Λ

In this section we develop near-exact distributions for the distribution of $W = -\log \Lambda$ and of Λ. From Theorem 3.1 we have that the characteristic function of W may written as

$$\Phi_{W_1}(t) \times \Phi_{2,W_a}(t)\,,$$

with $\Phi_{W_1}(t)$ and $\Phi_{2,W_a}(t)$ given respectively by Eqs. (13) and (14). Starting from this representation, we may say that the near-exact distributions are obtained by replacing a part of this characteristic function, in this case $\Phi_{2,W_a}(t)$ in Eq. (14), by an asymptotic approximating characteristic function, $\Phi_{W^*}(t)$, so that the resulting characteristic function

$$\Phi_{W_{1(t)}} \times \Phi_{W^*}(t) \tag{19}$$

corresponds to a known distribution with manageable density and cumulative distribution functions.

As already mentioned, $\Phi_{2,W_a}(t)$ is the characteristic function of the sum of $pq - \min(p,q)$ independent *Logbeta* $\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor,\ \ \frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor\right)$ random variables ($j = 1,\ldots,p; k = 1,\ldots,q; j \neq k$), multiplied by n. From the results in Tricomi and Erdélyi (1951) we know that it is possible to asymptotically approximate the characteristic function $\Phi_{2,W_a}(t)$ by the characteristic function of the sum of $pq - \min(p,q)$ independent infinite mixtures of gamma distributions, denoted by $\Gamma\left(\frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor + \ell, \frac{1}{n}\left(n-1+\left\lfloor \frac{k-j}{q} \right\rfloor\right)\right)$, ($\ell = 0,1,\ldots$). As remarked in Coelho et al. (2012), if the rate parameters in these gamma distributions were not functions of j or k this sum of mixtures would yield a simple mixture of gamma distributions. Having this in mind, we propose to use as an asymptotic replacement for $\Phi_{2,W_a}(t)$ in Eq. (14) the characteristic function

$$\Phi_{W^*}(t) = \sum_{i=0}^{m^*} \pi_i (\lambda^*)^{r+i} (\lambda^* - \mathrm{i}t)^{-(r+i)} \tag{20}$$

with

$$r = \sum_{j=1}^{p} \sum_{k=1}^{q} \frac{k-j}{q} - \left\lfloor \frac{k-j}{q} \right\rfloor = p\,\frac{q-1}{2} \tag{21}$$

corresponding to the sum of the second parameters of the logbeta distributions easily identified in $\Phi_{2,W_a}(t)$ 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 random variables whose characteristic function is $\Phi_{2,W_a}(t)$ in Eq. (14). The characteristic function $\Phi_{W*}(t)$ is then the characteristic function of a finite mixture of $m*+1$ Gamma random variables, with distribution $\Gamma\,(r+i,\lambda^*)$, $(i=0,\ldots,m^*)$. The weights $\pi_i(i=0,\ldots,m^*-1)$, are then determined in such a way that

$$\frac{\partial^h}{\partial t^h}\,\Phi_{W^*}(t)\bigg|_{t=0}=\frac{\partial^h}{\partial t^h}\,\Phi_{2,W_a}(t)\bigg|_{t=0},h=1,\ldots,m^* \tag{22}$$

with

$$\pi_{m^*}=1-\sum_{i=0}^{m^*-1}\pi_i.$$

Given the preceding construction, whenever r is not an integer we have as near-exact distributions for W mixtures of m^*+1 GNIG (generalized near-integer gamma) distributions (see Coelho 2004), and when the parameter r is an integer we have as near-exact distributions for W mixtures of m^*+1 GIG (generalized integer gamma) distributions (see Coelho 1998). The near-exact distributions for Λ are easily obtained by simple transformation. The next theorem concerns the development of the near-exact distributions when r is not an integer; when r is an integer we mainly just have to replace in the mixtures the GNIG distribution by the GIG distribution.

Theorem 4.1. *If we asymptotically replace $\Phi_{2,W_a}(t)$ in Eq. (14) by the characteristic function $\Phi_{W^*}(t)$ in Eq. (20) we obtain as near-exact probability density and cumulative distribution functions for Λ (using the notation in Appendix 1 of Coelho and Marques [2012])*

$$f_\Lambda\,(w)\approx\sum_{i=0}^{m^*}\pi_i\,f^{GNIG}\left(-\log w|\left\{\boldsymbol{r_a^*},\boldsymbol{r_b}\right\},r+i;\{\boldsymbol{l_a},\boldsymbol{l_b}\},\lambda^*;d\right)\frac{1}{w} \tag{23}$$

and

$$F_\Lambda\,(w)\approx 1-\sum_{i=0}^{m^*}\pi_i F^{GNIG}\left(-\log w|\left\{\boldsymbol{r_a^*},\boldsymbol{r_b}\right\},r+i;\{\boldsymbol{l_a},\boldsymbol{l_b}\},\lambda^*;d\right) \tag{24}$$

where

$$\boldsymbol{r_a^*}=\left\{r_{a,j}^*,j=1,\ldots,p-1\right\},\boldsymbol{r_b}=\left\{r_{b,j},j=1,\ldots,p-1\text{ and }j\neq q,\ldots,\alpha q\right\},$$

$$\boldsymbol{l_a}=\left\{\frac{n-1-j}{n},j=1,\ldots,p-1\right\},$$

$$\boldsymbol{l_b}=\left\{\frac{n-1-j/q}{n},j=1,\ldots,p-1\text{ and }j\neq q,\ldots,\alpha q\right\},$$

$r_{a,j}^*$ are given in Eq. (15), $r_{b,j}^*$ in Eq. (11), r in Eq. (21) noninteger, $\alpha=\left\lfloor\frac{p-1}{q}\right\rfloor$, 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 random variables whose characteristic function is $\Phi_{2,W_a}(t)$ in Eq. (14), and where $0 < w < 1$ represents the running value of Λ. The weights π_i, $i = 0, \ldots, m^*$ are determined in such way that the system of equations in Eq. (22) holds.

In the real case, in order to obtain near-exact distributions for W, we have to replace in the characteristic function of W, which is given by the product of the characteristic functions in Eqs. (16) and (17), the characteristic function $\Phi_{W_2}(t)$ by the characteristic function $\Phi_{W^*}(t)$ in Eq. (20), now considering

$$r = \left\lfloor \frac{p+1}{2} \right\rfloor \frac{q-1}{2} + k^*/2, \tag{25}$$

where $k^* = \left\lfloor \frac{\ell}{2} \right\rfloor$, ℓ denotes the number of $\underline{X}_k$'s with an odd number of variates, and 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 random variables whose characteristic function is $\Phi_{W_2}(t)$ in Eq. (17), as described in Coelho et al. (2010). The weights π_i $(i = 0, \ldots, m^* - 1)$, would be determined in a similar way,

$$\left.\frac{\partial^h}{\partial t^h}\Phi_{W^*}(t)\right|_{t=0} = \left.\frac{\partial^h}{\partial t^h}\Phi_{W_2}(t)\right|_{t=0}, h = 1, \ldots, m^* \quad \text{with} \quad \pi_{m^*} = 1 - \sum_{i=0}^{m^*-1} \pi_i, \tag{26}$$

in this case with $\Phi_{W_2}(t)$ in Eq. (17).

This way the near-exact distributions for W are also mixtures of GNIG distributions when r in Eq. (25) is not an integer, or mixtures of GIG distributions when r is an integer. From the near-exact density and cumulative distribution functions of W, by simple transformation, it is possible to obtain the near-exact probability density and cumulative distribution functions for Λ that will have, when r is not integer, the same structure as the ones in Eqs. (23) and (24), now with $r^*_{a,j}$ given in Eq. (18), $r_{a,j}$ and $r_{b,j}$ given by expressions (A.2) and (A.12) in Marques et al. (2012), r in Eq. (25), 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 random variables whose characteristic function is $\Phi_{W_2}(t)$ in Eq. (17).

In the next section we study the quality of the near-exact distributions developed.

5. Numerical Studies

5.1. Assessing the Quality of the Near-Exact Approximations

In this section we use a measure of proximity between distribution functions that is based on the proximity between the corresponding characteristic functions. The measure can be define as

$$\Delta = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{\phi_Y(t) - \phi_n(t)}{t} \right| \mathrm{d}t, \tag{27}$$

where $F_Y(y)$ is the distribution function of a continuous random variable defined on a support S, $\phi_Y(t)$ is its characteristic function, and $\phi_n(t)$ and $F_n(y)$ are respectively the approximating characteristic and distribution functions. The measure satisfies

$$\max_{y \in S} |F_Y(y) - F_n(y)| \leq \Delta.$$

In Tables 1 and 2, we used the exact characteristic function of W in Eq. (13) and (14) and the near-exact characteristic function in Eq. (19). From these tables we may observe that the values of the measure Δ are considerably low, which indicates high precision for the near-exact approximations, and that the values decrease even more when we consider increasing values of n, p, and q.

In Tables 3 and 4 we analyze what happens in the real case. In this case we used for the measure Δ the exact characteristic function of W given by the product of the characteristic functions in Eqs. (16) and (17) and the near-exact characteristic function in Eq. (19), now considering $\Phi_{W_1}(t)$ in Eq. (16) and, in the expression of $\Phi_{W^*}(t)$, the value of r given by Eq. (25), and with λ^* being the common rate parameter of a mixture of two gamma distributions whose first four moments match the first four moments of W_2. From these tables we may observe that the values of the measure Δ are also considerably low and that in this case the near-exact approximations also display the same kind of asymptotic behavior for increasing values of n, p, and q. Comparing Tables 1 and 2 with Tables 3 and 4 we notice that the values of Δ are slightly lower for the complex case, which may be explained by the fact that in this case we always have the exact distribution of the test statistic used to test the independence of several sets of variables.

5.2. *Power Studies*

To study the power of this test in the complex case we consider two complex multivariate normal populations with zero mean vector and covariance matrices that violate the null hypothesis in different ways. As such, we considered $q = 2$ (number of populations), and

Table 1
Values of Δ for the near-exact distributions of $W = -\log \lambda^*$ and for increasing values of n—Complex case

					Near-exact distribution Number of exact moments matched			
p	p_i	k	q	n	4	6	10	15
10	{3,4,3}	3	3	12	4.1×10^{-11}	1.5×10^{-14}	3.8×10^{-21}	1.0×10^{-28}
				15	2.9×10^{-11}	9.0×10^{-15}	1.5×10^{-21}	3.5×10^{-29}
				20	7.8×10^{-12}	1.5×10^{-15}	7.7×10^{-23}	1.2×10^{-30}
				30	7.2×10^{-13}	4.9×10^{-17}	3.7×10^{-25}	3.8×10^{-33}
				40	9.9×10^{-14}	1.3×10^{-18}	7.0×10^{-26}	5.2×10^{-35}
				50	1.5×10^{-14}	5.9×10^{-19}	9.7×10^{-27}	1.8×10^{-36}
				100	8.0×10^{-16}	2.0×10^{-20}	9.1×10^{-30}	3.8×10^{-41}
15	{2,3,5,3}	4	5	20	4.6×10^{-12}	6.4×10^{-16}	2.1×10^{-23}	2.2×10^{-32}
				30	1.1×10^{-12}	8.9×10^{-17}	1.0×10^{-24}	2.9×10^{-34}
				40	2.6×10^{-13}	1.3×10^{-17}	5.3×10^{-26}	5.0×10^{-36}
				50	7.8×10^{-14}	2.6×10^{-18}	4.7×10^{-27}	1.8×10^{-37}
				100	1.5×10^{-15}	1.3×10^{-20}	1.7×10^{-30}	3.8×10^{-42}

Table 2
Values of Δ for the near-exact distributions of $W = -\log \Lambda^*$ and for increasing values of p and q—Complex case

					Near-exact distribution Number of exact moments matched			
p	p_i	k	q	n	4	6	10	15
7	{3,4}	2	2	9	1.4×10^{-9}	2.9×10^{-12}	2.5×10^{-17}	1.2×10^{-22}
			3		4.0×10^{-11}	6.4×10^{-15}	1.9×10^{-20}	4.7×10^{-27}
			5		1.9×10^{-11}	7.6×10^{-15}	2.6×10^{-21}	5.7×10^{-29}
			10		2.9×10^{-12}	2.6×10^{-16}	2.7×10^{-24}	1.8×10^{-33}
			20		2.8×10^{-12}	2.5×10^{-16}	2.7×10^{-24}	4.9×10^{-34}
7	{3,4}	2	2	9	1.4×10^{-9}	2.9×10^{-12}	2.5×10^{-17}	1.2×10^{-22}
15	{10,5}			17	3.5×10^{-10}	3.9×10^{-13}	1.0×10^{-18}	3.2×10^{-25}
20	{5,15}			22	1.8×10^{-10}	1.5×10^{-13}	2.3×10^{-19}	4.0×10^{-26}
30	{15,15}			32	3.3×10^{-11}	1.5×10^{-14}	4.7×10^{-21}	9.7×10^{-29}

Table 3
Values of Δ for the near-exact distributions of $W = -\log \Lambda^*$ and for increasing values of n—Real case

					Near-exact distribution Number of exact moments matched			
p	p_i	k	q	n	4	6	10	15
10	{3,4,3}	3	3	12	1.5×10^{-10}	1.1×10^{-13}	1.3×10^{-19}	1.9×10^{-26}
				15	9.0×10^{-11}	5.3×10^{-14}	3.5×10^{-20}	2.7×10^{-27}
				20	2.1×10^{-11}	6.9×10^{-15}	1.4×10^{-21}	4.1×10^{-29}
				30	1.7×10^{-12}	2.3×10^{-16}	6.1×10^{-24}	6.3×10^{-32}
				40	2.7×10^{-13}	1.6×10^{-17}	2.2×10^{-26}	2.2×10^{-32}
				50	5.8×10^{-14}	1.6×10^{-18}	1.3×10^{-26}	1.6×10^{-35}
				100	7.5×10^{-17}	1.5×10^{-20}	1.8×10^{-29}	2.4×10^{-40}
15	{2,3,5,3}	4	5	20	2.0×10^{-11}	6.6×10^{-15}	1.5×10^{-21}	2.2×10^{-29}
				30	2.9×10^{-12}	4.8×10^{-16}	2.5×10^{-23}	4.0×10^{-32}
				40	6.6×10^{-13}	6.1×10^{-17}	1.0×10^{-24}	8.3×10^{-35}
				50	2.0×10^{-13}	1.2×10^{-17}	8.1×10^{-26}	1.3×10^{-35}
				100	5.2×10^{-15}	7.9×10^{-20}	3.2×10^{-29}	1.5×10^{-40}

four groups of variables, respectively, with $p_1 = 3$, $p_2 = 4$, $p_3 = 4$, and $p_4 = 5$ variables and samples of size $n = 25$, for each population. We then analyze the power of the test when the null hypothesis H_0 is violated either by violation of the equality of the covariance matrices or by violation of the independence between groups of variables, or both.

We take Σ_1 and Σ_2 as being the covariance matrices for the two populations, with

Table 4
Values of Δ for the near-exact distributions of $W = -\log \Lambda^*$ and for increasing values of p and q—Real case

					Near-exact distribution Number of exact moments matched			
p	p_i	k	q	n	4	6	10	15
7	{3,4}	2	2	9	9.0×10^{-8}	1.0×10^{-9}	3.3×10^{-13}	4.5×10^{-17}
			3		2.6×10^{-10}	6.5×10^{-13}	2.9×10^{-17}	7.4×10^{-21}
			5		3.0×10^{-10}	2.2×10^{-13}	1.3×10^{-18}	4.0×10^{-24}
			10		3.7×10^{-10}	3.0×10^{-13}	3.9×10^{-19}	2.5×10^{-26}
			20		1.6×10^{-10}	9.4×10^{-14}	6.1×10^{-20}	2.8×10^{-27}
7	{3,4}	2	2	9	9.0×10^{-8}	1.0×10^{-9}	3.3×10^{-13}	4.5×10^{-17}
15	{10,5}			17	4.3×10^{-9}	1.6×10^{-11}	5.1×10^{-16}	4.1×10^{-21}
20	{5,15}			22	7.9×10^{-10}	1.5×10^{-12}	1.2×10^{-17}	2.0×10^{-23}
30	{15,15}			32	1.2×10^{-10}	9.3×10^{-14}	1.3×10^{-19}	2.1×10^{-26}

$$\Sigma_1 = \alpha \Sigma_2 \quad \text{and} \quad \Sigma_2 = (bM) \left(\frac{1}{b} I_{16} \right)$$

where the matrix M is given in the appendix, for $a = 1, 2, 3, 4, 5, 6$, and $b = 9/10$, 1, $107/100$.

If we consider the matrix M with the following partition, corresponding to the different groups of variables,

$$M = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_1 \end{array} \\ \left[\begin{array}{cccc} M_{11} & M_{12} & M_{13} & \mathbf{0}_{3\times 5} \\ M'_{12} & M_{22} & M_{23} & \mathbf{0}_{4\times 5} \\ M'_{13} & M'_{23} & M_{33} & \mathbf{0}_{4\times 5} \\ \mathbf{0}_{5\times 3} & \mathbf{0}_{5\times 4} & \mathbf{0}_{5\times 4} & M_{44} \end{array} \right] \begin{array}{c} p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \end{array} \tag{28}$$

then we may study the following cases of violation of H_0:

Case 1: violation of the independence between groups 1 and 2 of variables, $M_{12} \neq \mathbf{0}$, $M_{13} = \mathbf{0}_{3\times 4}$, $M_{23} = \mathbf{0}_{4\times 4}$.
Case 2: violation of the independence between groups 2 and 3 of variables, $M_{23} \neq \mathbf{0}$, $M_{12} = \mathbf{0}_{3\times 4}$, $M_{13} = \mathbf{0}_{3\times 4}$.
Case 3: violation of the independence between groups 1 and 3 of variables, $M_{13} \neq \mathbf{0}$, $M_{12} = \mathbf{0}_{3\times 4}$, $M_{23} = \mathbf{0}_{4\times 4}$.
Case 4: violation of the independence between groups 1, 2 and 3 of variables, $M_{12} \neq \mathbf{0}$, $M_{13} \neq \mathbf{0}$, $M_{23} \neq \mathbf{0}$.

For each of these four cases the authors considered the six different values of a and the three different values of b indicated earlier, generating a total of seventy two different scenarios.

We may note that for a only values equal or greater than 1 were considered, since taking for a their reciprocals would give exactly the same values for power as well as for the pseudo p-values reported ahead in the text.

The use of a real covariance matrix for the multivariate complex normal distributions considered, which configures a particular case of these distributions, is due to the fact that when using a complex matrix it is somehow even more difficult to be able to set up a 16 × 16 covariance matrix that is positive definite and that remains positive-definite after zeroing some off-diagonal blocks.

For each of the 72 scenarios, 100,000 pseudo-random samples were considered for each population, and the values obtained for the power of the test, for significance levels of 0.10, 0.05, and 0.01, are shown in Table 5, rounded to three decimal places.

As we may see from the values in this table, the power obtained for the different scenarios is always very high, which shows the very good aptitude of the test to detect violations of the null hypothesis. We should note that for $a = 1$ the hypothesis of equality of the two covariance matrices is not violated, and even in this case the values obtained for power, although slightly lower, are still very high.

Concerning the violation of the null hypothesis, given the structure of the covariance matrix, with several groups, each of them with several variables, it may be not too easy to assess how far or how close one is to the null hypothesis. In order to give the reader a sense of this, the authors computed what would be the p-values for the likelihood ratio test statistic for each one of the 72 different scenarios, as if the matrix used as the population covariance matrix were a sample matrix. These pseudo p-values may be analyzed, for the complex case, in Table 6, where we may see that most values are quite close to 1. In the

Table 5
Power of the test for the different cases and scenarios in the complex case

		Case 1			Case 2			Case 3			Case 4		
		Signif. levels			Signif. levels			Signif. levels			Signif. levels		
a	b	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
1	0.90	0.992	0.979	0.905	0.861	0.761	0.496	0.915	0.841	0.608	1.000	1.000	1.000
	1.00	0.999	0.997	0.975	0.910	0.834	0.599	0.964	0.922	0.753	1.000	1.000	1.000
	1.07	1.000	1.000	0.995	0.941	0.883	0.679	0.989	0.971	0.876	1.000	1.000	1.000
2	0.90	1.000	1.000	0.999	0.997	0.993	0.958	0.999	0.997	0.977	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	0.999	0.996	0.976	1.000	0.999	0.992	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	0.999	0.998	0.985	1.000	1.000	0.998	1.000	1.000	1.000
3	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6
"*p*-Values" for the different cases and scenarios in the complex case

a	b	Case 1	Case 2	Case 3	Case 4
1	0.90	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}
	1.00	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}
	1.07	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	2.728×10^{-8}
2	0.90	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}
	1.00	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}
	1.07	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	1.150×10^{-12}
3	0.90	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	9.994×10^{-1}
	1.00	1.000×10^{0}	1.000×10^{0}	1.000×10^{0}	7.508×10^{-1}
	1.07	9.997×10^{-1}	1.000×10^{0}	1.000×10^{0}	4.000×10^{-20}
4	0.90	9.997×10^{-1}	1.000×10^{0}	1.000×10^{0}	5.567×10^{-1}
	1.00	9.805×10^{-1}	1.000×10^{0}	9.999×10^{-1}	1.921×10^{-2}
	1.07	6.105×10^{-1}	9.998×10^{-1}	9.723×10^{-1}	5.165×10^{-28}
5	0.90	7.386×10^{-1}	9.677×10^{-1}	9.454×10^{-1}	1.173×10^{-2}
	1.00	2.855×10^{-1}	9.011×10^{-1}	7.861×10^{-1}	1.061×10^{-5}
	1.07	1.606×10^{-2}	7.766×10^{-1}	2.428×10^{-1}	1.311×10^{-35}
6	0.90	5.841×10^{-2}	3.130×10^{-1}	2.400×10^{-1}	1.326×10^{-5}
	1.00	3.958×10^{-3}	1.619×10^{-1}	7.634×10^{-2}	8.430×10^{-10}
	1.07	2.164×10^{-5}	7.232×10^{-2}	2.769×10^{-3}	1.060×10^{-42}

real case, with the exception of the cases $a = 4, 5, 6$, and $b = 1.07$, for which the pseudo *p*-values, rounded to three decimal places, were 0.976, 0.763, and 0.362, all other pseudo *p*-values are equal to 1.000. These values show that in both the real and complex cases we are computing power under quite "adverse" cases, that is, cases quite close to the null hypothesis, for which this hypothesis would not be rejected.

Similar to the complex case, for the real case the same population covariance matrices were used and the same 72 scenarios were considered. The values of power in Table 7 are generally very high, with many values equal to 1, when rounded to three decimal places.

The values of power in Tables 5 and 7 show that the test proposed has very good power against a wide range of scenarios violating H_0, even for cases that lie quite close to the null hypothesis. It is interesting to note that the test shows somehow higher power when the underlying population has a complex multivariate normal distribution.

5.3. *Error Levels for the Test*

In trying to assess the actual error level of the test, the authors considered 27 different cases in each one of the real and complex cases. For both the real and complex cases, three different arrangements of groups of variables were considered, and for each one of these we consider 2, 5, and 10 populations and in each of these cases we took samples of size $p + 2$, $p + 10$, and $p + 50$, where p represents the total number of variables involved. For each one of these scenarios, 100,000 pseudo-random samples were generated.

Then, to estimate the error levels, we counted the number of rejections, using the near-exact quantiles for $\alpha = 0.05$ and $\alpha = 0.01$, from a near-exact distribution that matches the first four exact moments.

Table 7
Power of the test for the different cases and scenarios in the real case

		Case 1			Case 2			Case 3			Case 4		
		Signif. levels			Signif. levels			Signif. levels			Signif. levels		
a	b	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
1	0.90	0.813	0.691	0.411	0.561	0.412	0.175	0.618	0.468	0.215	0.995	0.987	0.939
	1.00	0.952	0.897	0.701	0.687	0.541	0.272	0.785	0.658	0.377	1.000	1.000	0.998
	1.07	0.997	0.990	0.944	0.790	0.664	0.385	0.960	0.911	0.727	1.000	1.000	1.000
2	0.90	0.978	0.949	0.823	0.905	0.825	0.593	0.930	0.863	0.648	1.000	0.999	0.995
	1.00	0.997	0.991	0.951	0.950	0.896	0.710	0.975	0.941	0.801	1.000	1.000	1.000
	1.07	1.000	1.000	0.996	0.975	0.942	0.803	0.998	0.993	0.958	1.000	1.000	1.000
3	0.90	1.000	0.999	0.994	0.998	0.995	0.968	0.999	0.996	0.977	1.000	1.000	1.000
	1.00	1.000	1.000	0.999	0.999	0.998	0.984	1.000	0.999	0.992	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	0.999	0.992	1.000	1.000	0.999	1.000	1.000	1.000
4	0.90	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	0.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.07	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

The results, rounded to three decimal places, may be analyzed in Tables 8 and 9, where the values of the measure Δ in Eq. (27) are also reported.

The very low values of this measure assure the quality of the approximation and that the near-exact quantiles match at least six to seven decimal places of the exact ones for $-\log\,\Lambda$. This shows that the results in Tables 8 and 9 would be the same had we used the exact quantiles.

6. Application

In this section we illustrate the application of this test in a change detection method for polarimetric synthetic aperture radar (SAR) data. In Conradsen et al. (2003), the authors show that the test of equality of two complex Wishart distributions, assuming a block-diagonal structure for both matrices can be applied to change detection in polarimetric SAR data. With our test procedure it is possible to test simultaneously the equality of the two complex Wishart distributions and the block diagonal structure. In Conradsen et al. (2003), the authors state that "when working with multilook fully polarimetric SAR data, an appropriate way of representing the backscattered signal consists of the so-called covariance matrix. For each pixel, this is a 3×3 Hermitian positive definite matrix that follows a complex Wishart distribution" and which is assumed in the paper to have the following block structure:

Table 8
Simulated error levels for different scenarios in the real case

	$p_i = \{7,6\}$								
	$n = 15$			$n = 23$			$n = 63$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.050	0.010	1.2×10^{-9}	0.049	0.010	1.8×10^{-10}	0.050	0.010	1.9×10^{-13}
$q = 5$	0.049	0.010	2.6×10^{-11}	0.049	0.009	1.1×10^{-11}	0.049	0.010	5.6×10^{-14}
$q = 10$	0.049	0.010	4.4×10^{-13}	0.049	0.010	2.3×10^{-13}	0.051	0.010	1.5×10^{-15}

	$p_i = \{6,4,9\}$								
	$n = 21$			$n = 29$			$n = 69$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.051	0.010	1.2×10^{-9}	0.051	0.010	4.7×10^{-10}	0.049	0.010	2.3×10^{-12}
$q = 5$	0.050	0.010	1.0×10^{-12}	0.050	0.010	9.5×10^{-13}	0.051	0.010	1.1×10^{-14}
$q = 10$	0.049	0.010	1.4×10^{-14}	0.049	0.009	1.4×10^{-14}	0.051	0.010	7.8×10^{-17}

	$p_i = \{3,4,4,5\}$								
	$n = 18$			$n = 26$			$n = 66$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.050	0.010	1.9×10^{-9}	0.049	0.010	5.3×10^{-10}	0.050	0.010	2.1×10^{-12}
$q = 5$	0.050	0.010	8.3×10^{-13}	0.050	0.010	5.2×10^{-13}	0.049	0.010	1.9×10^{-15}
$q = 10$	0.050	0.010	4.2×10^{-14}	0.050	0.010	3.3×10^{-14}	0.049	0.010	3.1×10^{-16}

Table 9

Simulated error levels for different scenarios in the complex case

	$p_i = \{7,6\}$								
	$n = 15$			$n = 23$			$n = 63$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.054	0.011	5.1×10^{-10}	0.054	0.010	1.4×10^{-10}	0.051	0.011	3.9×10^{-13}
$q = 5$	0.056	0.011	1.7×10^{-12}	0.054	0.011	8.6×10^{-13}	0.052	0.010	4.7×10^{-15}
$q = 10$	0.058	0.012	4.1×10^{-14}	0.059	0.012	2.2×10^{-14}	0.053	0.010	8.8×10^{-17}
	$p_i = \{6,4,9\}$								
	$n = 21$			$n = 29$			$n = 69$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.052	0.011	1.5×10^{-10}	0.052	0.010	9.1×10^{-11}	0.051	0.010	7.8×10^{-13}
$q = 5$	0.054	0.011	4.6×10^{-13}	0.053	0.010	4.7×10^{-13}	0.052	0.010	7.6×10^{-15}
$q = 10$	0.055	0.011	1.2×10^{-14}	0.055	0.011	1.4×10^{-14}	0.052	0.011	2.3×10^{-16}
	$p_i = \{3,4,4,5\}$								
	$n = 18$			$n = 26$			$n = 66$		
	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ	$\alpha = 0.05$	$\alpha = 0.01$	Δ
$q = 2$	0.053	0.011	3.3×10^{-10}	0.052	0.011	1.2×10^{-10}	0.051	0.011	6.2×10^{-13}
$q = 5$	0.054	0.011	5.5×10^{-13}	0.054	0.011	4.0×10^{-13}	0.052	0.011	3.1×10^{-15}
$q = 10$	0.055	0.011	3.1×10^{-14}	0.057	0.012	2.6×10^{-14}	0.053	0.010	3.1×10^{-16}

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}. \tag{29}$$

The same authors (Conradsen et al., 2003) state:

> The polarimetric SAR measures the amplitude and phase of backscattered signals in four combinations of the linear receive and transmit polarizations: horizontal-horizontal (HH), horizontal-vertical (HV), vertical-horizontal (VH), and vertical-vertical (VV). These signals form the complex scattering matrix that relates the incident and the scattered electric fields (Zyl & Ulaby, 1990). The inherent speckle in the SAR data can be reduced by spatial averaging at the expense of loss of spatial resolution. In this so-called multilook case, a more appropriate representation of the backscattered signal is the covariance matrix in which the average properties of a group of resolution cells can be expressed in a single matrix.

This covariance matrix, which the authors call "average covariance matrix," is defined as

$$C = \begin{pmatrix} < S_{hh}S_{hh}^{*} > & < S_{hh}S_{hv}^{*} > & < S_{hh}S_{vv}^{*} > \\ < S_{hv}S_{hh}^{*} > & < S_{hv}S_{hv}^{*} > & < S_{hv}S_{vv}^{*} > \\ < S_{vv}S_{hh}^{*} > & < S_{vv}S_{hv}^{*} > & < S_{vv}S_{vv}^{*} > \end{pmatrix},$$

where $< \cdot >$ denotes the inner product and "*" represents the complex conjugate. "S_{rt} is the complex scattering amplitude for receive polarization r and transmit polarization t (r and t are either h for horizontal or v for vertical). Reciprocity, which normally applies to natural targets, gives (in the backscattering direction using the backscattering alignment convention (Zyl and Ulaby, 1990)) and results in the covariance matrix C with rank 3." Yet according to the same authors, who refer Goodman (1963), "C follows a complex Wishart distribution."

Conradsen et al. (2003) use then the likelihood ratio test for the equality of two covariance matrices, to test the equality of covariance matrices corresponding to groups of pixels for two images of the same area, taken at two different times, assuming a two-block structure for the matrices.

In the absence of the original data, even after efforts to contact the authors, we decided to obtain two such matrices from synthetic data, based on a pseudo-random sample of dimension 9 from a 3-variate complex multivariate normal distribution with a Hermitian positive definite covariance matrix with the structure in Eq. (29). The reason to consider a pseudo-random sample of size 9 is because it is common to consider for each pixel the surrounding square of 3×3 pixels. The two sample covariance matrices obtained were

$$A_1 = \begin{pmatrix} 4.5778 & 3.6978 - 2.0889\mathrm{i} & 0.6133 - 1.4222\mathrm{i} \\ 3.6978 + 2.0889\mathrm{i} & 14.6578 & 2.4178 + 3.7956\mathrm{i} \\ 0.6133 + 14222\mathrm{i} & 2.4178 - 3.7956\mathrm{i} & 6.3556 \end{pmatrix}$$

and

$$A_2 = \begin{pmatrix} 10.9067 & 3.1022 - 8.0356\mathrm{i} & 1.1911 - 3.1733\mathrm{i} \\ 3.1022 + 8.0356\mathrm{i} & 18.2489 & 2.5867 + 0.0800\mathrm{i} \\ 1.1911 + 3.1733\mathrm{i} & 2.5867 - 00800\mathrm{i} & 14.8356 \end{pmatrix}.$$

In this example we have thus $n = 9$, $q = 2$ and $p = 3$, with $p_1 = 2$ and $p_2 = 1$ and a computed value of 0.000590915 for the statistic Λ_a in Eq. (6), which gives a near-exact p-value of 0.26068 which would lead to a nonrejection of the null hypothesis of equality of the two covariance matrices, as we would expect, since both pseudo sample covariance matrices were generated from the same covariance population matrix. This near-exact distribution for the likelihood ratio test statistic to test the equality of covariance matrices may be easily obtained from the near-exact distribution in Theorem 4.1, considering that in this case the characteristic function $\Phi_W(t)$ is equal to $\Phi_{W_a}(t)$ in Eq. (8).

However, Conradsen et al. (2003) enter some problems when, instead of assuming or testing the independence of the two blocks of variables, the first one with two variables and the second one with just one variable, they insert zeros in the off-diagonal blocks of their sample covariance matrices, before carrying out the test of equality. With the test we introduce in this article it would have been easy to carry out in a single step the test of equality of the two covariance matrices and the test of independence of the two blocks of variables. In this case, we obtain for the likelihood ratio statistic Λ in Eq. (7) the computed value of 0.0000565334, to which, for $m^* = 4$, corresponds a near-exact p-value of 0.30393, which would lead to a nonrejection of the null hypothesis H_0 in Eq. (2). We may note that, for $m^* = 4$ and for the values of n, p_1, p_2, and q stated earlier, we would have, according to Theorem 4.1, a near-exact distribution that would be a finite mixture of five GNIG distributions of depth $d = 4$, with shape parameters $\boldsymbol{r}_{\boldsymbol{a}}^* = \{2, 2\}, \boldsymbol{r}_{\boldsymbol{b}} = \{1\}$ and $r = 3/2$ and rate parameters $\boldsymbol{l}_{\boldsymbol{a}} = \{7/9, 2/3\}, \boldsymbol{l}_{\boldsymbol{b}} = \{5/6\}$, and $\lambda^* = 0.8985360217$ and the weights $\pi_i (i = 0, \ldots, m^*)$ defined as in section 4.

7. Conclusions

Using a decomposition of the null hypothesis of the multisample independence test into two null hypotheses we show that the exact distribution of the test statistic, when the populations have real or complex multivariate normal distributions, may be represented as the product of independent Beta random variables. From this representation and using the results in Coelho et al. (2012) and Marques et al. (2012) it is possible to derive an adequate factorization of the characteristic function of the logarithm of the test statistic which is the basis for the development of the near-exact distributions. These approximating distributions, developed for the test statistic of the multisample independence test, are based on mixtures of GNIG distributions or on mixtures of GIG distributions. The number of GNIG or GIG distributions in the mixture depends on the number of moments matched. We have verified that the greater the number of moments matched the better is the approximation. These distributions can be easily implemented and as such may be of practical use in the most different applications. Computational modules for the implementation of the GNIG and GIG distribution functions can be obtained from the webpages https://sites.google.com/site/nearexactdistributions and http://en.wikipedia.org/wiki/Generalized_integer_gamma_distribution. Results show the high accuracy of these approximations even when small sample sizes are considered. In addition the near-exact approximations also present good asymptotic properties. The test proposed shows very good power even for alternatives quite close to the null hypothesis, as shown by the simulation studies carried out in section 5.2.

Funding

This research was supported by Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) through the project PEst-OE/MAT/UI0297/2014 (Centro de Matemática e Aplicações–CMA-FCT/ UNL).

References

Anderson, T. W. 2003 *An introduction to multivariate statistical analysis*, 3rd ed. New York, NY: John Wiley & Sons.

Coelho, C. A. 1998. The generalized integer gamma distribution—A basis for distribution in multivariate statistics. *J. Multivariate Anal*., 64, 86–102.

Coelho, C. A. 2004. The generalized near-integer Gamma distribution, a basis for "near-exact" approximations to the distributions of statistics which are the product of an number of independent Beta random variables. *J. Multivariate Anal*., 89, 191–218.

Coelho, C. A., and F. J. Marques. 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, C. A., and F. J. Marques. 2010. Near-exact distributions for the independence and sphericity likelihood ratio test statistics. *J. Multivariate Anal*., 101, 583–593.

Coelho, C. A., and F. J. Marques. 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.

Coelho, C. A., B. C. Arnold, and F. J. Marques. 2010. Near-exact distributions for certain likelihood ratio test statistics. *J. Stat. Theory Pract*., 4, 711–725.

Coelho, C. A., and J. T. Mexia. 2010. *Product and ratio of generalized gamma-ratio random variables: Exact and near-exact distributions—Applications*. Saarbrücken, Germany: LAP–Lambert Academic Publishing AG & Co. KG.

Coelho, C. A., F. J. Marques, and B. C. Arnold. 2012. A general approach to the exact and near-exact distributions of the main likelihood ratio test statistics used in the complex multivariate normal setting. Preprint CMA #15/2011, CMA-FCT/UNL. http://www.cma.fct.unl.pt/sites/www.cma.fct.unl.pt/files/documentos/publicacoes/pdf_2011/CMA15-2011.pdf

Conradsen, K., A. A. Nielsen, J. Schou, and H. Skiver. 2003 A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data. *IEEE Trans. Geosci. Remote Sensing*, 41, 4–19.

Fang, C., P. R. Krishnaiah, and B. N. Nagarsenker. 1982. Asymptotic distributions of the likelihood ratio test statistics for covariance structures of the complex multivariate normal distributions, *J. Multivariate Anal*., 12, 597–611.

Geiger, D., and J. Pearl. 1993. Logical and algorithmic properties of conditional independence and graphical models. *Ann. Stat*., 21, 2001–2021.

Goodman, N. R. 1963. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Stat*., 34, 152–177.

Gupta, A. K. 1971. Distribution of Wilks' likelihood-ratio criterion in the complex case. *Ann. Inst. Stat. Math*., 23, 77–87.

Hannan, E. J. 1970. *Multiple time series*. New York, NY: Wiley.

Khatri, C. G. 1965. Classical statistical analysis based on a certain multivariate complex Gaussian distribution, *Ann. Math. Stat*., 36, 98–114.

Krishnaiah, P. R., J. C. Lee, and T. C. Chang. 1976. The distributions of the likelihood ratio statistics for tests of certain covariance structures of complex multivariate normal populations, *Biometrika*, 63, 543–549.

Marques, F. J., C. A. Coelho, and B. C. Arnold, 2011. A general near-exact distribution theory for the most common likelihood ratio test statistics used in multivariate analysis. *Test*, 20, 180–203.

Marques, F. J., and C. A. Coelho. 2012. The multi-sample independence test. *AIP Conf. Proc.*, 1479, 1129–1132.

Mathai, A. M. 1973. A few remarks about some recent articles on the exact distributions of multivariate test criteria: I. *Ann. Inst. Stat. Math.*, 25, 557–566.

Pannu, N. S., A. J. Coy, and J. Read. 2003. Application of the complex mul-tivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Crystallogr., Sect. D Biol.Crystallogr.*, 59, 1801–1808.

Pillai, K. C. S., and G. M. Jouris, 1971. Some distribution problems in the multivariate complex Gaussian case. *Ann. Math. Stat.*, 42, 517–525.

Tang, J., and A. K. Gupta. 1986. Exact distribution of certain general test statistics in multivariate analysis. *Aust. J. Stat.*, 28, 107–114.

Tricomi, F. G., and A. Erdélyi. 1951. The asymptotic expansion of a ratio of Gamma functions. *Pacific J. Math.*, 1, 133–142.

Zyl, J. J., and F. T. Ulaby. 1990. Scattering matrix representation for simple targets. In *Radar polarimetry for geoscience applications*, ed. F. T. Ulaby and C. Elachi, 17–52. Norwood, MA: Artech House.

Wooding, R. A. 1956. The multivariate distribution of complex normal variables. *Biometrika*, 43, 212–215.

Appendix: Covariance Matrix in Eq. (28) Used in the Power Studies

$$
M = \frac{1}{10}\left(\begin{array}{ccc|cccc|cccc|ccccc}
345 & 139 & 250 & 194 & 190 & 158 & 253 & 181 & 174 & 152 & 104 & 0 & 0 & 0 & 0 & 0 \\
139 & 175 & 157 & 108 & 105 & 152 & 121 & 142 & 91 & 105 & 74 & 0 & 0 & 0 & 0 & 0 \\
250 & 157 & 282 & 120 & 130 & 165 & 207 & 172 & 159 & 140 & 87 & 0 & 0 & 0 & 0 & 0 \\
\hline
194 & 108 & 120 & 293 & 188 & 88 & 144 & 95 & 98 & 85 & 82 & 0 & 0 & 0 & 0 & 0 \\
190 & 105 & 130 & 188 & 222 & 80 & 126 & 92 & 96 & 83 & 79 & 0 & 0 & 0 & 0 & 0 \\
158 & 152 & 165 & 88 & 80 & 248 & 146 & 168 & 119 & 120 & 44 & 0 & 0 & 0 & 0 & 0 \\
253 & 121 & 207 & 144 & 126 & 146 & 330 & 160 & 164 & 137 & 74 & 0 & 0 & 0 & 0 & 0 \\
\hline
181 & 142 & 172 & 95 & 92 & 168 & 160 & 264 & 145 & 145 & 80 & 0 & 0 & 0 & 0 & 0 \\
174 & 91 & 159 & 98 & 93 & 119 & 164 & 145 & 203 & 132 & 98 & 0 & 0 & 0 & 0 & 0 \\
152 & 105 & 140 & 85 & 83 & 120 & 137 & 145 & 132 & 135 & 59 & 0 & 0 & 0 & 0 & 0 \\
104 & 74 & 87 & 82 & 89 & 44 & 74 & 80 & 98 & 59 & 123 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 106 & 31 & 39 & 29 & 58 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 31 & 103 & 41 & 28 & 56 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 39 & 41 & 95 & 36 & 35 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 29 & 28 & 35 & 75 & 40 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 58 & 56 & 35 & 40 & 125
\end{array}\right)
$$