The Likelihood Ratio Test for Equality of Mean Vectors with Compound Symmetric Covariance Matrices

Carlos A. Coelho(✉)

Centro de Matemática e Aplicações – CMA/FCT-UNL,
Departamento de Matemática, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, Caparica, Portugal
cmac@fct.unl.pt

Abstract. The author derives the likelihood ratio test statistic for the equality of mean vectors when the covariance matrices are assumed to have a compound symmetric structure. Its exact distribution is then expressed in terms of a product of independent Beta random variables and it is shown that for some particular cases it is possible to obtain very manageable finite form expressions for the probability density and cumulative distribution functions for this distribution. For the other cases, given the intractability of the expressions for the exact distribution, very sharp near-exact distributions are developed. Numerical studies show the extreme good performance of these near-exact distributions.

Keywords: Beta distributions · Exact distribution · Likelihood ratio statistic · Near-exact distributions

1 Introduction

The likelihood ratio test for the equality of mean vectors, when the covariance matrices are assumed to be just positive-definite, but otherwise unstructured, is a well-known test in Multivariate Analysis, and the distribution of the associated test statistic has been extensively studied ([8, Chap. 9], [1, Chap. 8], [10, Chap. 10], [5,9]).

However, a similar test for cases where some common given structure is assumed for the covariance matrices is not available in the literature.

We say that a $p \times p$ positive-definite covariance matrix Σ_{CS} is compound symmetric if, for $-\frac{a}{p-1} < b < a$, it can be written as

$$\Sigma_{CS} = \begin{bmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ b & b & a & \dots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \dots & a \end{bmatrix} = (a-b)I_p + bE_{pp} = aI_p + b(E_{pp} - I_p)\,,$$

O. Gervasi et al. (Eds.): ICCSA 2017, Part V, LNCS 10408, pp. 20–32, 2017.
DOI: 10.1007/978-3-319-62404-4_2

where I_p represents the identity matrix of order p and E_{pp} a $p \times p$ matrix of 1's.

Let us suppose that $\underline{X}_k \sim N_p(\underline{\mu}_k, \Sigma_k)$, $k = 1, \ldots, q$, where Σ_k are assumed to be equal and compound symmetric. Let us further suppose that we have a sample of size $n_k > p$ from $\underline{X}_k$ $(k = 1, \ldots, q)$ and that these q samples are independent, with $n = \sum_{k=1}^{q} n_k$. We will be interested in the test to the null hypothesis of equality of the q mean vectors $\underline{\mu}_k$ $(k = 1, \ldots, q)$, that is, the null hypothesis

$$\begin{array}{l} H_0 : \underline{\mu}_1 = \ldots \underline{\mu}_q \\ \qquad \text{assuming } \Sigma_1 = \cdots = \Sigma_q (= \Sigma_{CS} \text{ non-specified}), \end{array} \tag{1}$$

where Σ_{CS} represents a compound symmetric matrix of order p.

The interest in this test comes from the fact that such covariance structure is quite common or quite commonly assumed for covariance matrices in many situations and for many statistical models. See for example [6] for references. Moreover, in case the assumption of such structure for the covariance matrices is correct, then not accounting for it when carrying out the tests for the mean vectors will lead to losses in power. Therefore it is of interest to investigate the test for equality of mean vectors when assuming this structure for the covariance matrices.

2 The Likelihood Ratio Test Statistic

The $(2/n)$-th power of the likelihood ratio test (l.r.t.) statistic to test the null hypothesis in (1) is

$$\Lambda = \frac{a_{11}^{**}(a^{**})^{p-1}}{c_{11}^{**}(c^{**})^{p-1}}, \tag{2}$$

where

$$a^{**} = \frac{1}{p-1}\sum_{j=2}^{p} a_{jj}^{**} \qquad \text{and} \qquad c^{**} = \frac{1}{p-1}\sum_{j=2}^{p} c_{jj}^{**}, \tag{3}$$

with a_{jj}^{**} and c_{jj}^{**}, the diagonal elements of the matrices

$$A^{**} = UAU' \qquad \text{and} \qquad C^{**} = U(A+B)U' \tag{4}$$

where U is a Helmert matrix of order p, which is a $p \times p$ orthogonal matrix with first row equal to $\frac{1}{\sqrt{p}} E_{1p}$ and i-th row $(i = 2, \ldots, p)$ equal to

$$\frac{1}{\sqrt{(i-1)i}}\Big[\underbrace{1, \ldots, 1}_{i-1}, -(i-1), \underbrace{0, \ldots, 0}_{p-i}\Big] \qquad (i = 2, \ldots, p).$$

We may obtain the l.r.t. statistic in (2) by establishing a parallel with the l.r.t. statistic used when the covariance matrices Σ_k are assumed to be just positive-definite, which is the statistic

$$\Lambda^* = \frac{|A|}{|A+B|} \tag{5}$$

where

$$A=\sum_{k=1}^{q}(n_k-1)S_k \qquad \text{and} \qquad B=\sum_{k=1}^{q}n_k(\underline{\overline{X}}_k-\underline{\overline{X}})(\underline{\overline{X}}_k-\underline{\overline{X}})' \tag{6}$$

where S_k and $\underline{\overline{X}}_k$ are respectively the sample covariance matrix and mean vector of the k-th sample and

$$\underline{\overline{X}}=\frac{1}{n}\sum_{k=1}^{q}n_k\,\underline{\overline{X}}_k\,,$$

and by seeing that the l.r.t. statistic to test H_0 in (1) will be in its beginning similar to the l.r.t. statistic in (5), to which we have to add the fact that now the matrices Σ_k are assumed to be compound symmetric.

If we take into account that all the diagonal elements of Σ_{CS} are equal among themselves and that also all the off-diagonal elements of Σ_{CS} are also equal among themselves, the maximum likelihood estimator (m.l.e.) of Σ_{CS}, under the alternative hypothesis

$$\begin{array}{l} H_1:\exists j,j'\in\{1,\dots,q\}:\underline{\mu}_j\neq\underline{\mu}_{j'}\,, \\ \qquad\quad \text{assuming } \Sigma_1=\cdots=\Sigma_q(=\Sigma_{CS}\text{ non-specified})\,, \end{array} \tag{7}$$

where $\Sigma_{CS}=aI_p+b(E_{pp}-I_p)$, is $A^*=[a^*_{ij}]$, with

$$a^*_{ii}=\widehat{a}\,\big|_{H_1}=\frac{1}{p}\sum_{j=1}^{p}a_{jj} \qquad \text{and} \qquad a^*_{ij}=\widehat{b}\,\big|_{H_1}=\frac{1}{p(p-1)}\sum_{\substack{i=1\\ i\neq j}}^{p}\sum_{j=1}^{p}a_{ij} \quad (i\neq j)\,,$$

while the m.l.e. of Σ_{CS} under the null hypothesis in (1) is $C^*=[c^*_{ij}]$, with

$$c^*_{ii}=\widehat{a}\,\big|_{H_0}=\frac{1}{p}\sum_{j=1}^{p}\big(a_{jj}+b_{jj}\big) \quad \text{and} \quad c^*_{ij}=\widehat{b}\,\big|_{H_0}=\frac{1}{p(p-1)}\sum_{\substack{i=1\\ i\neq j}}^{p}\sum_{j=1}^{p}\big(a_{ij}+b_{ij}\big) \;\; (i\neq j),$$

where a_{ij} and b_{ij} represent the running elements of the matrices A and B in (5) and (6). Then, the l.r.t. statistic to test H_0 in (1) may be written as

$$\Lambda=\frac{|A^*|}{|C^*|} \tag{8}$$

where

$$|A^*|=a^{**}_{11}\left(\frac{1}{p-1}\sum_{j=2}^{p}a^{**}_{jj}\right)^{p-1} \qquad \text{and} \qquad |C^*|=c^{**}_{11}\left(\frac{1}{p-1}\sum_{j=2}^{p}c^{**}_{jj}\right)^{p-1} \tag{9}$$

where, as in (2) and (3), a^{**}_{jj} and c^{**}_{jj} represent respectively the j-th diagonal element of the matrices A^{**} and C^{**} in (4), so that Λ in (8) may be written as in (2).

We may note that, while a^{**}_{11} and c^{**}_{11} are the m.l.e.'s of $a+(p-1)b$, respectively under H_1 in (7) and under H_0 in (1), a^{**} and c^{**} in (2) and (3) are the m.l.e.'s of $a-b$, respectively under H_1 in (7) and under H_0 in (1).

3 Characterization of the Distribution of the L.R.T. Statistic

In order to obtain the distribution of the l.r.t. statistic Λ in (8) or (2), under the null hypothesis (1), now one only has to notice that, under this null hypothesis,

$$U\Sigma_{CS}U' = \Delta = diag\big(a+(p-1)b, \underbrace{a-b,\dots,a-b}_{p-1}\big)\,,$$

so that, since A and B are independent, with

$$A \sim W_p(n-q,\Sigma_{CS}) \qquad \text{and} \qquad B \sim W_p(q-1,\Sigma_{CS})\,,$$

A^{**} and $B^{**} = UBU'$ are independent, with

$$A^{**} \sim W_p(n-q,\Delta) \qquad \text{and} \qquad B^{**} \sim W_p(q-1,\Delta)\,,$$

and as such,

$$C^{**} = A^{**} + B^{**} \sim W_p(n-1,\Delta)\,.$$

The diagonal elements of A^{**} are thus independent, as well as the diagonal elements of B^{**} and C^{**}, with

$$\frac{a_{jj}^{**}}{\delta_j} \sim \chi^2_{n-q}\,, \qquad \frac{b_{jj}^{**}}{\delta_j} \sim \chi^2_{q-1}\,, \qquad \frac{c_{jj}^{**}}{\delta_j} = \frac{a_{jj}^{**}}{\delta_j} + \frac{b_{jj}^{**}}{\delta_j} \sim \chi^2_{n-1}\,, \qquad (j=1,\dots,p) \tag{10}$$

for $\delta_1 = a+(p-1)b$ and $\delta_2 = \dots = \delta_p = a-b$.

Therefore, from (8), (9) and (10), we see that

$$\Lambda \stackrel{d}{\equiv} Y_1\,(Y_2)^{p-1}$$

where Y_1 and Y_2 are independent, with

$$Y_1 \sim Beta\left(\frac{n-q}{2},\frac{q-1}{2}\right) \qquad \text{and} \qquad Y_2 \sim Beta\left(\frac{(n-q)(p-1)}{2},\frac{(q-1)(p-1)}{2}\right).$$

The h-th moment of Λ in (2) may thus be written as

$$E(\Lambda^h) = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-q}{2}+h\right)}{\Gamma\left(\frac{n-q}{2}\right)\,\Gamma\left(\frac{n-1}{2}+h\right)}\,\frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-q)(p-1)}{2}+(p-1)h\right)}{\Gamma\left(\frac{(n-q)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}+(p-1)h\right)}\,, \tag{11}$$

which will be used in deriving an exact finite form for the distribution of Λ for odd q and to obtain sharp near-exact approximations for its distribution in the other cases.

4 The Exact Distribution of Λ for Odd q

For odd q, by using the relation

$$\frac{\Gamma(a+n)}{\Gamma(a)} = \prod_{j=0}^{n-1}(a+k) \tag{12}$$

and given that the expression in (11) remains valid for any complex h, one may write the characteristic function (c.f.) of $W = -\log \Lambda$ as

$$\begin{aligned}
\Phi_W(t) &= E\left(e^{\mathrm{i}tW}\right) = E(\Lambda^{-\mathrm{i}t}) \\
&= \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-q}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-q)(p-1)}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-q)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-(p-1)\mathrm{i}t\right)} \\
&= \left\{\prod_{j=0}^{\frac{q-1}{2}-1}\left(\frac{n-q}{2}+j\right)\left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1}\right\} \\
&\qquad \times \left\{\prod_{j=0}^{\frac{(q-1)(p-1)}{2}-1}\left(\frac{(n-q)(p-1)}{2}+j\right)\left(\frac{(n-q)(p-1)}{2}+j-(p-1)\mathrm{i}t\right)^{-1}\right\} \\
&= \left\{\prod_{j=0}^{\frac{q-1}{2}-1}\left(\frac{n-q}{2}+j\right)\left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1}\right\} \\
&\qquad \times \left\{\prod_{j=0}^{\frac{(q-1)(p-1)}{2}-1}\left(\frac{n-q}{2}+\frac{j}{p-1}\right)\left(\frac{n-q}{2}+\frac{j}{p-1}-\mathrm{i}t\right)^{-1}\right\} \\
&= \prod_{j=1}^{(q-1)(p-1)/2}\left(\frac{n-q}{2}+\frac{j-1}{p-1}\right)^{r_j}\left(\frac{n-q}{2}+\frac{j-1}{p-1}-\mathrm{i}t\right)^{r_j}
\end{aligned}$$

for

$$r_j = \begin{cases} 1\,, & j=1,\ldots,(q-1)(p-1)/2 \\ & j \neq (j-1)(p-1)+1 \ \text{ for } \ell=1,\ldots,(q-1)/2 \\ 2\,, & j=(\ell-1)(p-1)+1 \ \text{ for } \ell=1,\ldots,(q-1)/2\,, \end{cases} \tag{13}$$

which shows that for odd q the exact distribution of W is a GIG (Generalized Integer Gamma) distribution [3] and that of Λ an EGIG (Exponentiated Generalized Integer Gamma) distribution [2] of depth $(q-1)(p-1)/2$ with shape parameters r_j and rate parameters $\frac{n-q}{2}+\frac{j-1}{p-1}$ $(j=1,\ldots,(q-1)(p-1)/2)$.

The exact probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of Λ are thus given by

$$f_\Lambda(z) = f^{EGIG}\left(z \,\middle|\, \{r_j\}_{j=1,\ldots,g}\,; \left\{\frac{(n-q)}{2}+\frac{j-1}{p-1}\right\}_{j=1,\ldots,g}; g\right)$$

and

$$F_\Lambda(z) = F^{EGIG}\left(z \,\middle|\, \{r_j\}_{j=1,\ldots,g}\,; \left\{\frac{(n-q)}{2}+\frac{j-1}{p-1}\right\}_{j=1,\ldots,g}; g\right)$$

for $g=(q-1)(p-1)/2$ and r_j given by (13). See [2] for the notation used for the p.d.f. and c.d.f. of the EGIG distribution.

5 Near-Exact Distributions of Λ for Even q

5.1 Near-Exact Distributions for Even q and Even p

When q is even, the exact p.d.f. and c.d.f. of Λ are not manageable. As such in this case we will develop near-exact distributions for Λ, which are asymptotic distributions that, opposite to common asymptotic distributions, will be asymptotic not only for increasing sample sizes but also for increasing values of p and q. The case where q and p are both even is somewhat more complicated and this is the one we will address in this subsection.

From (11), for even q and even p we may write the c.f. of $W=-\log\Lambda$ as

$$
\begin{aligned}
\Phi_W(t) &= \frac{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-q}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}\right)\Gamma\left(\frac{(n-q)(p-1)}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-q)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}-(p-1)\mathrm{i}t\right)} \\
&\quad\times \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-(p-1)\mathrm{i}t\right)} \\
&= \left\{\prod_{j=0}^{\frac{q-2}{2}-1}\left(\frac{n-q}{2}+j\right)\left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1}\right\} \\
&\quad\times\left\{\prod_{j=0}^{\frac{q(p-1)-p}{2}-1}\left(\frac{(n-q)(p-1)}{2}+j\right)\left(\frac{(n-q)(p-1)}{2}+j-(p-1)\mathrm{i}t\right)^{-1}\right\} \\
&\quad\times \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-(p-1)\mathrm{i}t\right)} \\
&= \left\{\prod_{j=0}^{\frac{q-2}{2}-1}\left(\frac{n-q}{2}+j\right)\left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1}\right\} \\
&\quad\times\left\{\prod_{j=0}^{\frac{q(p-1)-p}{2}-1}\left(\frac{n-q}{2}+\frac{j}{p-1}\right)\left(\frac{n-q}{2}+\frac{j}{p-1}-\mathrm{i}t\right)^{-1}\right\} \\
&\quad\times \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-1)(p-1)}{2}-\frac{1}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-(p-1)\mathrm{i}t\right)}
\end{aligned}
$$

$$= \underbrace{\left\{ \prod_{j=1}^{((q-1)(p-1)-1)/2} \left(\frac{n-q}{2} + \frac{j-1}{p-1} \right)^{r_j^*} \left(\frac{n-q}{2} + \frac{j-1}{p-1} - \mathrm{i}t \right)^{-r_j^*} \right\}}_{\Phi_{W,1}(t)}$$
$$\times \underbrace{\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2} - \mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2} - \mathrm{i}t\right)} \frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2} - \frac{1}{2} - (p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-1)(p-1)}{2} - \frac{1}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2} - (p-1)\mathrm{i}t\right)}}_{\Phi_{W,2}(t)} \tag{14}$$

with

$$r_j^* = \begin{cases} 1\,, & j = 1, \ldots, ((q-1)(p-1)-1)/2 \\ & j \neq (\ell-1)(p-1)+1 \;\; \text{for } \ell = 1, \ldots, (q-2)/2 \\ 2\,, & j = (\ell-1)(p-1)+1 \;\; \text{for } \ell = 1, \ldots, (q-2)/2\,. \end{cases} \tag{15}$$

In (14) $\Phi_{W,1}(t)$ is the c.f. of a GIG distribution of depth $((q-1)(p-1)-1)/2$ with shape parameters r_j^* and rate parameters $\frac{n-q}{2} + \frac{j-1}{p-1}$ $(j = 1, ..., ((q-1)(p-1)-1)/2)$ and will be left untouched, while, based on the results in Sect. 5 of [11], which show that we can, for increasing values of a, approximate asymptotically a $Logbeta(a, b)$ distribution by an infinite mixture of $Gamma\,(b+k, a)$ $(k = 0, 1, \ldots)$ distributions, using a somewhat heuristic approach, we will asymptotically approximate $\Phi_{W,2}(t)$, which is the c.f. of a sum of two independent random variables whose exponential has a Beta distribution with a second parameter equal to 1/2, by

$$\Phi_2^*(t) = \sum_{k=0}^{m^*} \pi_k\, (\lambda^*)^{1+k} (\lambda^* - \mathrm{i}t)^{-(1+k)}$$

which is the c.f. of a finite mixture of Gamma distributions, where λ^* is the rate parameter in

$$\Phi^*(t) = \theta(\lambda^*)^{r_1}(\lambda^* - \mathrm{i}t)^{-r_1} + (1-\theta)(\lambda^*)^{r_2}(\lambda^* - \mathrm{i}t)^{-r_2}$$

which will be numerically computed together with θ, r_1 and r_2 in such a way that

$$\left.\frac{\partial^h}{\partial t^h}\Phi^*(t)\right|_{t=0} = \left.\frac{\partial^h}{\partial t^h}\Phi_{W,2}(t)\right|_{t=0}, \qquad h = 1, \ldots, 4\,.$$

The weights π_k, $k = 0, \ldots, m^*-1$, will then be computed in such a way that

$$\left.\frac{\partial^h}{\partial t^h}\Phi_2^*(t)\right|_{t=0} = \left.\frac{\partial^h}{\partial t^h}\Phi_{W,2}(t)\right|_{t=0}, \qquad h = 1, \ldots, m^*\,,$$

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

By doing this we obtain

$$\Phi_W^*(t) = \Phi_{W,1}(t)\, \Phi_2^*(t)$$

as a near-exact c.f. for W, which will yield as near-exact distributions for W mixtures with m^*+1 components, each of which is a GIG distribution of depth $((q-$

$1)(p-1)-1)/2+1$, with shape parameters r_j^* $(j=1,\ldots,((q-1)(p-1)-1)/2)$ and a last one equal to $k+1$, and corresponding rate parameters $(n-q)/2$ $+(j-1)/(p-1)$ and λ^*.

This gives near-exact distributions for Λ with p.d.f.

$$f_\Lambda(z)=\sum_{k=0}^{m^*}\pi_k\, f^{EGIG}\left(z\,\middle|\,\{r_j^*\}_{j=1,\ldots,g},k+1;\left\{\frac{n-q}{2}+\frac{j-1}{p-1}\right\}_{j=1,\ldots,g},\lambda^*;\,g+1\right)$$

and c.d.f.

$$F_\Lambda(z)=\sum_{k=0}^{m^*}\pi_k\left(1-F^{EGIG}\left(z\,\middle|\,\{r_j^*\}_{j=1,\ldots,g},k+1;\left\{\frac{n-q}{2}+\frac{j-1}{p-1}\right\}_{j=1,\ldots,g},\lambda^*;\,g+1\right)\right),$$

for $g=((q-1)(p-1)-1)/2$ and r_j^* given by (15).

This yields very manageable near-exact distributions for both W and Λ, which will match the first m^* exact moments of W and which, as it is shown in the next section, lie very close to the exact distribution and are asymptotic not only for increasing sample sizes but also for increasing values of p and q, that is, the number of variables in each vector $\underline{X}_k$ and the number of populations considered, opposite to the common asymptotic distributions which are asymptotic for increasing sample sizes, but which quickly degrade their performance for increasing values of p.

5.2 Near-Exact Distributions for Even q and Odd p

Indeed when q is even but p is odd, we may treat the c.f. of W still in a similar manner, but now taking advantage of the fact that in this case the c.f. of the negative logarithm of the r.v. Y_2 may be expressed as the c.f. of a GIG distribution. Since in this case $p-1$ is even, we may write

$$\begin{aligned}\Phi_W(t)&=\frac{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-q}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-q}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}\,\frac{\Gamma\left(\frac{(n-1)(p-1)}{2}\right)\Gamma\left(\frac{(n-q)(p-1)}{2}-(p-1)\mathrm{i}t\right)}{\Gamma\left(\frac{(n-q)(p-1)}{2}\right)\Gamma\left(\frac{(n-1)(p-1)}{2}-(p-1)\mathrm{i}t\right)}\\&\quad\times\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}\\&=\left\{\prod_{j=0}^{\frac{q-2}{2}-1}\left(\frac{n-q}{2}+j\right)\left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1}\right\}\\&\quad\times\left\{\prod_{j=0}^{\frac{(q-1)(p-1)}{2}}\left(\frac{(n-q)(p-1)}{2}+j\right)\left(\frac{(n-q)(p-1)}{2}+j-(p-1)\mathrm{i}t\right)^{-1}\right\}\\&\quad\times\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}\end{aligned}$$

$$
\begin{aligned}
&= \left\{ \prod_{j=0}^{\frac{q-2}{2}-1} \left(\frac{n-q}{2}+j\right) \left(\frac{n-q}{2}+j-\mathrm{i}t\right)^{-1} \right\} \\
&\quad \times \left\{ \prod_{j=0}^{\frac{(q-1)(p-1)}{2}} \left(\frac{n-q}{2}+\frac{j}{p-1}\right) \left(\frac{n-q}{2}+\frac{j}{p-1}-\mathrm{i}t\right)^{-1} \right\} \\
&\quad \times \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)} \\
&= \underbrace{\left\{ \prod_{j=1}^{(q-1)(p-1)/2} \left(\frac{n-q}{2}+\frac{j-1}{p-1}\right)^{r_j^*} \left(\frac{n-q}{2}+\frac{j-1}{p-1}-\mathrm{i}t\right)^{-r_j^*} \right\}}_{\Phi_{W,1}(t)} \\
&\quad \times \underbrace{\frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}-\mathrm{i}t\right)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-1}{2}-\mathrm{i}t\right)}}_{\Phi_{W,2}(t)},
\end{aligned}
\tag{16}
$$

with the r_j^* defined in a similar manner to that in (15), now going through $(q-1)(p-1)/2$, that is, with

$$
r_j^* = \begin{cases} 1\,, & j = 1, \ldots, (q-1)(p-1)/2 \\ & j \neq (\ell-1)(p-1)+1 \ \text{ for } \ell = 1, \ldots, (q-2)/2 \\ 2\,, & j = (\ell-1)(p-1)+1 \ \text{ for } \ell = 1, \ldots, (q-2)/2\,. \end{cases} \tag{17}
$$

In (16) $\Phi_{W,1}(t)$ is now the c.f. of a GIG distribution of depth $(q-1)(p-1)/2$ with shape parameters r_j^* and rate parameters $\frac{n-q}{2}+\frac{j-1}{p-1}$ $(j = 1, ..., (q-1)(p-1)/2)$ and, similarly to what we did before, will be left untouched. Once again based on the results in Sect. 5 of [11], we will asymptotically approximate $\Phi_{W,2}(t)$ which is the c.f. of a random variable whose exponential has a Beta distribution with a second parameter equal to 1/2, by

$$
\Phi_2^*(t) = \sum_{k=0}^{m^*} \pi_k \,(\lambda^*)^{k+1/2} (\lambda^* - \mathrm{i}t)^{-(k+1/2)}
$$

which is the c.f. of a finite mixture of *Gamma* $(k+1/2, \lambda^*)$ distributions, where λ^* may be either taken as $(n-2)/2$, or alternatively, be computed in a similar manner to that used for even q and even p. This second choice would, mainly for values of p higher than 3, give near-exact distributions that will lie a little bit closer to the exact distribution.

By proceeding in this way we obtain

$$
\Phi_W^*(t) = \Phi_{W,1}(t)\, \Phi_2^*(t)
$$

as a near-exact c.f. for W, which will now yield as near-exact distributions for W mixtures with m^*+1 components, each of which is a GNIG (Generalized Near-Integer Gamma) distribution of depth $(q-1)(p-1)/2+1$, with shape parameters r_j^* $(j = 1, \ldots, (q-1)(p-1)/2)$ and a last one equal to $k + 1/2$, and corresponding rate parameters $(n-q)/2 + (j-1)/(p-1)$ and λ^*. See [4] and [9, Appendix B] for the GNIG distribution and the expressions for its p.d.f. and c.d.f..

This gives near-exact distributions for Λ with p.d.f.

$$f_\Lambda(z) = \sum_{k=0}^{m^*} \pi_k f^{GNIG}\left(\log z \,\middle|\, \{r_j^*\}_{j=1,\ldots,g}, k+\tfrac{1}{2}; \left\{\frac{n-q}{2} + \frac{j-1}{p-1}\right\}_{j=1,\ldots,g}, \lambda^*; g+1\right) \frac{1}{z}$$

and c.d.f.

$$F_\Lambda(z) = \sum_{k=0}^{m^*} \pi_k \left(1 - F^{GNIG}\left(\log z \middle| \{r_j^*\}_{j=1,\ldots,g}, k+\tfrac{1}{2}; \left\{\frac{n-q}{2} + \frac{j-1}{p-1}\right\}_{j=1,\ldots,g}, \lambda^*; g+1\right)\right),$$

for $g = (q-1)(p-1)/2$ and r_j^* given by (17).

6 Numerical Studies

In order to evaluate the proximity of the near-exact distributions to the exact distribution we will use the measure

$$\Delta = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right| dt \tag{18}$$

Table 1. Values of the measure Δ in (18) for increasing values of p and increasing sample sizes $n = pq + \{2, 102, 502\}$, for near-exact distributions which match m^* exact moments

p	q	n	m^*			
			2	4	10	20
4	6	26	3.70×10^{-8}	1.05×10^{-11}	3.47×10^{-20}	6.24×10^{-28}
		126	3.29×10^{-10}	4.05×10^{-15}	1.11×10^{-26}	2.76×10^{-42}
		526	4.50×10^{-12}	3.14×10^{-18}	1.64×10^{-33}	5.32×10^{-55}
10	6	62	1.06×10^{-9}	8.14×10^{-15}	4.89×10^{-25}	7.13×10^{-39}
		162	5.98×10^{-11}	5.90×10^{-17}	1.22×10^{-29}	4.87×10^{-47}
		562	1.44×10^{-12}	1.09×10^{-19}	1.27×10^{-35}	2.19×10^{-58}
30	6	182	9.06×10^{-12}	2.98×10^{-19}	1.27×10^{-32}	3.05×10^{-52}
		282	2.44×10^{-12}	2.30×10^{-20}	9.77×10^{-35}	3.07×10^{-56}
		682	1.73×10^{-13}	1.42×10^{-22}	5.54×10^{-39}	2.42×10^{-64}

with

$$\Delta \geq \max_{w>0} |F_W(w) - F_W^*(w)| \qquad \text{and} \qquad \Delta \geq \max_{0<z<1} |F_{\Lambda^*}(z) - F_{\Lambda^*}^*(z)| \,, \quad (19)$$

and where $\Phi_W(t)$ and $\Phi_W^*(t)$ represent respectively the exact and the near-exact characteristic functions of W and $F_W(\cdot)$ and $F_W^*(\cdot)$ the corresponding cumulative distribution functions. For the derivation of the measure Δ in (18) and the relation in (19) see [7, Appendix A].

In Tables 1 and 2 we may analyze the values of the measure Δ in (18) for different even values of p and q and different sample sizes, and in Tables 3 and 4 the values of this measure for different odd values of p, with smaller values of Δ

Table 2. Values of the measure Δ in (18) for increasing values of q and increasing sample sizes $n = pq + \{2, 102, 502\}$, for near-exact distributions which match m^* exact moments

p	q	n	m^*: 2	4	10	20
4	6	26	3.70×10^{-8}	1.05×10^{-11}	3.47×10^{-20}	6.24×10^{-28}
		126	3.29×10^{-10}	4.05×10^{-15}	1.11×10^{-26}	2.76×10^{-42}
		526	4.50×10^{-12}	3.14×10^{-18}	1.64×10^{-33}	5.32×10^{-55}
4	16	66	3.55×10^{-10}	5.37×10^{-15}	4.01×10^{-26}	3.42×10^{-40}
		166	2.63×10^{-11}	6.93×10^{-17}	3.05×10^{-30}	3.12×10^{-48}
		566	7.08×10^{-13}	1.66×10^{-19}	5.14×10^{-36}	3.95×10^{-59}
4	36	146	8.47×10^{-12}	1.10×10^{-17}	7.06×10^{-32}	4.45×10^{-51}
		246	2.07×10^{-12}	1.04×10^{-18}	3.92×10^{-34}	2.95×10^{-55}
		646	1.30×10^{-13}	1.02×10^{-20}	1.46×10^{-38}	1.09×10^{-63}

Table 3. Values of the measure Δ in (18) for even q, increasing odd values of p and increasing sample sizes $n = pq + \{2, 102, 502\}$, for near-exact distributions which match m^* exact moments

p	q	n	m^*: 2	4	10	20
5	6	32	1.06×10^{-8}	2.54×10^{-12}	5.90×10^{-21}	1.30×10^{-30}
		132	1.47×10^{-10}	2.07×10^{-15}	2.74×10^{-27}	3.02×10^{-43}
		532	2.22×10^{-12}	1.90×10^{-18}	5.87×10^{-34}	1.05×10^{-55}
11	6	68	3.44×10^{-10}	9.27×10^{-15}	1.03×10^{-25}	1.07×10^{-39}
		168	2.25×10^{-11}	9.86×10^{-17}	5.47×10^{-30}	7.98×10^{-48}
		568	5.76×10^{-13}	2.20×10^{-19}	8.18×10^{-36}	8.44×10^{-59}
31	6	188	3.44×10^{-12}	4.56×10^{-18}	8.69×10^{-33}	8.83×10^{-53}
		288	9.54×10^{-13}	5.37×10^{-19}	7.97×10^{-35}	1.36×10^{-56}
		688	6.97×10^{-14}	6.86×10^{-21}	5.46×10^{-39}	1.68×10^{-64}

showing a better agreement between the near-exact and the corresponding exact distribution.

We may see how all the near-exact distributions exhibit extremely low values of the measure Δ and how they display a clear asymptotic behavior not only for increasing sample sizes but also for increasing values of p, the number of variables involved, as well as for increasing values of q, the number of populations involved.

Noticeably, even for very small sample sizes the near-exact distributions exhibit extremely low values of Δ, showing their extreme closeness to the exact distribution, even for these very small sample sizes.

As expected, near-exact distributions with higher values of m^* show lower values of the measure Δ, given that m^* is the number of exact moments of W that are matched by these near-exact distributions.

Table 4. Values of the measure Δ in (18) for odd p, increasing even values of q and increasing sample sizes $n = pq + \{2, 102, 502\}$, for near-exact distributions which match m^* exact moments

p	q	n	m^*			
			2	4	10	20
5	6	32	1.06×10^{-8}	2.54×10^{-12}	5.90×10^{-21}	1.30×10^{-30}
		132	1.47×10^{-10}	2.07×10^{-15}	2.74×10^{-27}	3.02×10^{-43}
		532	2.22×10^{-12}	1.90×10^{-18}	5.87×10^{-34}	1.05×10^{-55}
5	16	82	1.00×10^{-10}	1.24×10^{-15}	1.61×10^{-27}	2.05×10^{-43}
		182	1.02×10^{-11}	2.74×10^{-17}	3.85×10^{-31}	7.90×10^{-50}
		582	3.30×10^{-13}	8.90×10^{-20}	1.28×10^{-36}	3.42×10^{-60}
31	6	188	2.40×10^{-12}	2.53×10^{-18}	2.57×10^{-33}	1.05×10^{-53}
		288	7.15×10^{-13}	3.35×10^{-19}	2.99×10^{-35}	2.45×10^{-57}
		688	5.58×10^{-14}	4.77×10^{-21}	2.54×10^{-39}	4.32×10^{-65}

7 Conclusions

The method used to analyze and factorize the characteristic function of the negative logarithm of the likelihood ratio statistic, with the help of the relation (12), proved itself extremely useful. It not only enabled us to obtain, for odd q, that is, for an odd number of the populations involved, the exact distribution of the negative logarithm of the likelihood ratio statistic as a GIG distribution, which is a very manageable distribution, but also allowed for the development of very sharp near-exact distributions for even q. As noted in Subsect. 5.2, when q is even and p is odd, it is possible to obtain even sharper near-exact distributions than those obtained in Subsect. 5.1 for even p and even q. Although, as expected, the near-exact distributions display a much better closeness to the exact distribution as the value of m^*, the number of exact moments of W matched by these near-exact distributions, increases, even for very small values of m^* the near-exact distributions display a great closeness to the exact distribution and a clear asymptotic behavior not only for increasing sample sizes but also for increasing values of the number of populations involved, and opposite to the common asymptotic distributions, also for increasing numbers of variables involved, always with very good performances for very small samples sizes, even for large values of p, the number of variables involved.

Acknowledgements. Research supported by FCT–Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), project UID/MAT/00297/2013, through Centro de Matemática e Aplicações (CMA/FCT-UNL).

References

1. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley, Hoboken (2003)
2. Arnold, B.C., Coelho, C.A., Marques, F.J.: The distribution of the product of powers of independent Uniform random variables - a simple but useful tool to address and better understand the structure of some distributions. J. Multivariate Anal. **113**, 19–36 (2013)
3. Coelho, C.A.: The generalized integer Gamma distribution - a basis for distributions in Multivariate Statistics. J. Multivariate Anal. **64**, 86–102 (1998)
4. Coelho, C.A.: The generalized near-integer Gamma distribution: a basis for ear-exactapproximations to the distribution of statistics which are the product of an odd number of independent Beta random variables. J. Multivariate Anal. **89**, 191–218 (2004)
5. Coelho, C.A., Arnold, B.C., Marques, F.J.: Near-exact distributions for certain likelihood ratio test statistics. J. Stat. Theory Pract. **4**, 711–725 (2010)
6. Coelho, C.A., Roy, A.: Testing the hypothesis of a block compound symmetric covariance matrix for elliptically contoured distributions. TEST (in print)
7. Grilo, L.M., Coelho, C.A.: The exact and near-exact distributions for the statistic used to test the reality of covariance matrix in a Complex Normal distribution. In: Bebiano, N. (ed.) Applied and Computational Matrix Analysis, pp. 295–315. Springer, New York (2017)
8. Kshirsagar, A.M.: Multivariate Analysis. Marcel Dekker, New York (1972)
9. Marques, F.J., Coelho, C.A., Arnold, B.C.: A general near-exact distribution theory for the most common likelihood ratio test statistics used in Multivariate Analysis. TEST **20**, 180–203 (2011)
10. Muirhead, R.J.: Aspects of Multivariate Statistical Theory, 2nd edn. Wiley, Hoboken (2005)
11. Tricomi, F.G., Erdélyi, A.: The asymptotic expansion of a ratio of gamma functions. Pacific J. Math. **1**, 133–142 (1951)

http://www.springer.com/978-3-319-62403-7

Computational Science and Its Applications – ICCSA 2017
17th International Conference, Trieste, Italy, July 3-6, 2017, Proceedings, Part V

Gervasi, O.; Murgante, B.; Misra, S.; Borruso, G.; Torre, C.M.; Rocha, A.M.A.C.; Taniar, D.; Apduhan, B.O.; Stankova, E.; Cuzzocrea, A. (Eds.)

2017, XXXVI, 744 p. 265 illus., Softcover
ISBN: 978-3-319-62403-7