

The block-matrix sphericity test – exact and near-exact distributions for the test statistic

Filipe J. Marques, Carlos A. Coelho and Paula Marques

Abstract In this work near-exact distributions for the likelihood ratio test (l.r.t.) statistic to test the one sample block-matrix sphericity hypothesis are developed under the assumption of multivariate normality. Using a decomposition of the null hypothesis in two null hypotheses, one for testing the independence of the k groups of variables and the other one for testing the equality of the k block diagonal matrices of the covariance matrix, we are able to derive the expressions of the l.r.t. statistic, its h -th null moment, and the characteristic function (c.f.) of its negative logarithm. The decomposition of the null hypothesis induces a factorization on the c.f. of the negative logarithm of the l.r.t. statistic that enables us to obtain near-exact distributions for the l.r.t. statistic. Numerical studies using a measure based on the exact and approximating c.f.'s are developed. This measure is an upper bound on the distance between the exact and approximating distribution functions and it is used to assess the performance of the near-exact distributions and to compare these with the Box type asymptotic approximation in [3].

1 Introduction

The one sample block-matrix sphericity test is of great interest when we wish to test, under multivariate normality, if in a sequence of p random variables (r.v.'s) X_1, \dots, X_p we have k independent groups of p^* variables and if all of the k covariance matrices are equal. We show that we can split the null hypothesis of the block-matrix test in two null hypotheses, one for testing the independence among the k

Filipe J. Marques and Carlos A. Coelho
Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa,
Quinta da Torre 2829-516, Caparica, Portugal, e-mail: fjm@fct.unl.pt and cmac@fct.unl.pt

Paula Marques
Instituto Superior Dom Dinis, Av. 1 Maio, 2430, Marinha Grande, Portugal
e-mail: paulafartaria@gmail.com

of groups of variables and the other one for testing the equality of the k covariance matrices. The exact distribution of the likelihood ratio test (l.r.t.) statistic has a very complicated expression which makes its use very difficult in practice. Therefore the development of easy to use and yet highly accurate approximations becomes a good target. Our aim is to show that, based on a the decomposition of the null hypothesis of the one sample block-matrix sphericity, we are able to derive the expressions of the l.r.t statistic and also of its h -th null moment, and the characteristic function (c.f.) of its negative logarithm. The factorization induced on the c.f. of the logarithm of the l.r.t. statistic, by the decomposition of the null hypothesis, together with the results in [7] and [6] will allow us to develop very accurate near-exact distributions for the l.r.t statistic (see also [8]). In [4] the exact null distribution of the l.r.t. statistic when $k = 2$ is obtained using the inverse Mellin transform and the Meijer G-function what renders the quantile computations too hard even for small values of p^* , reinforcing the need for good manageable approximations. In [3] the authors present an asymptotic approximation based on Box's method (see [2]) which we will use to compare with the new approximations proposed.

2 The decomposition of the test null hypothesis

Let us consider a sample of size N taken from a p -variate normal population $N_p(\underline{\mu}, \Sigma)$. We intend to test the following null hypothesis

$$H_0 : \Sigma = \begin{pmatrix} \Delta & 0 & \dots & 0 \\ 0 & \Delta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta \end{pmatrix} (= I_k \otimes \Delta), \quad (\Delta \text{ not specified}) \quad (1)$$

where Δ is of order p^* , with $p = kp^*$.

The null hypothesis in (1) may be decomposed in two null hypotheses, more precisely

$$H_0 = H_{0b|0a} \circ H_{0a} \quad (2)$$

where, for

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1k} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \dots & \Sigma_{kk} \end{pmatrix}, \quad (3)$$

we have

$$H_{0a} : \Sigma = bdiag(\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{kk}), \quad (4)$$

the null hypothesis to test the independence among the k groups of p^* variables and

$$H_{0b|0a} : \Sigma_{11} = \Sigma_{22} = \dots = \Sigma_{kk} (= \Delta) , \quad (\Delta \text{ not specified}) \quad (5)$$

assuming H_{0a} true

the null hypothesis to test the equality of the k covariance matrices of order p^* .

3 The l.r.t. statistic, λ^* , and the h -th null moment of λ^*

The expressions of the l.r.t statistics, λ_a^* and $\lambda_{b|a}^*$, to test the null hypotheses in (4) and (5) respectively are given by (see [1])

$$\lambda_a^* = \frac{|A|^{n/2}}{\prod_{j=1}^k |A_{jj}|^{n/2}} \quad \text{and} \quad \lambda_{b|a}^* = \frac{(kn)^{kn p^*/2} \prod_{j=1}^k |A_{jj}|^{n/2}}{\prod_{j=1}^k n^{p^* n/2} |A^*|^{nk/2}} \quad (6)$$

where $n = N - 1$, $A = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})'$, A_{jj} is the j -th diagonal matrix of order p^* of A and $A^* = A_{11} + \dots + A_{kk}$. Using the decomposition in (2) we may obtain the expression for the l.r.t. statistic, λ^* , to test H_0 in (1) as the product of the expressions of the l.r.t. statistics in (6) (see Lemma 10.3.1 in [1])

$$\lambda^* = \lambda_a^* \lambda_{b|a}^* = \frac{|A|^{n/2}}{\left| \frac{1}{k} \sum_{j=1}^k A_{jj} \right|^{nk/2}}. \quad (7)$$

Given the independence of the l.r.t. statistics, λ_a^* and $\lambda_{b|a}^*$, in (6), under H_0 in (1), the expression of the h -th null moment of λ^* may be obtained as the product of the expressions of the h -th null moments of λ_a^* and $\lambda_{b|a}^*$ (see [1]), therefore

$$\begin{aligned} E [(\lambda^*)^h] &= E \left[(\lambda_{b|a}^* \lambda_a^*)^h \right] = E \left[(\lambda_{b|a}^*)^h \right] \times E \left[(\lambda_a^*)^h \right] \\ &= \frac{(nk)^{nk p^* h/2}}{\prod_{j=1}^k n^{p^* n h/2}} \frac{\Gamma_{p^*} \left(\frac{nk}{2} \right)}{\Gamma_{p^*} \left(\frac{nk}{2} (1+h) \right)} \prod_{j=1}^k \frac{\Gamma_{p^*} \left(\frac{n}{2} (1+h) \right)}{\Gamma_{p^*} \left(\frac{n}{2} \right)} \end{aligned} \quad (8)$$

$$\begin{aligned} &\times \frac{\Gamma_{p^* k} \left(\frac{1}{2} n + \frac{1}{2} h n \right)}{\Gamma_{p^* k} \left(\frac{1}{2} n \right)} \prod_{i=1}^k \frac{\Gamma_{p^*} \left(\frac{n}{2} \right)}{\Gamma_{p^*} \left(\frac{n}{2} (1+h) \right)} \\ &= k^{\frac{1}{2} p^* k n h} \prod_{j=1}^{p^* k} \frac{\Gamma \left(\frac{1}{2} (n + n h - j + 1) \right)}{\Gamma \left(\frac{1}{2} (n - j + 1) \right)} \prod_{i=1}^{p^*} \frac{\Gamma \left(\frac{1}{2} (n k - i + 1) \right)}{\Gamma \left(\frac{1}{2} (n k + n k h - i + 1) \right)}. \end{aligned} \quad (9)$$

4 The c.f. of $W = -\log \lambda^*$

The expression of the c.f. of the r.v. $W = -\log \lambda^*$ may be derived from the expression of the h -th null moment of λ^* , noticing that $E(e^{itW}) = E((\lambda^*)^{-it})$

$$\Phi_W(t) = k^{-\frac{1}{2}p^*knt} \prod_{j=1}^{p^*k} \frac{\Gamma(\frac{1}{2}(n - nit - j + 1))}{\Gamma(\frac{1}{2}(n - j + 1))} \prod_{j=1}^{p^*} \frac{\Gamma(\frac{1}{2}(nk - j + 1))}{\Gamma(\frac{1}{2}(nk - nkit - j + 1))}.$$

If we consider the decomposition in (8) and (9) we may rewrite the c.f. of W as the product of the c.f.'s of $W_1 = -\log \lambda_a^*$ and $W_2 = -\log \lambda_{b|a}^*$

$$\Phi_W(t) = \underbrace{\frac{\Gamma_{p^*k}(\frac{1}{2}n - \frac{1}{2}itn)}{\Gamma_{p^*k}(\frac{1}{2}n)} \prod_{i=1}^k \frac{\Gamma_{p^*}(\frac{n}{2})}{\Gamma_{p^*}(\frac{n}{2}(1 - it))}}_{\Phi_{W_1}(t)} \quad (10)$$

$$\times \underbrace{\frac{(nk)^{-nkp^*it/2}}{\prod_{j=1}^k n^{-p^*nit/2}} \frac{\Gamma_{p^*}(\frac{nk}{2})}{\Gamma_{p^*}(\frac{nk}{2}(1 - it))} \prod_{j=1}^k \frac{\Gamma_{p^*}(\frac{n}{2}(1 - it))}{\Gamma_{p^*}(\frac{n}{2})}}_{\Phi_{W_2}(t)}. \quad (11)$$

4.1 Factorizations of the c.f.'s of $W_1 = -\log \lambda_a^*$ and $W_2 = -\log \lambda_{b|a}^*$

With the final goal of developing near-exact distributions for the l.r.t. statistic, λ^* , in (7) (see Subsection 5) we will use factorizations of the c.f.'s of $W_1 = -\log \lambda_a^*$ and $W_2 = -\log \lambda_{b|a}^*$. These factorizations, already obtained in [6] and [7], show that the exact distribution of both W_1 and W_2 may be represented in the form of the sum of two independent r.v.'s, one with a Generalized Integer Gamma (GIG) distribution (see [5]) and the other one with the distribution of the sum of independent Logbeta distributions, eventually multiplied by a constant. These similarities in the structure of the c.f.'s of W_1 and W_2 will be of great use to achieve our goal.

4.1.1 The c.f. of $W_1 = -\log \lambda_a^*$

In [6] the author shows a possible factorization for the c.f. of $W_1 = -\log \lambda_a^*$, in the form

$$\Phi_{W_1}(t) = \underbrace{\prod_{j=2}^{p-1} \left(\frac{n-j}{n}\right)^{r_j^*} \left(\frac{n-j}{n} - it\right)^{-r_j^*}}_{\Phi_{1,1}(t)} \underbrace{\left\{ \frac{\Gamma(\frac{n}{2}) \Gamma(\frac{n}{2} - \frac{1}{2} - \frac{n}{2}it)}{\Gamma(\frac{n}{2} - \frac{n}{2}it) \Gamma(\frac{n}{2} - \frac{1}{2})} \right\}^{m^*}}_{\Phi_{1,2}(t)} \quad (12)$$

with $m^* = k$ if p^* is odd and $m^* = 0$ if p^* is even; the parameters r_j^* are the parameters r_j given by expressions (A.3) and (A.4) in [9]. The c.f. $\Phi_{1,1}(t)$ is the c.f. of the sum of $p - 2$ r.v.'s with Gamma distribution, with integers shape parameters, r_j^* , and with rates parameters $\frac{n-j}{n}$ ($j = 2, \dots, p - 1$), that is, the c.f. of a GIG distribution with depth $p - 2$. The c.f. $\Phi_{1,2}(t)$ is the c.f. of the sum of m^* independent r.v.'s with Logbeta distribution multiplied by $\frac{n}{2}$ and with parameters $\frac{n}{2} - \frac{1}{2}$ and $\frac{n}{2}$.

4.1.2 The c.f. of $W_2 = -\log \lambda_{b|a}^*$

In [7] the authors derive the following factorization for the c.f. of $\Phi_{W_2}(t)$

$$\begin{aligned} \Phi_{W_2}(t) = & \underbrace{\prod_{j=1}^{p-1} \left(\frac{n-j}{n} \right)^{r_j} \left(\frac{n-j}{n} - it \right)^{-r_j}}_{\Phi_{2,1}(t)} \\ & \times \prod_{j=1}^{\lfloor p/2 \rfloor} \prod_{k=1}^q \frac{\Gamma(a_j + b_{jk})}{\Gamma(a_j + b_{jk}^*)} \frac{\Gamma(a_j + b_{jk}^* - nit)}{\Gamma(a_j + b_{jk} - nit)} \\ & \times \underbrace{\left(\prod_{k=1}^q \frac{\Gamma(a_p + b_{pk})}{\Gamma(a_p + b_{pk}^*)} \frac{\Gamma(a_p + b_{pk}^* - \frac{n}{2}it)}{\Gamma(a_p + b_{pk} - \frac{n}{2}it)} \right)^{p \perp\!\!\!\perp 2}}_{\Phi_{2,2}(t)} \end{aligned} \quad (13)$$

where $a_j, b_{jk}, b_{jk}^*, a_p, b_{pk}$ and b_{pk}^* are given in (3.1) and (3.2) in [7], r_j are given in expressions (3.3)-(3.5) also in [7] and where

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

The c.f. $\Phi_{2,1}(t)$ is the c.f. of the sum of $p - 1$ r.v.'s with Gamma distribution, with integers shape parameters, r_j , and with rates parameters $\frac{n-j}{n}$ ($j = 1, \dots, p - 1$), that is, the c.f. of a GIG distribution with depth $p - 1$. The c.f. $\Phi_{2,2}(t)$ is the c.f. of the sum of $\lfloor p/2 \rfloor \times q + q \times p \perp\!\!\!\perp 2$ independent r.v.'s with Logbeta distribution, the first $\lfloor p/2 \rfloor \times q$ ones multiplied by n and the remaining ones by $\frac{n}{2}$.

5 Near-exact distributions for W and λ^*

Using the similarities observed on the factorizations of the c.f.'s of W_1 e W_2 given in Subsection 4.1 we may now rewrite the c.f. of $W = -\log \lambda^*$.

Theorem 1. *The c.f. of $W = -\log \lambda^*$ may be represented in the form*

$$\Phi_W(t) = \underbrace{\prod_{j=1}^{p-1} \left(\frac{n-j}{n}\right)^{v_j} \left(\frac{n-j}{n} - it\right)^{-v_j}}_{\Phi_{W_1^*}(t)} \underbrace{\Phi_{1,2}(t) \times \Phi_{2,2}(t)}_{\Phi_{W_2^*}(t)} \quad (14)$$

with $\Phi_{1,2}(t)$ and $\Phi_{2,2}(t)$ given respectively in (12) and (13) and where v_j are given by

$$v_j = \begin{cases} r_j & j = 1 \\ r_j + r_j^* & j = 2, \dots, p-1, \end{cases} \quad (15)$$

with r_j^* equal to the parameters r_j given by expressions (A.3) and (A.4) in [9] and r_j given in expressions (3.3)-(3.5) in [7].

The near-exact c.f.'s will thus have the form

$$\underbrace{\Phi_{W_1^*}(t)}_{\text{GIG distribution}} \times \Phi_{ne}(t) \quad (16)$$

where $\Phi_{W_1^*}(t)$ is given by (14) and $\Phi_{ne}(t)$ is the c.f. that we will use to approximate the c.f. $\Phi_{W_2^*}(t)$ in (14). Since a Logbeta distribution is indeed an infinite mixture of Gamma distributions, we propose $\Phi_{ne}(t)$ to be the c.f. of a single Gamma r.v. or a mixture of 2 or 3 Gamma r.v.'s. The parameters in $\Phi_{ne}(t)$ are evaluated so that we can ensure that $\Phi_{W_2^*}(t)$ and $\Phi_{ne}(t)$ have the same first 2, 4 or 6 derivatives at $t = 0$, that is, we want to ensure the exact and the approximating distributions to have the same first 2, 4 or 6 moments. Thus, we will have

$$\Phi_{ne}(t) = \sum_{\ell=1}^{h/2} \theta_\ell \mu^{\delta_\ell} (\mu - it)^{-\delta_\ell} \quad (17)$$

with weights $\theta_\ell > 0$ ($\ell = 1, \dots, h/2$ with $h = 2, 4$ or 6) and $\sum_{\ell=1}^{h/2} \theta_\ell = 1$, and with

$$\left. \frac{d^j}{dt^j} \Phi_{W_2^*}(t) \right|_{t=0} = \left. \frac{d^j}{dt^j} \Phi_{ne}(t) \right|_{t=0}, \quad j = 1, \dots, h \quad (18)$$

for $h = 2, 4$ or 6 , depending on $\Phi_{ne}(t)$ being the c.f. of a single r.v. or a mixture of 2 or 3 Gamma r.v.'s with the same rate parameters. Using this approach we obtain, for $h = 2$, as near-exact distribution for W a single Generalized Near-Integer Gamma (*GNIG*) distribution (see [6]) or, for $h = 4$ or 6 , a mixture of 2 or 3 *GNIG* distributions. By simple transformation it is easy to obtain near-exact distributions for λ^* .

Theorem 2. *The near-exact distributions for λ^* are either a exponential GNIG distribution or a mixture of 2 or 3 exponential GNIG distributions of depth p and for $h = 2, 4$ or 6 with p.d.f. for λ^* given by (using the notation of Appendix B in [9])*

$$\sum_{\ell=1}^{h/2} \theta_\ell f^{GNIG} \left(-\log w | v_1, \dots, v_{p-1}, \delta_\ell; \frac{n-1}{n}, \dots, \frac{n-p+1}{n}, \mu; p \right) \frac{1}{w}$$

and c.d.f given by

$$1 - \sum_{\ell=1}^{h/2} \theta_{\ell} F^{GNIG}\left(-\log w | \mathbf{v}_1, \dots, \mathbf{v}_{p-1}, \delta_{\ell}; \frac{n-1}{n}, \dots, \frac{n-p+1}{n}, \mu; p\right)$$

with $\mathbf{v}_1, \dots, \mathbf{v}_{p-1}$ given in (15), and where, for $h = 2, 4$ or 6 , the parameters θ_{ℓ} , δ_{ℓ} and μ are obtained as the numerical solution of the system of equations in (18), with $\theta_{h/2} = 1 - \sum_{\ell=1}^{h/2-1} \theta_{\ell}$.

6 Numerical studies

To assess the closeness of the near-exact distributions to the exact distribution we will use the measure

$$\Delta = \int_{-\infty}^{+\infty} \left| \frac{\Phi_W(t) - \Phi_{app}(t)}{t} \right| dt,$$

where $\Phi_W(t)$ and $\Phi_{app}(t)$ represent respectively the exact and the approximate c.f. of the r.v. $W = -\log \lambda^*$. For further details on this measure see [7]. We will denote by *GNIG*, *M2GNIG* and *M3GNIG* the near-exact distributions corresponding, respectively, to a GNIG or to a mixture of 2 or 3 GNIG distributions and by *Box* the asymptotic approximation in [3].

Table 1 Values of Δ for the approximating distributions for $W = -\log \lambda^*$

p	p^*	k	n	<i>GNIG</i>	<i>M2GNIG</i>	<i>M3GNIG</i>	<i>Box</i>
9	3	3	11	4.5×10^{-6}	3.6×10^{-9}	4.0×10^{-12}	5.8×10^{-2}
12	3	4	14	1.2×10^{-6}	4.3×10^{-10}	2.9×10^{-13}	1.2×10^{-1}
15	3	5	17	4.3×10^{-7}	8.2×10^{-11}	3.8×10^{-14}	2.1×10^{-1}
21	3	7	23	8.3×10^{-8}	1.7×10^{-12}	9.0×10^{-16}	4.1×10^{-1}
27	3	9	29	2.5×10^{-8}	6.3×10^{-13}	5.2×10^{-17}	6.4×10^{-1}

Table 2 Values of Δ for the approximating distributions for $W = -\log \lambda^*$

p	p^*	k	n	<i>GNIG</i>	<i>M2GNIG</i>	<i>M3GNIG</i>	<i>Box</i>
8	4	2	10	3.1×10^{-6}	2.9×10^{-10}	9.1×10^{-13}	3.3×10^{-2}
8	4	2	50	2.1×10^{-7}	6.9×10^{-13}	1.8×10^{-14}	3.2×10^{-5}
8	4	2	100	5.5×10^{-8}	4.3×10^{-14}	3.2×10^{-18}	3.3×10^{-6}

From Tables 1, 2 and 3 we may observe not only the very good asymptotic properties of the near-exact distributions proposed for increasing values of n and p but also the very good results of the measure Δ for the near-exact distributions, specially when compared with the ones presented by the asymptotic approximation denoted by *Box* and given in [3].

Table 3 Values of Δ for the approximating distributions for $W = -\log \lambda^*$

p	p^*	k	n	$GNIG$	$M2GNIG$	$M3GNIG$	Box
9	3	3	11	4.5×10^{-6}	3.6×10^{-9}	4.0×10^{-12}	5.8×10^{-2}
9	3	3	50	4.6×10^{-7}	2.7×10^{-11}	1.2×10^{-14}	9.0×10^{-5}
9	3	3	100	1.2×10^{-7}	2.0×10^{-12}	1.5×10^{-15}	9.2×10^{-6}

7 Conclusions

We have shown that, based on a the decomposition of the null hypothesis of the one sample block-matrix sphericity, we may derive the expressions of the l.r.t statistic, its h -th null moment, and also of the c.f. of its logarithm. This decomposition induces a factorization on the c.f. of the l.r.t. statistic which together with the results obtained in [7] and [6] allow us to develop very accurate near-exact distributions for the l.r.t statistic and for the logarithm of the l.r.t. statistic. These near-exact distributions are very accurate approximations and reveal at the same time very good asymptotic properties.

Acknowledgements This research was partially financially supported by the Portuguese Foundation for Science and Technology, through CMACentro de Matemática e Aplicações de Universidade Nova de Lisboa (Financiamento Base 2010 ISFL-1-297 from FCT/MCTES/PT).

References

1. Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd ed., J. Wiley & Sons, New York.
2. Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317-346.
3. Chao, C. C. & Gupta, A. K. (1991). Testing of homogeneity of diagonal blocks with blockwise independence. *Commun. Stat., Theory Methods*, **20**, 1957-1969.
4. Cardeño, L. & Nagar, D. K. (2001). Testing Block Sphericity of a Covariance Matrix. *Divulgaciones Matemáticas*, **9**, 25-34.
5. Coelho, C. A. (1998). The Generalized Integer Gamma Distribution – A basis for distributions in Multivariate Statistics. *Journal of Multivariate Analysis*, **64**, 86-102.
6. 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 odd number of independent Beta random variables. *Journal of Multivariate Analysis*, **89**, 191-218.
7. Coelho CA, Marques FJ (2009) Near-exact distributions for the likelihood ratio test statistic for testing equality of several variance-covariance matrices (Revisited). *The New University of Lisbon, Mathematics Department, Technical Report 9/2009*. (<http://www.dmat.fct.unl.pt/fct/servlet/DownloadFile?file=1162&sub=investiga&modo=prepub>).
8. Coelho, C. A. & Marques, F. J. (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 and Its Applications*, **430**, 2592-2606..
9. Marques, F. J. & Coelho, C. A. (2011). A general near-exact distribution theory for the most common likelihood ratio test statistics used in Multivariate Statistics. *Test*, in print.