O teste de esfericidade para várias amostras

Como usar a decomposição da hipótese nula na construção de aproximações quase-exactas para a estatística de teste

Filipe J. Marques
Dep. de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa e Centro de Matemática e Aplicações - fjm@fct.unl.pt

Carlos A. Coelho
Dep. de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa e Centro de Matemática e Aplicações - cmac@fct.unl.pt

Resumo: Uma decomposição da hipótese nula do teste de esfericidade para várias amostras é utilizada, não só para obter de uma forma simples a estatística de razão de verosimilhanças do teste e a expressão do h-ésimo momento nulo, mas também, para através da decomposição induzida na função característica do logaritmo da estatística, obter distribuições quase-exactas para a estatística de teste que correspondam a funções densidade e distribuição que permitam, de uma forma rápida e fácil, a obtenção de quantis e p-values. A distribuição exacta da estatística de razão de verosimilhanças não tem utilização prática devido à sua estrutura complexa e os resultados fornecidos, em termos de qualidade/precisão, pelas aproximações assimptóticas disponíveis na literatura podem ser francamente melhorados se optarmos antes pela utilização de aproximações quase-exactas. Para estudar a qualidade das novas aproximações quase-exactas e da aproximação assimptótica obtida por Moschopoulos em [9] é utilizada uma medida baseada nas fórmulas de inversão.

Palavras–chave: distribuição gama quase-inteira generalizada, misturas, distribuições quase-exactas, distribuições assimptóticas.

Abstract: A decomposition of the null hypothesis of the multi-sample sphericity test is used, not only to obtain the likelihood ratio test statistic and the expression for the null moments of the test statistic but also, using the induced factorization of the characteristic function of the logarithm of the statistic, to obtain near-exact distributions for this test statistic, that correspond to density and distribution functions that allow the computation of quantiles and p-values in a easy and practical way. The exact distribution of the test statistic has a complex structure and is not possible to use in practical terms while the precision/quality of the results given by the asymptotic approximations available in the literature can be substantially improved if instead we use near-exact approximations. A measure based on the inversion formulas is used in order to assess the quality of the new near-exact approximations and to compare them with the asymptotic approximation obtained by Moschopoulos in [9]

Keywords: multi-sample sphericity test, generalized near-integer gamma distribution, mixtures, near-exact distributions, asymptotic distributions.

1 Introdução

A distribuição exacta da estatística de razão de verosimilhanças do teste de esfericidade para várias amostras tem uma expressão complexa sendo praticamente impossível de utilizar na prática. Por outro lado, como veremos, as aproximações conhecidas não revelam a necessária precisão que as torne fiáveis em termos práticos. Desta forma justifica-se a obtenção de novas aproximações, nomeadamente aproximações quase-exactas, que sejam práticas e simultaneamente revelem elevado grau de precisão.

Distribuições quase-exactas para a estatística de razão de verosimilhanças do teste de esfericidade foram obtidas em [8]. Os autores comparam as aproximações quase-exactas com aproximações baseadas no método de Box (veja-se [2]) e no método ponto-sela (veja-se [3]), nos vários estudos numéricos apresentados pode-se verificar que as aproximações quase-exactas demonstram em todos casos ser mais precisas do que as restantes aproximações. Em [6] são obtidas distribuições quase-exactas para a estatística de razão de verosimilhanças do teste de igualdade de matrizes de covariância, neste trabalho são também apresentados estudos numéricos que permitem verificar a elevada qualidade das aproximações quase-exactas desenvolvidas nesse trabalho. Nos trabalhos desenvolvidos em [8] e [6] são obtidas factorizações das funções características do logaritmo da estatísticas de razão de verosimilhanças associadas aos testes de esfericidade e de igualdade de várias matrizes de covariância que serão utilizadas para obter novas aproximações quase-exactas para a estatística de razão de verosimilhanças do teste de esfericidade para várias amostras.

2 A estatística do teste de esfericidade para várias amostras e a expressão dos momentos nulos

Consideremos q populações multivariadas normais $N_p(\underline{\mu}_j\,,\Sigma_j)$, $j=1,\dots,q$ e q amostras independentes, sendo N_j a dimensão da j-ésima amostra ($j=1,\dots,q$). Pretendemos testar a hipótese nula

$$H_0:\ \Sigma_1=\Sigma_2=\ldots=\Sigma_q=I_p\,\sigma^2 \qquad (\,\sigma^2 \text{ não especificada})\,. \tag{1}$$

Com o objectivo de obter aproximações quase-exactas vamos olhar de uma forma diferente para H_0, isto é, vamos decompor esta hipótese da seguinte forma

$$H_0=H_{0b|0a}\circ H_{0a} \tag{2}$$

onde as hipóteses nulas parciais são

$$H_{0a}:\ \Sigma_1=\Sigma_2=\ldots=\Sigma_q(=\Sigma)\quad,\quad (\,\Sigma \text{ não especificada}) \tag{3}$$

e

$$\begin{aligned} H_{0b|0a}:\quad & \Sigma=\sigma^2 I_p \qquad (\sigma^2 \text{ não especificada})\\ & \text{assumindo } \Sigma_1=\ldots=\Sigma_q(=\Sigma)\,. \end{aligned} \tag{4}$$

Com base na decomposição em (2) podemos facilmente obter a expressão da estatística (modificada) de razão de verosimilhanças do teste como o produto das estatísticas (modificadas) de razão de verosimilhanças (veja-se Lema 10.3.1 em [1]) utilizadas para testar as hipóteses nulas em (3) e (4), respectivamente λ_a^* e $\lambda_{b|a}^*$, dadas em [1] e [10]. Assim, a estatística de teste é dada por

$$\begin{aligned}
\lambda^* &= \lambda_{b|a}^* \lambda_a^* \\
&= \frac{|A|^{\frac{n^*}{2}}}{(tr\ A)^{n^* p/2}}\, p^{n^* p/2} \frac{(n^*)^{n^* p/2}}{\prod_{j=1}^{q} (n_j)^{pn_j/2}} \frac{\prod_{j=1}^{q} |A_j|^{n_j/2}}{|A|^{n^*/2}} && (5) \\
&= \frac{(pn^*)^{pn^*/2}}{\prod_{j=1}^{q} n_j^{pn_j/2}} \frac{\prod_{j=1}^{q} |A_j|^{n_j/2}}{(trA)^{pn^*/2}} && (6)
\end{aligned}$$

onde A_j é a matriz de somas e produtos cruzados dos desvios para a média, obtidos da j-ésima amostra, $n_j = N_j - 1$ (número de graus de liberdade da distribuição Wishart de A_j), $A = A_1 + \ldots + A_j$ e $n^* = n_1 + \ldots + n_j$ (número de graus de liberdade da distribuição Wishart de A).

Dado que as estatísticas λ_a^* e $\lambda_{b|a}^*$ em (5) são independentes sob H_0 em (1), podemos obter a expressão do h-ésimo momento nulo de λ^* como o produto das expressões dos h-ésimos momentos nulos de λ_a^* e $\lambda_{b|a}^*$ (veja-se [1] e [10]), obtendo assim

$$\begin{aligned}
E\left[(\lambda^*)^h\right] &= E\left[\left(\lambda_{b|a}^*\right)^h\right] \times E\left[(\lambda_a^*)^h\right] \\
&= \frac{p^{hpn^*/2}\,\Gamma\left(\frac{n^* p}{2}\right)}{\Gamma\left(\frac{n^* p}{2} + \frac{phn^*}{2}\right)} \frac{\Gamma_p\left(\frac{n^*}{2} + \frac{n^* h}{2}\right)}{\Gamma_p\left(\frac{n^*}{2}\right)} && (7) \\
&\quad \times \frac{(n^*)^{n^* ph/2}}{\prod_{j=1}^{q} n_j^{pn_j h/2}} \frac{\Gamma_p\left(\frac{n^*}{2}\right)}{\Gamma_p\left(\frac{n^*}{2}(1+h)\right)} \prod_{j=1}^{q} \frac{\Gamma_p\left(\frac{n_j}{2}(1+h)\right)}{\Gamma_p\left(\frac{n_j}{2}\right)} && (8) \\
&= \frac{(pn^*)^{n^* ph/2}}{\prod_{j=1}^{q} n_j^{pn_j h/2}} \frac{\Gamma\left(\frac{n^* p}{2}\right)}{\Gamma\left(\frac{n^* p}{2} + \frac{phn^*}{2}\right)} \prod_{j=1}^{q} \frac{\Gamma_p\left(\frac{n_j}{2}(1+h)\right)}{\Gamma_p\left(\frac{n_j}{2}\right)} && (9)
\end{aligned}$$

onde $\Gamma(.)$ é a função gama e $\Gamma_p(.)$ denota a função gama multivariada (veja-se [1]).

3 A função característica de $W = -\log \lambda^*$

As funções gama em (9) estão bem definidas para qualquer h estritamente complexo, pelo que podemos usar a expressão dos momentos de λ^* para obter a função característica da variável aleatória $W = -\log \lambda^*$, assim

$$\begin{aligned} \Phi_W(t) &= E\left[\mathrm{e}^{\mathrm{i}Wt}\right] = E\left[(\lambda^*)^{-\mathrm{i}t}\right] \\ &= \frac{(pn^*)^{-n^* p\mathrm{i}t/2}}{\prod_{j=1}^{q} n_j^{-pn_j\mathrm{i}t/2}} \frac{\Gamma\left(\frac{n^* p}{2}\right)}{\Gamma\left(\frac{n^* p}{2} - \frac{p\mathrm{i}tn^*}{2}\right)} \prod_{j=1}^{q} \frac{\Gamma_p\left(\frac{n_j}{2}(1-\mathrm{i}t)\right)}{\Gamma_p\left(\frac{n_j}{2}\right)} \ . \end{aligned} \tag{10}$$

De forma a podermos desenvolver aproximações quase-exactas para W vamos usar a factorização de $\Phi_W(t)$ induzida pela decomposição da hipótese nula em (1). Este processo está bem descrito em [7]. Assim, usando as expressões em (7) e (8), substituindo h por $-\mathrm{i}t$, para $n_j = n$ $(j = 1, \ldots, q)$, temos

$$\Phi_W(t) = \underbrace{\frac{(n^*)^{-n^* p\mathrm{i}t/2}}{\prod_{j=1}^{q} n^{-pn\mathrm{i}t/2}} \frac{\Gamma_p\left(\frac{n^*}{2}\right)}{\Gamma_p\left(\frac{n^*}{2}(1-\mathrm{i}t)\right)} \prod_{j=1}^{q} \frac{\Gamma_p\left(\frac{n}{2}(1-\mathrm{i}t)\right)}{\Gamma_p\left(\frac{n}{2}\right)}}_{\Phi_1(t)} \tag{11}$$

$$\times \underbrace{\frac{p^{-\mathrm{i}tpn^*/2}\,\Gamma\left(\frac{n^* p}{2}\right)}{\Gamma\left(\frac{n^* p}{2} - \frac{p\mathrm{i}tn^*}{2}\right)} \frac{\Gamma_p\left(\frac{n^*}{2} - \frac{n^*\mathrm{i}t}{2}\right)}{\Gamma_p\left(\frac{n^*}{2}\right)}}_{\Phi_2(t)} \ . \tag{12}$$

A função característica $\Phi_1(t)$ é a função característica da variável aleatória $-\log \lambda_a^*$, a qual ainda pode ser factorizada no produto de duas funções características. Em [6] os autores mostram que a função característica $\Phi_1(t)$ pode ser escrita como

$$\begin{aligned} \Phi_1(t) &= \Phi_{1,1}(t) \times \Phi_{1,2}(t) \\ &= \underbrace{\prod_{j=1}^{p-1} \left(\frac{n-j}{n}\right)^{r_j} \left(\frac{n-j}{n} - it\right)^{-r_j}}_{\Phi_{1,1}(t)} \\ &\quad \underbrace{\times \prod_{j=1}^{\lfloor p/2 \rfloor} \prod_{i=1}^{q} \frac{\Gamma(a_j + b_{ij})}{\Gamma(a_j + b_{ij}^*)} \frac{\Gamma(a_j + b_{ij}^* - nit)}{\Gamma(a_j + b_{ij} - nit)} \times \left(\prod_{i=1}^{q} \frac{\Gamma(a_p + b_{ip})}{\Gamma(a_p + b_{ip}^*)} \frac{\Gamma\left(a_p + b_{ip}^* - \frac{n}{2}it\right)}{\Gamma\left(a_p + b_{ip} - \frac{n}{2}it\right)} \right)^{p \perp\!\!\!\perp 2}}_{\Phi_{1,2}(t)} \end{aligned} \tag{13}$$

onde $\Phi_{1,1}(t)$ é a função característica de uma distribuição Gama Inteira Generalizada (GIG) de profundidade $p-1$ (veja-se [4]), ou seja, correspondente à soma de $p-1$ variáveis aleatórias independentes com distribuição Gama, com parâmetros de forma r_j inteiros dados pelas expressões (31) a (33) obtidas em [6], e onde $\Phi_{1,2}(t)$ é função característica correspondente à soma de um total de $\lfloor p/2 \rfloor \times q + q \times (p \perp\!\!\!\perp 2)$ variáveis aleatórias independentes com distribuição Logbeta multiplicadas por n ou $n/2$, sendo $p \perp\!\!\!\perp 2$ o resto da divisão inteira de p por 2, sendo ainda

$$a_j = n + 1 - 2j \;, \qquad b_{ij} = 2j - 1 + \frac{i-2j}{q} \;, \qquad b^*_{ij} = \lfloor b_{ij} \rfloor \tag{14}$$

e

$$a_p = \frac{n+1-p}{2} \;, \qquad b_{ip} = \frac{pq-q-p+2i-1}{2q} \;, \qquad b^*_{ip} = \lfloor b_{ip} \rfloor \;. \tag{15}$$

A função característica de $-\log \lambda^*_{b|a}$ é $\Phi_2(t)$ em (12). Em [8] os autores mostram que pode ser factorizada no produto de duas funções características,

$$\begin{aligned} \Phi_2(t) &= \Phi_{2,1}(t) \times \Phi_{2,2}(t) \\ &= \underbrace{\prod_{j=1}^{p-1} \left(\frac{n-\frac{j}{q}}{n}\right)^{s_{j+1,p}} \left(\frac{n-\frac{j}{q}}{n} - \mathrm{i}t\right)^{-s_{j+1,p}}}_{\Phi_{2,1}(t)} \\ &\quad \times \underbrace{\prod_{j=2}^{p} \frac{\Gamma(a_{j,p}+b^*_j+c_j)}{\Gamma(a_{j,p}+b^*_j)} \frac{\Gamma(a_{j,p}+b^*_j-\frac{nq}{2}\mathrm{i}t)}{\Gamma(a_{j,p}+b^*_j+c_j-\frac{nq}{2}\mathrm{i}t)}}_{\Phi_{2,2}(t)} , \end{aligned} \tag{16}$$

onde $\Phi_{2,1}(t)$ é a função característica correspondente à soma de $p-1$ variáveis aleatórias independentes com distribuição Gama, ou seja, uma distribuição GIG de profundidade $p-1$ com parâmetros de forma $s_{j+1,p}$ inteiros dados pela expressão (14) em [8], e onde $\Phi_{2,2}(t)$ é função característica correspondente à soma de $p-1$ variáveis aleatórias independentes com distribuição Logbeta multiplicadas por $nq/2$, com

$$a_{j,p} = \frac{nq+1}{2} - \frac{j}{2} \;, \qquad b_j = \frac{j-1}{p} + \frac{j-1}{2} \;, \qquad b^*_j = \lfloor b_j \rfloor \;. \tag{17}$$

As expressões das funções características $\Phi_{2,1}(t)$ e $\Phi_{2,2}(t)$ são obtidas a partir da expressão (11) em [8] substituindo n por nq e tendo em conta que os autores em [8] não usam a estatística $\lambda^*_{b|a}$ mas sim a estatística $(\lambda^*_{b|a})^{2/n^*}$.

Teorema 1. A função característica $\Phi_W(t)$, para $W = -\log\lambda^*$, pode ser escrita como

$$\begin{aligned}\Phi_W(t) &= \underbrace{\prod_{j=1}^{p-1}\left(\frac{n-j}{n}\right)^{r_j^+}\left(\frac{n-j}{n}-\mathrm{i}t\right)^{-r_j^+}}_{\Phi_{W_1}(t)} \\ &\times \underbrace{\prod_{\substack{j=1\\ j\neq q,\ldots,\alpha q}}^{p-1}\left(\frac{n-\frac{j}{q}}{n}\right)^{s_{j+1,p}}\left(\frac{n-\frac{j}{q}}{n}-\mathrm{i}t\right)^{-s_{j+1,p}}}_{\Phi_{W_2}(t)} \\ &\times \underbrace{\Phi_{1,2}(t)\times\Phi_{2,2}(t)}_{\Phi_{W_3}(t)}\end{aligned} \tag{18}$$

onde $\alpha = \left\lfloor \frac{p-1}{q}\right\rfloor$ e

$$r_j^+ = \begin{cases} r_j + s_{q\times j+1,p} & j = 1,\ldots\alpha \\ r_j & j = \alpha+1,\ldots,p-1\end{cases} \tag{19}$$

com r_j e $s_{j+1,p}$ dados respectivamente por (31) a (33) em [6] e (14) em [8].

Demonstração. Verifica-se que α variáveis aleatórias com distribuição Gama cujas funções características são factores de $\Phi_{2,1}(t)$, mais precisamente, as que têm parâmetro de escala igual a $(n-j/q)/n$ para $j = q, 2q, \ldots, \alpha\, q$, têm o mesmo parâmetro de escala que algumas das variáveis aleatórias com distribuição Gama com funções características em $\Phi_{1,1}(t)$. Então agrupamos em $\Phi_{W_1}(t)$ todas as funções características em $\Phi_{1,1}(t)$ mais as funções características em $\Phi_{2,1}(t)$ que correspondem a variáveis aleatórias com distribuição Gama cujo parâmetro de escala é dado por $(n-j/q)/n$ para $j = q, 2q, \ldots, \alpha\, q$. Em $\Phi_{W_2}(t)$ deixamos as restantes funções características de $\Phi_{2,1}(t)$. ∎

4 Distribuições quase-exactas para $W = -\log\lambda^*$ e λ^*

A obtenção das distribuições quase-exactas para W é feita a partir da factorização da função característica em (18), aproximando assimptóticamente $\Phi_{W_3}(t)$ por outra função característica, que neste caso corresponderá à função característica de uma distribuição Gama ou à mistura de duas ou três distribuições Gama. A aproximação da distribuição da soma de variáveis aleatórias independentes com distribuição Logbeta por uma Gama ou uma mistura de Gamas está devidamente justificada em [6], [7] e [8]. Com este procedimento vamos obter como distribuições quase-exactas uma distribuição Gama Quase-Inteira Generalizada (GQIG) (veja-se [5]) ou uma mistura de duas ou três distribuições

GQIG. Estas distribuições são distribuições altamente manejáveis e que ao mesmo tempo apresentam valores muito próximos da distribuição exacta que pretendemos aproximar. Assim, vamos aproximar $\Phi_{W_3}(t)$ por $\Phi^*_{W_3}(t)$ dada por

$$\Phi^*_{W_3}(t) = \sum_{k=1}^{h/2} p_k \, \lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k} \,, \tag{20}$$

com pesos $p_k > 0$ tais que

$$\sum_{k=1}^{h/2} p_k = 1 \,, \tag{21}$$

e onde $h = 2, 4$ ou 6, conforme o caso em que $\Phi^*_{W_3}(t)$ seja a função característica correspondente a uma variável aleatória Gama ($h = 2$) ou à mistura de duas Gamas ($h = 4$) ou à misturas de três Gamas ($h = 6$). Os parâmetros em (20) são obtidos através da resolução do sistema de equações

$$\left. \frac{d^j}{dt^j} \, \Phi^*_{W_3}(t) \right|_{t=0} = \left. \frac{d^j}{dt^j} \, \Phi_{W_3}(t) \right|_{t=0} \,, \qquad j = 1, \ldots, h \,, \tag{22}$$

de forma a que os primeiros h momentos da distribuição quase-exacta igualem os primeiros h momentos exactos. O processo pode ser facilmente descrito a partir do seguinte esquema:

$$\Phi_W(t) = \underbrace{\Phi_{W_1}(t) \times \Phi_{W_2}(t)}_{\text{distribuição GIG}} \times \underbrace{\Phi_{W_3}(t)}_{\text{Soma de distribuições Logbeta independentes}}$$

$$\text{substituição assimptótica para } \Phi_{W_3}(t) \longrightarrow \quad \Big\downarrow \; \textstyle\sum_{k=1}^{h/2} p_k \, \lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k}$$

$$\begin{aligned} \Phi_W(t) &\approx \Phi_{W_1}(t) \times \Phi_{W_2}(t) \times \sum_{k=1}^{h/2} p_k \, \lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k} \\ &= \underbrace{\sum_{k=1}^{h/2} p_k \, \Phi_{W_1}(t) \times \Phi_{W_2}(t) \; \times \; \lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k}}_{\text{distribuição GQIG ou mistura de distribuições GQIG}} \end{aligned}$$

Teorema 2. Se substituirmos assimptoticamente $\Phi_{W_3}(t)$ em (18) por $\Phi^*_{W_3}(t)$ em (20) obtemos como distribuições quase-exactas para W uma distribuição GQIG ou uma mistura de duas ou três distribuições GQIG com profundidade $2(p-1)-\alpha+1$ para $\alpha=\left\lfloor \frac{p-1}{q}\right\rfloor$, consoante consideremos $h=2,4$ ou 6, com função densidade de probabilidade

$$\sum_{\nu=1}^{h/2} p_\nu\, f^{\,GNIG}\Big(w|r_1^+,\dots,r_{p-1}^+,\underbrace{s_{2,p},\dots,s_{p,p}}_{\text{excepto } s_{j+1,p}\ ,\ j=q,\dots,\alpha q},s_\nu; \frac{n-1}{n},\dots,\frac{n-p+1}{n},\underbrace{\frac{n-1/q}{n},\dots,\frac{n-(p-1)/q}{n}}_{\text{excepto } \frac{n-j/q}{n}\ ,\ j=q,\dots,\alpha q},\lambda\Big) \tag{23}$$

e função distribuição

$$\sum_{\nu=1}^{h/2} p_\nu\, F^{\,GNIG}\Big(w|r_1^+,\dots,r_{p-1}^+,\underbrace{s_{2,p},\dots,s_{p,p}}_{\text{excepto } s_{j+1,p}\ ,\ j=q,\dots,\alpha q},s_\nu; \frac{n-1}{n},\dots,\frac{n-p+1}{n},\underbrace{\frac{n-1/q}{n},\dots,\frac{n-(p-1)/q}{n}}_{\text{excepto } \frac{n-j/q}{n}\ ,\ j=q,\dots,\alpha q},\lambda\Big) \tag{24}$$

onde r_j^+ e $s_{j,p}$ são dados respectivamente em (19) e na expressão (14) em [8], e onde para $h=2$

$$\lambda=\frac{m_1}{m_2-m_1^2} \qquad \text{e} \qquad s_1=\frac{m_1^2}{m_2-m_1^2} \tag{25}$$

com

$$m_j=\mathrm{i}^{-j}\,\frac{\partial^j}{\partial t^j}\,\Phi_{W_3}(t)\Big|_{t=0}\,, \qquad j=1,2\,,$$

e para $h=4$ ou $h=6$ os valores de p_ν, s_ν e λ são as soluções numéricas dos sistemas de equações em (22), isto é das equações

$$\frac{d^j}{dt^j}\,\Phi^*_{W_3}(t)\Big|_{t=0}=\frac{d^j}{dt^j}\,\Phi_{W_3}(t)\Big|_{t=0}\,, \qquad j=1,\dots,h$$

com

$$p_{h/2}=1-\sum_{k=1}^{h/2-1}p_k\,.$$

Demonstração. Vamos considerar apenas o caso $h = 6$ uma vez que os outros se demonsram de forma análoga. Se na função característica de $\Phi_W(t)$ em (18) substituirmos $\Phi_{W_3}(t)$ por $\Phi_{W_3}(t)^*$ em (20)

$$\begin{aligned}
\Phi_{W}(t) \quad &\approx \quad \Phi_{W_1}(t) \times \Phi_{W_2}(t) \times \underbrace{\sum_{k=1}^{3} p_k \, \lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k}}_{\Phi^*_{W_3}(t)} \\
&\approx \quad \sum_{k=1}^{3} p_k \; \underbrace{\underbrace{\Phi_{W_1}(t) \times \Phi_{W_2}(t)}_{\text{distribuição GIG}} \times \underbrace{\lambda^{s_k} \, (\lambda - \mathrm{i}t)^{-s_k}}_{\text{distribuição Gama}}}_{\text{distribuição GQIG}}
\end{aligned}$$

ou seja, a função característica da mistura de três distribuições GQIG de profundidade $2(p-1)-\alpha+1$ com $\alpha = \left\lfloor \frac{p-1}{q} \right\rfloor$ e com correspondente função distribuição dada por (24). Os parâmetros p_ν, s_ν e λ são obtidos de forma a igualar os seis primeiros momentos de W, ou seja

$$\frac{d^j}{dt^j} \, \Phi^*_{W_3}(t) \bigg|_{t=0} = \frac{d^j}{dt^j} \, \Phi_{W_3}(t) \bigg|_{t=0} , \qquad j = 1, \ldots, 6 \,. \blacksquare$$

Note-se que a partir das distribuições quase-exactas obtidas para W podemos facilmente obter a distribuições quase-exactas para λ^*.

Corolário 2.1 Distribuições quase-exactas para a estatística de razão de verosimilhanças λ^* podem ser obtidas da seguinte forma

$$\begin{aligned}
f_{\lambda^*}(\ell) \approx \sum_{\nu=1}^{h/2} p_\nu \, f^{\,GNIG}\Big(-\log \ell | r_1^+, \ldots, r_{p-1}^+, \; &\underbrace{s_{2,p}, \ldots, s_{p,p}}_{\text{excepto } s_{j+1,p} \;,\; j=q,\ldots,\alpha q} \;, s_\nu; \\
&\frac{n-1}{n}, \ldots, \frac{n-p+1}{n}, \underbrace{\frac{n-1/q}{n}, \ldots, \frac{n-(p-1)/q}{n}}_{\text{excepto } \frac{n-j/q}{n} \;,\; j=q,\ldots,\alpha q}, \lambda \Big) \frac{1}{\ell} \qquad (26)
\end{aligned}$$

e

$$\begin{aligned}
F_{\lambda^*}(\ell) \approx 1 - \sum_{\nu=1}^{h/2} p_\nu \, F^{\,GNIG}\Big(-\log \ell | r_1^+, \ldots, r_{p-1}^+, \; &\underbrace{s_{2,p}, \ldots, s_{p,p}}_{\text{excepto } s_{j+1,p} \;,\; j=q,\ldots,\alpha q} \;, s_\nu; \\
&\frac{n-1}{n}, \ldots, \frac{n-p+1}{n}, \underbrace{\frac{n-1/q}{n}, \ldots, \frac{n-(p-1)/q}{n}}_{\text{excepto } \frac{n-j/q}{n} \;,\; j=q,\ldots,\alpha q}, \lambda \Big) \qquad (27)
\end{aligned}$$

para $h=2$, $h=4$ ou $h=6$, $0<\ell<1$ e onde os parâmetros em (26) e (27) são iguais aos das expressões (23) e (24) do Teorema 2.

Demonstração. Uma vez que no Teorema 2 são obtidas distribuições quase-exactas para a variável aleatória $W=-\log\lambda^*$, para obtermos as correspondentes distribuições quase-exactas de λ^* só é preciso ter em conta que

$$F_{\lambda^*}(\ell)=1-F_W(-\log\ell)$$

onde $F_{\lambda^*}(\cdot)$ representa a função distribuição de λ^* e $F_W(\cdot)$ a função distribuição de W. ■

5 Estudos numéricos

Com o objectivo de avaliar a qualidade das distribuições quase-exactas vamos utilizar um medida de proximidade baseada na fórmula de inversão das funções características e que nos dá um "upper bound" para o módulo da diferença entre as respectivas funções distribuição. Assim, seja Y uma variável aleatória com função distribuição $F_Y(y)$ e função característica $\Phi_Y(t)$, e sejam $\Phi_n(t)$ e $F_n(y)$ as funções característica e distribuição da variável aleatória X_n. Esta medida é

$$\Delta=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\Phi_Y(t)-\Phi_n(t)}{t}\right|\mathrm{d}t\,,\quad \text{com}\quad \max_{y\in S}|F_Y(y)-F_n(y)|\;\leq\;\Delta\;.$$

A medida Δ já foi utilizada em [6] e [8] para avaliar a qualidade das distribuições quase-exactas obtidas para a estatística do teste de esfericidade e de igualdade de matrizes de covariância, respectivamente. Nas Tabelas e Figuras apresentadas vamos denotar por GQIG, M2GQIG e M3GQIG respectivamente as distribuições quase-exactas GQIG, mistura de duas distribuições GQIG e mistura de três distribuições GQIG. De forma a melhor podermos avaliar a qualidade destas novas aproximações vamos utilizar como base de comparação a aproximação assimptótica obtida por Moschopoulos em [9], e que denotaremos por Box. Na Tabela 1 apresentamos os valores da medida Δ para o caso em que p e n estão fixos e q aumenta e ainda o caso em que p e n aumentam, mantendo-se fixo o valor de q e a diferença $n-p$. Podemos observar que os valores apresentados pelas aproximações quase-exactas são consideravelmente menores do que os valores obtidos para a aproximação de Moschopoulos, com especial relevância para a distribuição M3GQIG que apresenta em todos os casos valores consideravelmente mais baixos. Note-se ainda que ao contrário da aproximação assimptótica de Moschopoulos as aproximações quase-exactas apresentam boas propriedades assimptóticas para crescentes valores de p e q.

Na Tabela 2 apresentamos casos em que os valores de p e q se mantêm inalterados com aumentos sucessivos do valor de n. Mais uma vez as aproximações quase-exactas apresentam valores de Δ substancialmente menores, continuando a revelar o seu carácter assimptótico agora para valores crescentes de n. Como

Tabela 1: Valores da medida Δ para valores crescentes de q e p.

p	q	n	M3GQIG	M2GQIG	GQIG	Box
7	6	9	1.5×10^{-11}	8.9×10^{-9}	7.2×10^{-6}	1.7×10^{0}
7	13	9	1.9×10^{-12}	2.0×10^{-9}	3.0×10^{-6}	2.6×10^{0}
7	15	9	1.3×10^{-12}	1.5×10^{-9}	2.5×10^{-6}	2.8×10^{0}
9	6	11	8.7×10^{-13}	1.3×10^{-9}	2.4×10^{-6}	2.8×10^{0}
15	6	17	3.4×10^{-14}	1.2×10^{-10}	5.4×10^{-7}	6.8×10^{0}

já era de esperar a aproximação assimptótica de Moschopoulos melhora os seus valores à medida que os valores de n aumentam ficando todavia muito aquém dos valores apresentados pelas aproximações quase-exactas.

Tabela 2: Valores da medida Δ para valores crescentes de n.

p	q	n	M3GQIG	M2GQIG	GQIG	Box
7	6	9	1.5×10^{-11}	8.9×10^{-9}	7.2×10^{-6}	1.7×10^{0}
7	6	60	9.3×10^{-15}	2.3×10^{-11}	3.1×10^{-7}	2.3×10^{-2}
7	6	120	7.8×10^{-16}	2.1×10^{-12}	7.9×10^{-8}	5.4×10^{-3}
7	6	200	1.2×10^{-16}	3.8×10^{-13}	2.8×10^{-8}	1.9×10^{-3}

6 Conclusão

A decomposição da hipótese nula em hipóteses nulas parciais condicionalmente independentes é um processo fácil de implementar, o qual pode ser utilizado com sucesso na obtenção de distribuições quase-exactas para a estatística de razão de verosimilhanças utilizada para testar a hipótese nula em (1). As aproximações quase-exactas revelam boas propriedades assimptóticas não só para valores crescentes da dimensão da amostra mas também no que diz respeito ao número de variáveis e ao número de populações em estudo.

A aproximação assimptótica de Moschopoulos revela valores, para a medida Δ, bastante superiores aos das aproximações quase-exactas. Os valores são de tal forma elevados que, nomeadamente quando são superiores a 1, esta aproximação não deverá sequer corresponder à distribuição de uma variável aleatória.

As distribuições quase-exactas não só se revelam fáceis de utilizar como, dado o seu elevado grau de precisão, isto é, de proximidade à distribuição exacta, convertem-se numa ferramenta muito útil em termos práticos.

Agradecimentos

Este trabalho de investigação foi financiado pela Fundação para a Ciência e Tecnologia (FCT) através do Centro de Matemática e Aplicações (CMA) da Universidade Nova de Lisboa.

Referências

[1] Anderson, T. W., (1958). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. New York: J. Wiley & Sons.

[2] Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.

[3] Butler, R. and Huzurbazar, S. and Booth, J. (1993). Saddlepoint approximations for tests of block independence, sphericity and equal variances and covariances. *Journal of the Royal Statistical Society*, Series B, 55, 171-183.

[4] Coelho, C. A. (1998). The Generalized Integer Gamma Distribution – A basis for distributions in Multivariate Statistics. *Journal of Multivariate Analysis*, 64, 86-102.

[5] 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.

[6] Coelho, C. A. e Marques, F. J. (2007). Near-exact approximations for the likelihood ratio test statistic for testing equality of several variance-covariance matrices. *Universidade Nova de Lisboa, Departamento de Matemática*, Relatório Técnico 12/2007.

[7] Coelho, C. A. and Marques, F. J. (2008). 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*, in print.

[8] Marques, F. J. e Coelho, C. A. (2008). Near-exact distributions for the sphericity likelihood ratio test statistic. *Journal of Statistical Planning and Inference*, Vol. 138, p. 726-741.

[9] Moschopoulos, P. G. (1988). Asymptotic expansions of the non-null distribution of the likelihood ratio criterion for multisample sphericity. American Journal of Mathematical and Management Sciences, Vol. 8, p. 135-163.

[10] Muirhead, R. J. (1982). *Aspects of Multivariate Statiscal Theory.* New York: J. Wiley & Sons.