Testes sobre a estrutura de matrizes de covariância

Filipe J. Marques, *fjm@fct.unl.pt*
Carlos A. Coelho, *cmac@fct.unl.pt*

Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT NOVA)
Centro de Matemática e Aplicações (CMA)

1. Introdução

A estrutura da matriz de covariância pode revelar caraterísticas importantes de uma determinada distribuição ou, no caso amostral, da estrutura dos dados. Vários modelos nas mais diversas áreas de investigação assumem como pressupostos estruturas para a matriz de covariância dos erros que podem ser simples ou ter alguma complexidade. Por este motivo, é importante ter ferramentas que nos permitam realizar, com a precisão adequada, testes sobre estruturas de matrizes de covariância. Se considerarmos uma população $N_p(\underline{\mu}, \Sigma)$, temos como alguns exemplos de estruturas mais simples:

1. Independência: $\Sigma = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$

2. Esférica: $\Sigma = \sigma^2 I_p$

3. Igualdade de variâncias e de covariâncias: $\Sigma = \sigma^2 \left((1-\rho)I_p + \rho E_{pp}\right)$ (onde $-\frac{1}{p-1} < \rho < 1$ e E_{pp} é uma matriz de ordem p com todas as entradas iguais a 1)

4. Circular: $\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_3 & \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$ (para $p = 6$)

5. Indepêndencia de grupos de variáveis: $\Sigma = bdiag(\Sigma_{11}, \dots, \Sigma_{kk}, \dots, \Sigma_{mm})$, onde Σ_{kk} é uma matriz de ordem p_k, com $p_1 + \cdots + p_k + \cdots + p_m = p$.

Claro que a estrutura de Σ pode-se tornar mais complexa por composição das estruturas acima. O interesse no estudo destas estruturas ditas mais complexas é hoje em dia potenciado pela também complexidade de novos modelos, nomeadamente modelos mistos. Veremos mais à frente como podem ser feitos testes a este tipo de estruturas.

Para realizar testes, quer a estruturas mais simples quer a estruturas complexas das matrizes de covariância, é possível deduzir as estatísticas de razão de verosimilhanças, de forma mais ou menos trabalhosa, contudo a questão coloca-se nas distribuições exatas destas estatísticas, as quais são normalmente de estrutura demasiado elaborada, o que torna difícil a sua implementação computacional e por isso pouco úteis na prática. Em geral, as estatísticas de razão de verosimilhanças, usadas em testes sobre a estrutura de matrizes de covariância, têm uma distribuição igual à do produto de variáveis aleatórias independentes com distribuição Beta. Existe uma vasta literatura sobre este tópico onde constam diferentes representações para esta distribuição como são os casos das representações em série (Tang e Gupta, 1984; Moschopoulos, 1986), das representações através de funções G de Meijer (Meijer, 1946; Nagar et al.,1985) ou funções H de Fox (Fox, 1961; Springer, 1979; Carter e Springer, 1977), entre outras. Contudo, hoje em dia, com toda capacidade computacional existente, ainda pode ser um problema obter quantis ou *p-values* precisos para estas distribuições. Em Coelho e Alberto (2012) os autores apresentam uma revisão de literatura muito detalhada sobre produto de variáveis aleatórias independentes com distribuição Beta. Neste artigo os autores desenvolvem distribuições quase-exatas precisas e computacionalmente implementáveis para produto de variáveis aleatórias independentes com distribuição Beta. No que diz respeito a testes sobre a estrutura de matrizes de covariância é bem conhecido que a distribuição do logaritmo da estatística de razão de verosimilhanças pode ser aproximada por um qui-quadrado, eventualmente multiplicado por um fator de correção. Estas aproximações podem ser melhoradas se considerarmos as aproximações obtidas por Box (1949) que são usualmente apresentadas como misturas de duas distribuições Gama. Contudo, o desempenho destas aproximações é limitado, principalmente se considerarmos cenários extremos como aqueles em que temos amostras de dimensão reduzida e/ou um número elevado de variáveis. Uma alternativa diferente são as aproximações ponto-de-sela (Daniels, 1954; Booth *et. al*, 1995). Contudo estas têm a desvantagem de não produzirem uma expressão nem para a função densidade nem para a função distribuição, mas apenas aproximações para pontos específicos, e a literatura mostra que estas podem ser francamente melhoradas. Mais recentemente, surgiram as aproximações quase-exatas (Coelho, 2004) que têm sido bastante utilizadas para aproximar a distribuição de estatísticas de razão de verosimilhanças utilizadas para realizar testes sobre a estrutura de matrizes de covariância e também em problemas relacionados com a distribuição de produtos, somas e combinações lineares de variáveis aleatórias. As aproximações quase-exatas podem ser utilizadas em estruturas simples como as já apresentadas ou em estruturas mais complexas. O procedimento para o desenvolvimento destas aproximações será apresentado em detalhe na secção seguinte.

2. Testes sobre matrizes de covariância com estruturas complexas

Muitas estruturas complexas podem ser interpretadas como composições de testes mais simples. Por exemplo, o teste de esfericidade apresentado anteriormente pode ser visto como a composição de dois testes; o teste à independência de várias variáveis e o teste de igualdade de variâncias, aliás em Anderson (2003) o autor utiliza esta mesma estratégia para obter a estatística de razão de verosimilhanças do teste. Em Coelho e Marques (2009) os autores mostram com é possível desenvolver distribuições quase-exatas para estruturas ditas complexas. A ideia geral é a seguinte: suponhamos que pretendemos testar uma determinada estrutura complexa e especificada na hipótese nula H_0 versus a correspondente hipótese alternativa H_1, a ideia fundamental é tentar decompor, de forma adequada, a hipótese nula inicial numa sequência de hipóteses nulas parciais. Suponhamos então que é possível fazer a decomposição de H_0 em *m* hipóteses nulas parciais, que podem ter que obedecer a uma determinada ordem, e cuja decomposição pode ser apresentada através da seguinte notação

$$H_0 \equiv H_{0m|1,\cdots,m-1} \circ \dots \circ H_{02|1} \circ \ H_{01}$$

como referido em Coelho e Marques (2009) esta notação representa que testar H_0 é equivalente a testar sequencialmente as *m* hipóteses $H_{0j|1,\cdots,j-1}$ $(j \ = 1, \dots, m)$, testando primeiro H_{01}, em seguida $H_{02|1}$, depois $H_{03|1,2}$, e assim sucessivamente, onde testar $H_{0j|1,\dots,j-1}$ representa testar H_{0j} assumindo que as hipótese H_{01} até $H_{0,j-1}$ não são rejeitadas. Note-se que, de uma forma geral, fazendo uma decomposição adequada de H_0 tem-se, sob esta hipótese nula, que as estatísticas de razão de verosimilhanças $\Lambda_{j|1,\cdots,\mathrm{j}-1}$ usadas para testar as hipóteses parciais $H_{0j|1,\dots,j-1}$ $(j \ = 1, \dots, m)$ são independentes. Tendo por base esta decomposição a estatística de razão de verosimilhanças, Λ, usada para testar a hipótese nula global H_0 é dada por

$$\Lambda = \prod_{j=1}^{m} \Lambda_{j|1,\cdots,\mathrm{j}-1}.$$

Tendo em conta a independência das estatísticas $\Lambda_{j|1,\cdots,j-1}$ sob H_0 podemos determinar a expressão do h-ésimo momento de Λ como o produto dos h-ésimos momentos das estatísticas $\Lambda_{j|1,\cdots,j-1}$ ou seja

$$E[\Lambda^h] = \prod_{j=1}^m E[\Lambda_{j|1,\cdots,j-1}^h].$$

A partir desta última expressão é possível obter a função caraterística da variável aleatória $W = -\log\Lambda$ da seguinte forma

$$\Phi_W(t) = E\left[\mathrm{e}^{\mathrm{i}tW}\right] = E\left[\Lambda^{-\mathrm{i}t}\right] = \prod_{j=1}^m E\left[\Lambda_{j|1,\cdots,j-1}^{-\mathrm{i}t}\right] = \prod_{j=1}^m E\left[\mathrm{e}^{\mathrm{i}tW_{j|1,\cdots,j-1}}\right] = \prod_{j=1}^m \Phi_{W_{j|1,\cdots,j-1}}(t)\,, t \in \mathbb{R}$$

onde $\Phi_{W_{j|1,\cdots,j-1}}(t)$ representa a função caraterística de $\mathrm{W}_{j|1,\ldots,j-1} = -\log\Lambda_{j|1,\ldots,j-1}, j = 1, \ldots, m$. A fatorização obtida deste modo para a função caraterística de W é o procedimento base para o desenvolvimento das aproximações quase-exatas para W e para Λ. O passo seguinte para a construção destas aproximações é obter uma nova fatorização da função caraterística de W de forma a que se aproximarmos um dos fatores por outra função característica possamos obter uma nova função caraterística à qual corresponda uma distribuição conhecida e fácil de utilizar na prática. Apresentamos na secção seguinte um exemplo deste procedimento.

3. Exemplo

Para ilustrar o procedimento descrito na secção anterior vamos apresentar sumariamente o teste estudado em (Marques e Coelho, 2015). Por uma questão de simplicidade vamos omitir algumas expressões podendo estas ser consultadas com detalhe na referência acima. Suponhamos então que, dada uma amostra extraída de uma população $N_p(\underline{\mu}, \Sigma)$ estamos interessados em testar a seguinte hipótese nula

$$H_0 \colon \Sigma = \begin{pmatrix} \Sigma_{11} & 0 & 0 \\ 0 & \Sigma_{22} & 0 \\ 0 & 0 & \Sigma_{33} \end{pmatrix} \qquad (1)$$

isto é, pretende-se testar se a matriz de covariância Σ tem uma estrutura diagonal por blocos em que Σ_{11} é uma matriz de ordem $p_1$sem uma estrutura especifica, Σ_{22} , de ordem p_2, tem uma estrutura esférica, ou seja, $\Sigma_{22} = \sigma^2 I_{p_2}$ (Anderson, 2003; Marques e Coelho, 2008) e Σ_{33}, de ordem p_3 tem uma estrutura circular representada por Σ_C (Olkin e Press, 1969; Marques e Coelho, 2013) e onde $p = p_1 + p_2 + p_3$.

É importante referir que o pressuposto de normalidade, em alguns casos, poder ser estendido a outras distribuições, por exemplo em Anderson et al. (1986) os autores, para uma classe de distribuições elípticas, obtém as estatísticas de razão de verosimilhanças para alguns testes sobre estruturas de matrizes de covariância e referem que a distribuição é a mesma que a do caso Normal.

Considerando o procedimento apresentado na secção anterior, vamos decompor a hipótese nula em (1) em três hipóteses nulas parciais, a primeira utilizada para testar a independência dos três grupos de variáveis

$$H_{01} \colon \Sigma_{ij} = 0\,, \qquad i \neq j, \qquad i,j = 1, \ldots, 3 \qquad (2)$$

a segunda para testar a estrutura esférica do segundo bloco diagonal da matriz de covariância de ordem p_2

$$H_{02|1} \colon \Sigma_{22} = \sigma^2 I_{p_2} \text{ (assumindo que } H_{01} \text{ não é rejeitada)} \qquad (3)$$

e a terceira para testar a estrutura circular do terceiro bloco diagonal de ordem p_3

$$H_{03|1}: \Sigma_{33} = \Sigma_C \text{ (assumindo que } H_{01} \text{ não é rejeitada).} \tag{4}$$

Assim, com base no Lema 10.3.1 apresentado em Anderson (2003), a estatística de razão de verosimilhanças, Λ, usada para testar H_0 em (1) é dada pelo produto das estatísticas de razão de verosimilhanças utilizadas para testar as hipóteses nulas parciais apresentadas em (2), (3) e (4). Pelo que, usando as expressões das estatísticas de teste utilizadas para testar H_{01}, $H_{02|1}$ e $H_{03|1}$designadas respetivamente por Λ_1, $\Lambda_{2|1}$e $\Lambda_{3|1}$ e dadas em Marques e Coelho (2015), Anderson (2003, sec. 9.2, 10.7) e Olkin e Press (1969, sec. 3.3) obtem-se

$$\Lambda = \Lambda_1 \times \Lambda_{2|1} \times \Lambda_{3|1}.$$

Pode encontrar todos os detalhes sobre a expressão de Λ na expressão (4) em Marques e Coelho (2015). Dada a independência das estatísticas Λ_1, $\Lambda_{2|1}$e $\Lambda_{3|1}$, sob H_0, a expressão do h-ésimo momento pode ser obtida como o produto das expressões dos h-ésimos momentos das estatísticas Λ_1, $\Lambda_{2|1}$e $\Lambda_{3|1}$, disponíveis em Marques e Coelho (2015), Anderson (2003, sec. 9.3, 10.7) e Olkin e Press (1969, sec. 3.3). Assim,

$$\mathrm{E}[\Lambda^h] = E[{\Lambda_1}^h] \times E[{\Lambda_{2|1}}^h] \times E[{\Lambda_{3|1}}^h].$$

Consideremos agora a variável aleatória $W = -\log \Lambda$, cuja função caraterística é dada por

$$\begin{aligned}\Phi_W(t) = E\left[\mathrm{e}^{\mathrm{i}tW}\right] = E\left[\Lambda^{-\mathrm{i}t}\right] = E[{\Lambda_1}^{-\mathrm{i}t}] \times E[{\Lambda_{2|1}}^{-\mathrm{i}t}] \times E[{\Lambda_{3|1}}^{-\mathrm{i}t}] \\ = \Phi_{W_1}(t) \times \Phi_{W_{2|1}}(t) \times \Phi_{W_{3|1}}(t)\end{aligned}$$

onde Φ_{W_1}, onde $\Phi_{W_{2|1}}$ e onde $\Phi_{W_{3|1}}$ são, respetivamente, as funções caraterísticas das variáveis aleatórias $W_1 = -\log \Lambda_1$, $W_{2|1} = -\log \Lambda_{2|1}$ e $W_{3|1} = -\log \Lambda_{3|1}$. Como já referido, o objetivo agora é encontrar uma fatorização de Φ_W de forma que, mantendo a maior parte intacta, e aproximando um dos fatores por outra função característica possamos obter uma nova função caraterística à qual corresponda uma distribuição conhecida e manejável. Em Marques e Coelho (2015) os autores mostram que é possível escrever Φ_W da seguinte forma:

$$\Phi_W(t) = \Phi_{W_1^*}(t) \times \Phi_{W_2^*}(t) \tag{6}$$

onde $\Phi_{W_1^*}$ é a função característica da soma de um dado número de variáveis aleatórias independentes com distribuição Gama com parâmetros de forma inteiros, o que corresponde a uma distribuição designada por Gama Inteira Generalizada (GIG) obtida em Coelho (1998) e $\Phi_{W_2^*}$ é a função caraterística da soma, de um dado número, de variáveis aleatórias com distribuição Logbeta (note-se que se X tem distribuição Beta de parâmetros a e b então dizemos que $-\log X$ tem uma distribuição Logbeta com os mesmos parâmetros). Usando os resultados em Tricomi e Erdélyi (1951) sabemos que uma simples distribuição Logbeta pode ser aproximada por uma mistura infinita de distribuições Gama, pelo que a abordagem seguida passa por aproximar a função característica $\Phi_{W_2^*}$ em (6) por uma mistura de distribuições Gama cuja função caraterística é dada por

$$\Phi_{\widetilde{W}_2}(t) = \sum_{j=0}^{m} \pi_j \, \lambda^{r+j} (\lambda - \mathrm{i}t)^{-(r+j)} \tag{7}$$

de forma a que $\widetilde{W}_2$ tenha os mesmos m primeiros momentos de W_2^*. Obtem-se assim como função caraterística aproximada de Φ_W

$$\Phi_W(t) \approx \Phi_{NE}(t) = \Phi_{W_1^*}(t) \times \Phi_{\widetilde{W}_2}(t).$$

No que se segue designaremos a função caraterística Φ_{NE} como função caraterística quase-exata. Na expressão de $\Phi_{\widetilde{W}_2}$ em (7) o parâmetro λ é a taxa de uma mistura de duas distribuições Gama que acerta os primeiros quatro momentos de W_2^* e r é igual à soma dos segundos parâmetros das distribuições Logbeta que caraterizam a distribuição de $\Phi_{W_2^*}$ em (6) para mais detalhes veja-se Coelho et al. (2010).

Fixados os parâmetros λ e r os pesos π_j são determinados de forma a que $\widetilde{W}_2$ tenha os mesmos m primeiros momentos de W_2^*, ou seja, são as soluções do seguinte sistema de equações

$$\frac{\partial^h}{\partial t^h}\Phi_{W_2^*}(t)\bigg|_{t=0} = \frac{\partial^h}{\partial t^h}\Phi_{\widetilde{W}_2}(t)\bigg|_{t=0}, \qquad h = 1, \ldots, m, \qquad \text{com } \pi_m = 1 - \sum_{j=0}^{m-1}\pi_j.$$

Note-se que este sistema de equações é de resolução simples com um software de cálculo matemático.

Finalmente, seguindo esta construção, obtemos como função caraterística quase-exata

$$\Phi_{NE}(t) = \sum_{j=0}^{m}\pi_j\left\{\Phi_{W_1^*}(t)\,\lambda^{r+j}(\lambda - \mathrm{i}t)^{-(r+j)}\right\}. \qquad (8)$$

Para um valor de j fixo a expressão $\Phi_{W_1^*}(t)\,\lambda^{r+j}(\lambda - \mathrm{i}t)^{-(r+j)}$ corresponde à função característica da soma de duas variáveis aleatórias independentes; W_1^* com distribuição GIG e uma variável aleatória com distribuição Gama com taxa λ e parâmetro de forma $r+j$. Se r for um número inteiro a soma destas duas variáveis aleatórias continua a ter uma distribuição GIG, se por outro lado r não for inteiro a distribuição da soma é uma Gama Quase-Inteira Generalizada (GQIG) obtida em Coelho (2004). Pelo que a distribuição correspondente à função caraterística Φ_{NE} em (8) é uma mistura de distribuições GIG ou uma mistura de distribuições GQIG consoante r seja inteiro ou não.

Em geral, as aproximações obtidas através deste processo apresentam elevado grau de precisão e são assimptóticas não só relativamente ao tamanho da amostra mas também a outros parâmetros envolvidos, como por exemplo o número de variáveis. Para avaliar as qualidade destas aproximações, em Marques e Coelho (2015), os autores utilizam uma medida de proximidade dada por

$$\Delta = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\Phi_W(t) - \Phi_{NE}(t)}{t}\right|\;\mathrm{d}t\,. \qquad (8)$$

Esta medida, baseada nas funções características exata e aproximada, fornece um valor numérico para o limite superior da distância entre a função distribuição exata e a aproximada. Podem observar-se, a partir da Tabela 1 em Marques e Coelho (2015), os valores da medida em diferente cenários. Estes valores ilustram a qualidade das aproximações e também as suas propriedades assimptóticas.

Referências

Anderson, T. W. (2003) - *An Introduction to Multivariate Statistical Analysis*. 3rd ed., J. Wiley & Sons, New York.

Anderson, T., Fang, K., Hsu, H. (1986) - Maximum-Likelihood Estimates and Likelihood-Ratio. Criteria for Multivariate Elliptically Contoured Distributions. *The Canadian Journal of Statistics*,14, 55-59.

Booth, J. G., Butler, R. W., Huzurbazar, S., Wood, A. T. A. (1995) - Saddlepoint approximations for p-values of some tests of covariance matrices. *Journal of Statistical Computation and Simulation*, 53, 165-180.

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

Carter, B. D., Springer, M. D. (1977) - The distribution of products, quotients and powers of independent H-function variates. *SIAM J. Appl. Math.* 33, 542-558.

Coelho, C. A. (1998) - The Generalized Integer Gamma Distribution - A Basis for Distributions in Multivariate Statistics. *Journal of Multivariate Analysis*, 64, 86–102.

Coelho, C. A. (2004) - The Generalized Near-Integer Gamma Distribution: A Basis for 'Near-Exact' Approximations to the Distribution of Statistics which are the Product of an Odd Number of Independent Beta Random Variables. *Journal of Multivariate Analysis*, 89, 191-218.

Coelho, C. A., Arnold, B. C., Marques, F. J. (2010) - Near-exact distributions for certain likelihood ratio test statistics. *Journal of Statistical Theory and Practice* 4, 711-725.

Coelho, C. A., Alberto, R. P. (2012) - On the Distribution of the Product of Independent Beta Random Variables - Applications. *Technical Report, CMA*, 12.

Daniels, H. E. (1954) - Saddlepoint Approximations in Statistics. *Ann. Math. Statist.*, 25, 631-650.

Fox, C. (1961) - The G and H functions as symmetrical kernels. *Trans. Amer. Math. Soc.*, 98, 395-429

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

Marques, F. J., Coelho, C. A. (2015) - Testing elaborate block-structures in covariance matrices by splitting the null hypothesis - an overview. *Proceedings of the 60th ISI World Statistics Congress, 26-31 July 2015, Rio de Janeiro, Brazil*, 1-6.

Meijer, C. S. (1946) - On the G-function I–VIII. *Proc. Koninklijk Nederlandse Akademie van Weteenschappen* 49, 227-237, 344-356, 457-469, 632-641, 765-772, 936-943, 1063-1072, 1165-1175.

Moschopoulos, P. G. (1986) - New Representations for the Distribution Function of a Class of Likelihood Ratio Criteria. *Journal of Statistical Research*, 20, 13-20.

Nagar, D. K., Jain S. K., Gupta A. K. (1985) - Distribution of LRC for testing sphericity of a complex multivariate Gaussian model. *Internat. J. Math. & Mathematical Sci.*, 8, 555–562.

Springer, M. D. (1979) - *The Algebra of Random Variables*. New York: J. Wiley & Sons.

Tang, J., Gupta, A. K. (1984) - On the distribution of the product of independent beta random variables. *Statistics & Probability Letters*, 2, 165-168.

Tricomi, F. G., Erdélyi, A. (1951) - The asymptotic expansion of a ratio of Gamma functions. *Pacific Journal of Mathematics* 1, 133-142.