Combinando testes de Mardia e BHEP na avaliação duma hipótese multivariada de normalidade

Carlos Tenreiro, *tenreiro@mat.uc.pt*

CMUC, Departamento de Matemática, Universidade de Coimbra

1. Introdução

Sendo $X_1,\ldots,X_n,\ldots$ uma sucessão de cópias independentes dum vector d-dimensional e absolutamente contínuo X, com densidade de probabilidade f, desconhecida, o problema do teste duma hipótese multivariada de normalidade (MVN) é o de, com base em $X_1,\ldots,X_n$, testar a hipótese

$$H_0 : f \in \mathscr{N}_d,$$

contra uma hipótese alternativa geral, onde $\mathscr{N}_d$ é a família das densidades de probabilidade normais sobre $\mathbb{R}^d$. Este é um problema clássico na literatura estatística sobre o qual muito trabalho tem sido desenvolvido como atestam Mecklin e Mundfrom (2000) que referem a existência de cerca de cinquenta procedimentos para testar uma hipótese MVN. Apesar disso, este assunto continua a despertar o interesse dos investigadores como confirmam os trabalhos mais recentes de Liang et al. (2005), Mecklin e Mundfrom (2005), Székely e Rizzo (2005), Sürücü (2006), Arcones (2007), Farrel et al. (2007), Chiu e Liu (2009), Liang et al. (2009), Tenreiro (2009, 2011) e Ebner (2012). O facto de muito dos métodos estatísticos multivariados como a ANOVA, regressão multivariada, análise discriminante ou correlação canónica, dependerem da aceitação duma hipótese MVN pode explicar este interesse continuado. Para mais bibliografia sobre o tema veja-se Csörgő (1986) e os artigos de revisão de Henze (2002) e Mecklin e Mundfrom (2004).

Neste texto, baseado, no essencial, nos nossos trabalhos acima citados, iremos centrar a nossa atenção em testes duma hipótese MVN que gozam da propriedade natural de serem invariantes para transformações de localização e de escala dos dados. Sendo $T_n = T_n(X_1,\ldots,X_n)$ a estatística associada a um tal teste, vale assim a igualdade

$$T_n(AX_1+b,\ldots,AX_n+b) = T_n(X_1,\ldots,X_n),$$

para toda a matriz não-singular A e todo o vector $b \in \mathbb{R}^d$. Apesar de grande parte dos testes propostos na literatura não satisfazerem a propriedade anterior, satisfazem-na alguns dos mais utilizados testes de normalidade, como são os casos dos testes clássicos de Mardia e dos testes BHEP (Baringhaus-Henze-Epps-Pulley) a que faremos referência detalhada neste texto. Atendendo à propriedade de invariância, os pontos críticos destes procedimentos de teste podem ser estimados através de experiências de Monte Carlo sob H_0. Sobre outros testes invariantes para transformações afins dos dados veja-se Henze (2002) e Székely e Rizzo (2005).

2. Os testes de Mardia

De entre a vasta família de procedimentos de testes para uma hipótese MVN, os testes de Mardia (1970), baseados em medidas multivariadas de assimetria e curtose, desempenham um papel importante, estando entre os testes mais recomendados para testar uma hipótese MVN (ver Romeu e Ozturk, 1993; Mecklin e Mundfrom, 2005, e as referências bibliográficas respectivas). Denotando por

$$\bar{X}_n = n^{-1} \sum_{j=1}^{n} X_j \quad \text{e} \quad S_n = n^{-1} \sum_{j=1}^{n} (X_j - \bar{X}_n)(X_j - \bar{X}_n)',$$

a média e a matriz de covariância amostrais, respectivamente, as estatísticas MS (multivariate skewness) e MK (multivariate kurtosis) de Mardia são definidas por

$$\text{MS} = nb_{1,d}$$

e

$$\text{MK} = \sqrt{n}\,|\,b_{2,d} - d(d+2)|,$$

com

$$b_{1,d} = \frac{1}{n^2} \sum_{j,k=1}^{n} (Y_j' Y_k)^3 \quad \text{e} \quad b_{2,d} = \frac{1}{n} \sum_{j=1}^{n} (Y_j' Y_j)^2,$$

onde

$$Y_j = S_n^{-1/2}(X_j - \bar{X}_n),\, j = 1, \ldots, n,$$

são os resíduos standardizados e $S_n^{-1/2}$ é a raiz quadrada definida positiva da inversa da matriz de covariância amostral. Sob a hipótese H_0, valem as convergências em distribuição

$$nb_{1,d} \xrightarrow{d} 6\chi^2_{d(d+1)(d+2)/6}$$

e

$$\sqrt{n}\,(\,b_{2,d} - d(d+2)) \xrightarrow{d} N(0, 8d(d+2))$$

(Mardia, 1970), e, assim, o teste MS rejeita H_0 para valores grandes de $b_{1,d}$ enquanto que o teste MK rejeita H_0 para valores pequenos ou grandes de $b_{2,d}$.

Apesar de serem invariantes para transformações afins dos dados, os testes de Mardia, tal como a quase totalidade dos testes de normalidade propostos na literatura, não são convergentes para todas as distribuições alternativas $f \notin \mathcal{N}_d$. Mesmo para amostras de grande tamanho, potencialmente infinitas, existem distribuições alternativas que não são detectadas por tais testes. Denotando por

$$\beta_{1,d} = \mathrm{E}((X_1 - \mu)'\Sigma^{-1}(X_2 - \mu))^3 \quad \text{e} \quad \beta_{2,d} = \mathrm{E}((X_1 - \mu)'\Sigma^{-1}(X_1 - \mu))^2,$$

os parâmetros correspondentes às medidas amostrais anteriores de assimetria e de curtose, onde μ e Σ são a média e a matriz de covariância de X, Baringhaus e Henze (1992) mostraram que se $\mathrm{E}(X'X)^3 < \infty$, o teste baseado em MS é convergente se e só se $\beta_{1,d} > 0$, e Henze (1994) provou que se $\mathrm{E}(X'X)^4 < \infty$, o teste baseado em MK é convergente se e só se $\beta_{2,d} \neq d(d+2)$. Assim, apesar dos testes baseados em MS e MK poderem possuir uma potência elevada para alternativas com $\beta_{1,d} > 0$ ou $\beta_{2,d} \neq d(d+2)$, ambos os testes podem apresentar um fraco comportamento para alternativas com os mesmos coeficientes de assimetria e curtose que a distribuição normal multivariada, isto é, para distribuições $f \notin \mathscr{N}_d$ com $\beta_{1,d} = 0$ e $\beta_{2,d} = d(d+2)$. Este problema pode ser também sentido em outros testes baseados em estatísticas que combinam as medidas anteriores de assimetria e curtose de forma a produzirem um teste com boas propriedades globais (*omnibus test*), como são os casos dos testes propostos por Mardia e Foster (1983), Horswell e Looney (1992) ou Doornik e Hansen (1994).

3. Os testes BHEP

Em alternativa aos testes de Mardia, ou a combinações destes como as que acabámos de referir, podemos optar por utilizar testes convergentes para todas as distribuições alternativas, como são os casos dos testes BHEP (Baringhaus–Henze–Epps–Pulley), introduzidos por Baringhaus e Henze (1988) e Henze e Zirkler (1990), e que estendem o teste de normalidade de Epps e Pulley (1983) ao contexto multivariado. A estatística de teste BHEP é baseada na distância L_2 ponderada entre a função característica amostral associada aos resíduos standardizados

$$\Psi_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(i t' Y_j\right), \quad t \in \mathbb{R}^d,$$

e a função característica Φ da distribuição normal standard em $\mathbb{R}^d$ com densidade

$$\phi(x) = (2\pi)^{-d/2}\exp(-x'x/2), \quad x \in \mathbb{R}^d.$$

A função de peso é dada por

$$t \to |\Phi_h(t)|^2 = \exp(-h^2 t' t),$$

onde Φ_h é a função característica de $\phi_h(\cdot) = \phi(\cdot/h)/h^d$ e h é um número real estritamente positivo que deve ser escolhido pelo utilizador. Assim, o teste BHEP é baseado na estatística

$$\mathrm{B}(h) = n\int |\Psi_n(t) - \Phi(t)|^2 |\Phi_h(t)|^2 dt = (2\pi)^d \frac{1}{n}\sum_{i,j=1}^{n} Q(Y_i, Y_j; h),$$

com

$$Q(u,v;h) = \phi_{(2h^2)^{1/2}}(u-v) - \phi_{(1+2h^2)^{1/2}}(u) - \phi_{(1+2h^2)^{1/2}}(v) + \phi_{(2+2h^2)^{1/2}}(0),$$

para $u, v \in \mathbb{R}^d$. A simplicidade da expressão anterior para $\mathrm{B}(h)$, justifica a escolha da função de peso considerada. Reparemos ainda que a estatística de teste depende das observações através das quantidades $||Y_i - Y_j||^2$ e $||Y_i||^2$, onde $||\cdot||$ representa a norma euclidiana em $\mathbb{R}^d$. Tais quantidades dependem apenas de S_n^{-1}, não sendo assim sequer necessário calcular $S_n^{-1/2}$ para obter $\mathrm{B}(h)$. Para uma referência recente sobre testes de ajustamento baseados na função característica ver Jiménez-Gamero et al. (2009).

O comportamento assintótico de $\mathrm{B}(h)$ sob a hipótese nula, sob uma alternativa fixa e sob uma sucessão de alternativas locais, foi estudado por diversos autores como são os casos de Baringhaus e Henze (1988), Csörgő (1989), Henze e Zirkler (1990) e Henze e Wagner (1997). Em particular, para cada $h > 0$, $\mathrm{B}(h)$ possui como distribuição nula assintótica uma soma ponderada de qui-quadrados independentes sendo o teste associado convergente para toda a alternativa fixa.

É interessante notar que no caso da densidade f ser de quadrado integrável, a estatística de teste anterior pode também ser interpretada como sendo baseada na distância L_2 entre o estimador da densidade de Parzen-Rosenblatt obtido a partir dos resíduos standardizados, com núcleo (kernel) $K = \phi$ e janela (*bandwidth*) h, e a convolução $K_h * \phi$, que pode ser vista como uma aproximação de ϕ quando h tende para zero (ver Henze e Zirkler, 1990; Bowman e Foster, 1993; Fan, 1998). Neste sentido, quando tomamos $h = h_n \to 0$, $n \to \infty$, o teste baseado em $\mathrm{B}(h)$ deve ser interpretado como um teste baseado na densidade de probabilidade de X e não na sua função característica. Deixando o parâmetro h de ser fixo e passando a desempenhar o papel de janela do estimador do núcleo da densidade, as propriedades assintóticas do teste baseado em $\mathrm{B}(h)$ são distintas das que acima descrevemos. No entanto, o teste resultante continua a ser convergente para toda a distribuição alternativa. Os testes de ajustamento baseados no estimador do núcleo da densidade de probabilidade foram primeiramente estudados por Bickel e Rosenblatt (1973). Para mais detalhes sobre estes testes vejam-se os trabalhos de Rosenblatt (1975), Fan (1994), Tenreiro (1996, 2007), Gouriéroux e Tenreiro (2001) e Henze (2002).

4. A selecção do parâmetro h nos testes BHEP

Apesar dos testes BHEP serem convergentes para todas as distribuições alternativas, independentemente do valor tomado para o parâmetro $h > 0$, a sua potência depende fortemente da escolha de h (cf. Henze e Wagner, 1997; Tenreiro, 2009). As diferentes escolhas de h consideradas na literatura foram analisadas em Tenreiro (2009) que, com base num vasto estudo de simulação para dimensões $2 \leq d \leq 15$, sugere duas escolhas empíricas para h. Assim, a janela

$$h = h_{\mathrm{L}} := 0.448 + 0.026\,d$$

mostrou ser adequada para alternativas com 'caudas leves' ou alternativas aproximadamente simétricas, e a janela

$$h = h_{\mathrm{P}} := 0.928 + 0.049\,d,$$

mostrou-se adequada para alternativas com 'caudas pesadas' ou alternativas moderadamente assimétricas. Estas escolhas estão de acordo com uma interpretação heurística das propriedades de potência do teste em termos da janela h. Para valores grandes de h, a ponderação $t \to \exp(-h^2 t't)$ coloca a maior parte da sua massa numa vizinhança da origem onde a função característica reflecte o comportamento da cauda da distribuição, sendo por isso de esperar que o teste BHEP seja sensível a alternativas com 'caudas pesadas'. Por razões análogas, será de esperar que o teste possa ser mais sensível para alternativas com 'caudas leves' para valores pequenos de h. Na ausência de informação adicional sobre o tipo de alternativa em causa, recomenda-se a utilização da janela combinada

$$h = \bar{h} := \frac{1}{2} h_{\mathrm{L}} + \frac{1}{2} h_{\mathrm{P}},$$

que conduz a um teste com boas propriedades de potência para um vasto conjunto de distribuições alternativas (cf. Tenreiro, 2009).

5. Combinando testes de Mardia e BHEP

Apesar das boas propriedades reveladas pelo teste BHEP baseado em $\mathrm{B}(\bar{h})$, para diversas distribuições alternativas este teste é claramente superado por um dos testes de Mardia. O teste MS revela-se particularmente eficiente na detecção de alternativas assimétricas ou com 'caudas pesadas', enquanto que o teste MK mostra-se especialmente eficaz na detecção de alternativas com 'caudas leves' (cf. Henze e Zirkler, 1990; Romeu e Ozturk, 1993). A ideia de combinar os testes de Mardia e BHEP de forma a obter um teste que possa usufruir das boas propriedades de cada um dos testes intervenientes na combinação surge assim de forma natural, sendo a mesma analisada em Tenreiro (2011) utilizando um método, considerado em Fromont e Laurent (2006), que pode ser interpretado como um melhoramento do método de Bonferroni clássico.

Sendo $T_{n,h}, h \in H$, um conjunto finito de estatísticas de testes invariantes para transformações afins dos dados, o teste múltiplo proposto rejeita a hipótese MVN se pelo menos uma das estatísticas $T_{n,h}$ for maior que o seu quantil de ordem $1 - u_{n,\alpha}$ sob a hipótese nula, onde $u_{n,\alpha}$ é calibrado de forma que o teste múltiplo tenha um nível de significância não superior ao nível nominal α fixado à partida. Mais precisamente, sendo $c_{n,h}(u)$ o quantil de ordem $1-u$ da estatística de teste $T_{n,h}$ sob H_0 e considerando a estatística corrigida

$$\mathbf{T}_n(u) = \max_{h \in H} \left(T_{n,h} - c_{n,h}(u) \right),$$

o teste múltiplo rejeita a hipótese H_0 sempre que

$$\mathbf{T}_n(u_{n,\alpha}) > 0$$

onde

$$u_{n,\alpha} = \sup \{ u \in]0,1[: \mathrm{P}_0(\mathbf{T}_n(u) > 0) \leq \alpha \},$$

e P_0 é a distribuição normal standard sobre $\mathbb{R}^d$. Na prática, o nível $u_{n,\alpha}$, segundo o qual cada um dos testes baseados nas estatísticas $T_{n,h}$ é aplicado, é estimado através de experiências de Monte Carlo sob H_0.

Em Tenreiro (2011) o método geral anterior é utilizado para combinar os testes baseados nas estatísticas $T_{n,1} = \mathrm{MS}$, $T_{n,2} = \mathrm{MK}$, $T_{n,3} = \mathrm{B}(h_{\mathrm{L}})$ e $T_{n,4} = \mathrm{B}(h_{\mathrm{P}})$. Outras combinações de testes invariantes duma hipótese MVN são naturalmente possíveis. Dum ponto de vista teórico, o teste resultante da combinação dos quatro testes anteriores é convergente para toda distribuição alternativa e, para n fixo, o seu nível de significância não é superior ao nível nominal $\alpha \in]0,1[$ fixado à partida. O desempenho a distância finita do procedimento múltiplo foi avaliado sob H_0, revelando o teste possuir um nível de significância efectivo muito próximo do nível nominal α, e sob um vasto conjunto de distribuições alternativas que são habitualmente consideradas na literatura em estudos deste tipo. Como seria de esperar, fixada uma distribuição alternativa, o teste múltiplo proposto nunca é o melhor dos testes incluídos na combinação. No entanto, ele herda as boas propriedades de cada um dos testes envolvidos no procedimento múltiplo, revelando um bom desempenho para todas as alternativas consideradas. Tendo em conta que numa situação real a formulação duma hipótese alternativa é em geral impossível, a propriedade anterior é muito interessante não sendo a mesma partilhada por nenhum dos testes envolvidos na combinação considerada. Além disso, o teste mostra um bom desempenho global quando comparado com os mais recomendados testes duma hipótese MVN, o que nos leva a considerá-lo uma alternativa válida aos diversos testes propostos na literatura. Uma função escrita em R para implementar este teste está disponível no endereço http://www.mat.uc.pt/∼tenreiro/publications.

Bibliografia

Arcones, M.A., 2007. Two tests for multivariate normality based on the characteristic function. Math. Methods Statist. 16, 177–201.

Baringhaus, L., Henze, N., 1988. A consistent test for multivariate normality based on the empirical characteristic function. Metrika 35, 339–348.

Baringhaus, L., Henze, N., 1992. Limit distributions for Mardia's measure of multivariate skewness. Ann. Statist. 20, 1889–1902.

Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Ann. Statist. 1, 1071–1095.

Bowman, A.W., Foster, P.J., 1993. Adaptive smoothing and density-based tests of multivariate normality. J. Amer. Statist. Assoc. 88, 529–537.

Chiu, S.N., Liu, K.I., 2009. Generalized Cramér-von Mises goodness-of-fit tests for multivariate distributions. Comput. Statist. Data Anal. 53, 3817–3834.

Csörgő, S., 1986. Testing for normality in arbitrary dimension. Ann. Statist. 14, 708–723.

Csörgő, S., 1989. Consistency of some tests for multivariate normality. Metrika 36, 107–116.

Doornik, J.A., Hansen, H., 1994. An omnibus test for univariate and multivariate normality. Working Paper, Nuffield College, Oxford.

Ebner, B., 2012. Asymptotic theory for the test for multivariate normality by Cox and Small. J. Multivariate Anal. 111, 368–379.

Epps, T.W., Pulley, L.B., 1983. A test for normality based on the empirical characteristic function. Biometrika 70, 723–726.

Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel method. Econometric Theory 10, 316–356.

Fan, Y., 1998. Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. Econometric Theory 14, 604–621.

Farrel, P.J., Salibian-Barrera, M., Naczk, K., 2007. On tests for multivariate normality and associated simulation studies. J. Stat. Comput. Simul. 77, 1065–1080.

Fromont, M., Laurent, B., 2006. Adaptive goodness-of-fit tests in a density model. Ann. Statist. 34, 680–720.

Gouriéroux, C., Tenreiro, C., 2001. Local power properties of kernel based goodness of fit tests. J. Multivariate Anal. 78, 161–190.

Henze, N., Zirkler, B., 1990. A class of invariante consistent tests for multivariate normality. Comm. Stat. Theory Methods 19, 3595–3617.

Henze, N., 1994. On Mardia's kurtosis test for multivariate normality. Comm. Statist. Theory Methods 23, 1047–1061.

Henze, N., 2002. Invariant tests for multivariate normality: a critical review. Statist. Papers 43, 467–506.

Henze, N., Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivariate Anal. 62, 1–23.

Horswell, R.L., Looney, S.W., 1992. A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis. J. Stat. Comput. Simul. 42, 21–38.

Jiménez-Gamero, M.D., Alba-Fernández, V., Muñoz-García, J., Chalco-Cano, Y., 2009. Goodness-of-fit tests based on empirical characteristic functions. Comput. Statist. Data Anal. 53, 3957–3971.

Liang, J., Pan, W.S.Y., Yang, Z.-H., 2005. Characterization-based Q-Q plots for testing multinormality. Statis. Probab. Lett. 70, 183–190.

Liang, J., Tang, M.-L., Chan, P.S., 2009. A generalized Shapiro–Wilk W statistic for testing high-dimensional normality. Comput. Statist. Data Anal. 53, 3883–3891.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

Mardia, K.V., Foster, K., 1983. Omnibus tests of multinormality based on skewness and kurtosis. Comm. Statist. Theory Methods 12, 207–221.

Mecklin, C.J., Mundfrom, D.J., 2000. Comparing of the power of classical and newer tests of multivariate normality. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 24–28, 2000.

Mecklin, C.J., Mundfrom, D.J., 2004. An appraisal and bibliography of tests for multivariate normality. Int. Stat. Rev. 72, 123–138.

Mecklin, C.J., Mundfrom, D.J., 2005. A Monte Carlo comparison of Type I and Type II error rates of tests of multivariate normality. J. Stat. Comput. Simul. 75, 93–107.

Romeu, J.L., Ozturk, A., 1993. A comparative study of goodness-of-fit tests for multivariate normality. J. Multivariate Anal. 46, 309–334.

Rosenblatt, M., 1975. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. Ann. Statist. 3, 1–14.

Sürücü, B., 2006. Goodness-of-fit tests for multivariate distributions. Comm. Statist. Theory Methods 35, 1319–1331.

Székely, G.J., Rizzo, M.L., 2005. A new test for multivariate normality. J. Multivariate Anal. 93, 58–80.

Tenreiro, C. 1996. Tests d'ajustement à une densité fondés sur un estimateur non paramétrique à noyau pour des observations dépendantes. Ann. Économ. Statist. 43, 129–148.

Tenreiro, C., 2007. On the asymptotic behaviour of location-scale invariant Bickel-Rosenblatt tests. J. Statist. Plann. Inference 137, 103–116. Erratum: 139, 2115, 2009.

Tenreiro, C., 2009. On the choice of the smoothing parameter for the BHEP goodness-of-fit test. Comput. Statist. Data Anal. 53, 1038–1053.

Tenreiro, C., 2011. An affine invariant multiple test procedure for assessing multivariate normality. Comput. Statist. Data Anal. 55, 1980–1992.