多変量標本歪度による統計量の分布について

東京理科大・理　　岡本 直也
東京理科大・理・院　小泉 和之
東京理科大・理　　瀬尾　隆

1. はじめに

多次元データが多変量正規分布に従っているという多変量正規性の検定法の中に，多変量歪度と多変量尖度を用いた手法がある．Mardia (1970) と Srivastava (1984) は，それぞれ異なる多変量標本歪度と多変量標本尖度の定義を与え，その標本分布について議論している．Seo and Ariga (2006) では，多変量標本尖度を用いた統計量の分布関数に対する漸近展開の結果から，新たな正規性検定統計量を導出した．本研究では，Srivastava (1984) によって定義された多変量標本歪度に対する期待値と分散を与え，多変量標本歪度を用いた新たな正規性検定統計量を導出した．また，導出された検定統計量の近似上側パーセント点を Cornish-Fisher 展開により求めた．

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ を互いに独立な多変量正規分布 $N_p(\boldsymbol{\mu}, \Sigma)$ に従う p 次元確率ベクトルとする．また，標本平均ベクトル $\overline{\boldsymbol{x}}$ と標本分散共分散行列 S をそれぞれ，

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{x}_j,$$

$$S = \frac{1}{N}\sum_{j=1}^{N} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})'$$

と定義する．標本分散共分散行列は，直交行列 $H = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$，$D_\omega = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_p)$（対角成分は S の固有値）を用いて $S = HD_\omega H'$ と表される．

このとき，多変量標本歪度 $b_{1,p}^2$ は，$y_{ij} = \boldsymbol{h}_i' \boldsymbol{x}_j (i = 1, 2, \ldots, p, j = 1, 2, \ldots, N)$，$\overline{y}_i = (1/N)\sum_{j=1}^{N} y_{ij}$ を用いて次のように定義される．

$$b_{1,p}^2 = \frac{1}{p}\sum_{i=1}^{p}\left\{\frac{1}{\omega_i^{\frac{3}{2}}}\sum_{j=1}^{N}\frac{(y_{ij}-\overline{y}_i)^3}{N}\right\}^2.$$

Srivastava (1984) は $b_{1,p}^2$ の漸近的な期待値とカイ二乗近似統計量を次のように導出している．

$$\mathrm{E}[b_{1,p}^2] = \frac{6}{N}, \qquad \frac{Np}{6} b_{1,p}^2 \sim \chi_p^2.$$

2. 改良正規性検定統計量

本研究では，標本数が少ない場合でも近似精度のよいカイ二乗近似統計量を導出するため，$m_{\nu i} = N^{-1}\sum_{j=1}^{N}(y_{ij} - \overline{y}_i)^\nu$ に関するいくつかの期待値を求め，$b_{1,p}^2$ の期待値と分散を次のように導出した．

$$\mathrm{E}[b_{1,p}^2] = \frac{6(N-2)}{(N+1)(N+3)},$$

$$\mathrm{Var}[b_{1,p}^2] = \frac{72N(N-2)(N^3+37N^2+11N-313)}{p(N+1)^2(N+3)^2(N+5)(N+7)(N+9)}.$$

この結果より，十分大きな N に対して，改良正規性検定統計量

$$T \equiv \frac{p(N+1)(N+3)}{6(N-2)} b_{1,p}^2 \sim \chi_p^2$$

を導出した.

3. 近似上側パーセント点

次に，Cornish-Fisher 展開を用いて T の分布の上側パーセント点を近似的に導出するため，$(T-\mathrm{E}[T])/\sqrt{\mathrm{Var}[T]}$ の特性関数を展開することにより，1 次から 5 次キュムラントをそれぞれ,

$$\kappa_1 = 0, \quad \kappa_2 = 1, \quad \kappa_3 = \frac{2\sqrt{2}}{\sqrt{p}} + \frac{54\sqrt{2}}{N\sqrt{p}} + \frac{783\sqrt{2}}{N^2\sqrt{p}} + O(N^{-3}),$$

$$\kappa_4 = \frac{12}{p} + \frac{864}{Np} + \frac{45360}{N^2p} + O(N^{-3}), \quad \kappa_5 = \frac{48\sqrt{2}}{p\sqrt{p}} + \frac{6480\sqrt{2}}{Np\sqrt{p}} + \frac{638280\sqrt{2}}{N^2p\sqrt{p}} + O(N^{-3})$$

と求めた．また，標準正規分布に従う確率変数を z とし,

$$B_1(z) = \frac{1}{6}\kappa_3(z^2-1), \quad B_2(z) = \frac{1}{24}\kappa_4(z^3-3z) - \frac{1}{36}\kappa_3^2(2z^3-5z),$$

$$B_3(z) = \frac{1}{120}\kappa_5(z^4-6z^2+3) - \frac{1}{24}\kappa_3\kappa_4(z^4-5z^2+2) + \frac{1}{324}\kappa_3^3(12z^4-53z^2+17)$$

とおくことにより，統計量 T の分布の上側 100α パーセント点を

$$C_\alpha \approx p + \sqrt{\mathrm{Var}[T]}\,(z_\alpha + B_1(z_\alpha) + B_2(z_\alpha) + B_3(z_\alpha))$$

（z_α は標準正規分布の上側 100α パーセント点）と導出した.

4. 近似精度

得られた正規性検定統計量 T のカイ二乗分布への近似精度を調べるためにモンテカルロ・シミュレーションを行い，検定統計量の帰無分布に対する期待値，分散，上側パーセント点，下側パーセント点により検定統計量の漸近的性質を数値的に評価した．また，シミュレーションにより求められた T の分布の上側パーセント点と Cornish-Fisher 展開により導出された近似上側パーセント点の誤差を評価した．その結果，比較的小さな標本数の場合であっても，検定統計量 T は Srivastava (1984) の検定統計量よりも正確な検定が行えることを確認した.

参考文献

[1] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

[2] Seo, T. and Ariga, M. (2006). On the distribution of kurtosis test for multivariate normality. Technical Report No.06-03, Statistical Research Group, Hiroshima University.

[3] Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics* & *Probability letters*, **2**, 263–267.