数理解析研究所講究録
第345巻 1979年 64-71

64

多次元正規性の検定について

塩野義解析センター　上坂浩之
後藤昌司

1. はじめに

多くの統計解析法は正規性の仮定の下に展開されているが、現実のデータはむしろ正規でない場合が多い。手法の有効性と妥当性に関する知識を得るために、その前提となる条件と現実のデータとの溝を測ることは有益である。プログラム・パッケージNISANではこれを重視してdata investigationの機能をとりあげており、(Asano et al.〔1〕)、その1つとして正規性の検定が考えられている(Goto, Uesaka & Asano〔4〕)。この研究はNISANの方法論研究の1つとして進めているものである。

正規性検定では1次元分布の場合、その3次と4次のキュミュラントにもとづく検定の有効なことが広く認められている。同様に多次元分布においても、3次と4次のキュミュラントの利用は有益と思われる。実際、これまでに提案されている多次元正規性の検定のうち、いくつかは直接あるいは間接にキュミュラントを扱っている。Mardia〔7〕,〔8〕は多次元の歪度と尖度を、HotellingのT^2統計量のロバストネスの問題に関連して定義し、多変量正規性の検定への応用を論じている。Dahiya & Gurland〔3〕は2変数の場合に、各周辺分布の3次と4次のキュミュラントがすべてゼロであることを検定する方法を提案した。また竹内〔9〕

は平均ベクトルと分散共分散行列の実現値上に条件づけられた3次キュミュラントの不偏推定量の条件つき分散と共分散を与え、これにもとづく検定を提案した。Cox & Small〔2〕は正規分布を特徴づける回帰の線形性を検定する種々の方法を論じた。これらの検定は3次あるいは4次のキュミュラントに直接関係しており、大標本の下での近似分布が与えられている。Mardiaの検定は1次変換に対して不変であるが、Dahiya & Gurlandおよび竹内の検定は特定のキュミュラントだけを用いるため座標変数に依存する。これに対して、すべての3次、あるいは4次のキュミュラントを推定し、それらの分散共分散行列を求め2次形式の統計量をつくるならば正則1次変換に対して不変な方式が得られる。さらに、個々のキュミュラントの大きさを見ることにより変数あるいは変数の組の特徴を知ることが可能になる。次にこの検定法とその若干の性質を述べる。

2. キュミュラント検定

p次元確率ベクトル$\boldsymbol{X}_\ell = (X_{\ell 1}, \cdots, X_{\ell p})'$, $\ell = 1, \cdots, n$ は互いに独立に同一のp次元分布に従うとする。標本平均ベクトルを$\overline{\boldsymbol{X}} = (\overline{X}_1, \cdots, \overline{X}_p)'$, 標本分散共分散行列を $S = (s_{ij})$, ここに

$$s_{ij} = \frac{1}{n} \sum_{\ell=1}^{n} (X_{\ell i} - \overline{X}_i)(X_{\ell j} - \overline{X}_j), \quad 1 \leq i, j \leq p$$

とする。$\boldsymbol{X}_\ell$ の母分散共分散行列を $\Sigma = (\sigma_{ij})$, 3次キュミュラントを $\kappa_{i_1 i_2 i_3}$, その不偏推定量を $k_{i_1 i_2 i_3}$, $1 \leq i_1 \leq i_2 \leq i_3 \leq p$ とし、$\boldsymbol{\kappa}_3 = (\kappa_{i_1 i_2 i_3}, 1 \leq i_1 \leq i_2 \leq i_3 \leq p)'$, $\boldsymbol{k}_3 = (k_{i_1 i_2 i_3}, 1 \leq i_1 \leq i_2 \leq i_3 \leq p)'$ とかく。

$$k_{i_1 i_2 i_3} = \frac{n}{(n-1)(n-2)} \sum_{\ell=1}^{n} (X_{\ell i_1} - \overline{X}_{i_1})(X_{\ell i_2} - \overline{X}_{i_2})(X_{\ell i_3} - \overline{X}_{i_3})$$

である。いま $\Sigma_3 = (\sigma_{i_1 i_2 i_3, j_1 j_2 j_3})$

$$\sigma_{i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3} = \Sigma_3' \sigma_{i_1\,j_1}\sigma_{i_2\,j_2}\sigma_{i_3\,j_3}$$

とおく、ここに Σ_3' は3次の置換 $\begin{pmatrix} i_1\,i_2\,i_3 \\ j_1\,j_2\,j_3 \end{pmatrix}$ 全体にわたる和を表わす。正規性の仮定の下で

$$E(k_{i_1\,i_2\,i_3}) = \kappa_{i_1\,i_2\,i_3} = 0$$

$$\mathrm{cov}(k_{i_1\,i_2\,i_3}, k_{j_1\,j_2\,j_3}) = \frac{n}{(n-1)(n-2)}\sigma_{i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3}$$

である。$\sqrt{n}\,\boldsymbol{k}_3$ は漸近的に $N(\boldsymbol{O}, \Sigma_3)$ に従う。σ_{ij} を S_{ij} でおきかえて Σ_3 の推定量とし、$\hat{\Sigma}_3$ とおく。$\hat{\Sigma}_3$ は $n\to\infty$ のとき Σ_3 に確率収束する。それ故 $n\boldsymbol{k}_3'\hat{\Sigma}_3^{-1}\boldsymbol{k}_3$ は $n\to\infty$ のとき、正規性の仮定の下で自由度 $f = p(p+1)(p+2)/6$ のカイ二乗分布に従う。実際には $\boldsymbol{k}_3$ のの分散共分散行列 $\{n/(n-1)(n-2)\}\Sigma_3$ の推定量を用いて、

$$Q_3 = \frac{(n-1)(n-2)}{n}\boldsymbol{k}_3'\hat{\Sigma}_3^{-1}\boldsymbol{k}_3$$

と定義する。

竹内〔9〕は $\overline{\boldsymbol{X}}$ と S の実現値上に条件づけられた $k_{i_1\,i_2\,i_3}$ と $k_{j_1\,j_2\,j_3}$ の共分散を求めている。これは次式で与えられる。

$$\mathrm{cov}(k_{i_1\,i_2\,i_3}, k_{j_1\,j_2\,j_3} \mid \overline{\boldsymbol{X}}, S)$$
$$= \frac{n(n-1)}{(n-2)^2(n+1)(n+3)}\{(n+1)A(i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3)$$
$$- 2B(i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3)\}$$

ここに

$$A(i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3) = S_{i_1\,j_1}S_{i_2\,j_2}S_{i_3\,j_3} + S_{i_1\,j_1}S_{i_2\,j_3}S_{i_3\,j_2}$$
$$+ S_{i_1\,j_2}S_{i_2\,j_1}S_{i_3\,j_3} + S_{i_1\,j_2}S_{i_2\,j_3}S_{i_3\,j_1}$$
$$+ S_{i_1\,j_3}S_{i_2\,j_2}S_{i_3\,j_1} + S_{i_1\,j_3}S_{i_2\,j_1}S_{i_3\,j_2}$$

$$B(i_1\,i_2\,i_3\,,\,j_1\,j_2\,j_3) = S_{i_1\,i_2}S_{j_1\,j_2}S_{i_3\,j_3} + S_{i_1\,i_2}S_{i_1\,j_3}S_{i_3\,j_2}$$
$$+ S_{i_1\,i_2}S_{j_2\,j_3}S_{i_3\,j_1} + S_{i_1\,i_3}S_{j_1\,j_2}S_{i_2\,j_3}$$

$$+ S_{i_1 i_3} S_{j_1 j_3} S_{i_2 j_2} + S_{i_1 i_3} S_{j_2 j_3} S_{i_2 j_1}$$

$$+ S_{i_2 i_3} S_{j_1 j_2} S_{i_1 j_3} + S_{i_2 i_3} S_{j_1 j_3} S_{i_1 j_2}$$

$$+ S_{i_2 i_3} S_{j_2 j_3} S_{i_1 j_1}$$

である。この条件つき分散共分散行列を $\{ n/(n-1)(n-2) \} \hat{\Sigma}_3$ の代わりに用いた2次形式統計量を Q_3^* とする。Q_3^* も $n \to \infty$ のとき自由度 f のカイ二乗分布に従う。

4次キュミュラントに関する検定も全く同じようにして進められる。4次キュミュラントを $\kappa_{i_1 i_2 i_3 i_4}$，その不偏推定量を $k_{i_1 i_2 i_3 i_4}$ とかく，$1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq p$。これらをそれぞれベクトルで $\boldsymbol{\kappa}_4 = (\kappa_{i_1 i_2 i_3 i_4},\ 1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq p)'$，$\boldsymbol{k}_4 = (k_{i_1 i_2 i_3 i_4},\ 1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq p)'$ とかく。$k_{i_1 i_2 i_3 i_4}$ は

$$k_{i_1 i_2 i_3 i_4} = \frac{n^2}{(n-1)(n-2)(n-3)} \{ (n+1) m_{i_1 i_2 i_3 i_4} - (n-1) B \}$$

ここに

$$B = S_{i_1 i_2} S_{i_3 i_4} + S_{i_1 i_3} S_{i_2 i_4} + S_{i_1 i_4} S_{i_2 i_3}$$

$$m_{i_1 i_2 i_3 i_4} = \frac{1}{n} \Sigma_{\ell=1}^{n} (X_{\ell i_1} - \overline{X}_{i_1})(X_{\ell i_2} - \overline{X}_{i_2})(X_{\ell i_3} - \overline{X}_{i_3})(X_{\ell i_4} - \overline{X}_{i_4})$$

である。また

$$\sigma_{i_1 i_2 i_3 i_4,\ j_1 j_2 j_3 j_4} = \Sigma_4' \sigma_{i_1 j_1} \sigma_{i_2 j_2} \sigma_{i_3 j_3} \sigma_{i_4 j_4}$$

とおく，ここに Σ_4' は4次の置換 $\begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ j_1 & j_2 & j_3 & j_4 \end{pmatrix}$ 全体にわたる和を表わす。正規性の仮定の下で

$$\mathrm{cov}(k_{i_1 i_2 i_3 i_4}, k_{j_1 j_2 j_3 j_4}) = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sigma_{i_1 i_2 i_3 i_4,\ j_1 j_2 j_3 j_4}$$

である。$\Sigma_4 = (\sigma_{i_1 i_2 i_3 i_4,\ j_1 j_2 j_3 j_4})$ とおき σ_{ij} を S_{ij} でおきかえて Σ_4 の推定量とし $\hat{\Sigma}_4$ とする。こうして4次キュミュラントの検定統計量

$$Q_4 = \frac{(n-1)(n-2)(n-3)}{n(n+1)} \boldsymbol{k}_4 \hat{\Sigma}_4^{-1} \boldsymbol{k}_4$$

を定義する。Q_4 は $n \to \infty$ のとき、正規性の仮定の下で自由度 $f = n(n+1)(n+2)(n+3)/24$ の

カイ二乗分布に従う。

3. キュミュラント検定の性質

多変量正規分布の1次変換に対する不変性から、多次元正規性検定の1次変換に対する不変性は重要である。観測された変量にとくに興味のある場合、あるいはそれらの特定のキュミュラントに興味のある場合はそれらに対する直接的な検定を採用すればよい。ここではまず前節で定義した Q_3 と Q_4 が正則1次変換の下で不変であることを示す。まず次の補題を証明する。

補題　　p次元確率ベクトル $\boldsymbol{X}$ は k（≥ 4）次キュミュラントをもつとする。$A=(a_{ij})$ は正則な p×p 実行列とし、$\boldsymbol{Y}=A\boldsymbol{X}$ を考える。このとき $\boldsymbol{Y}$ の3次，4次キュミュラント $\kappa^*_{i_1 i_2 i_3}$ と $\kappa^*_{i_1 i_2 i_3 i_4}$ はそれぞれ $\boldsymbol{X}$ の3次と4次のキュミュラントにより

$$\kappa^*_{i_1 i_2 i_3}=\sum^{p}_{j_1,j_2,j_3=1} a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3}\,\kappa_{j_1 j_2 j_3}$$

$$\kappa^*_{i_1 i_2 i_3 i_4}=\sum^{p}_{j_1,j_2,j_3,j_4=1} a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3} a_{i_4 j_4}\,\kappa_{j_1 j_2 j_3 j_4}$$

と表わされる。

証明　$\boldsymbol{X}$ のキュミュラント母関数を $\varphi(\boldsymbol{t})$，$\boldsymbol{Y}$ のそれを $\Psi(\boldsymbol{u})$ とすると $\Psi(\boldsymbol{u})=\varphi(A'\boldsymbol{u})$ である。したがって $\boldsymbol{t}=A'\boldsymbol{u}$ とおいて $\partial^3\varphi(\boldsymbol{u})/\partial u_{i_1}\partial u_{i_2}\partial u_{i_3}$ を求めると

$$\frac{\partial^3\Psi(\boldsymbol{u})}{\partial u_{i_1}\partial u_{i_2}\partial u_{i_3}}=\sum^{p}_{j_1,j_2,j_3=1}\frac{\partial^3\varphi(\boldsymbol{t})}{\partial t_{j_1}\partial t_{j_2}\partial t_{j_3}}\cdot\frac{\partial t_{j_1}}{\partial u_{i_1}}\frac{\partial t_{j_2}}{\partial u_{i_2}}\frac{\partial t_{j_3}}{\partial u_{i_3}}$$

$$=\sum^{p}_{j_1,j_2,j_3=1} a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3}\frac{\partial^3\varphi(\boldsymbol{t})}{\partial t_{j_1}\partial t_{j_2}\partial t_{j_3}}$$

である。$\boldsymbol{u}=0$ とおいて両辺を比較し所期の結果が従う。4次キュミュラントについても同様にして示される。

補題とキュミュラント不偏推定量 $k_{i_1 i_2 i_3}$ の一意性により，$\kappa^*_{i_1 i_2 i_3}$ の不偏推定量 $k^*_{i_1 i_2 i_3}$ は

$\Sigma^{p}_{j_1,j_2,j_3=1}\ a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3}\ k_{j_1 j_2 j_3}$ に等しいことがわかる。$\gamma_{i_1 i_2 i_3, j_1 j_2 j_3} = a_{i_1 j_1} a_{i_2 j_2} a_{i_3 j_3}$ とおいて行列 $\Gamma = (\gamma_{i_1 i_2 i_3,\ j_1 j_2 j_3})$ を定義する。$\boldsymbol{k}^{*}_{3} = \Gamma \boldsymbol{k}_3$ である。したがって確率ベクトル $\boldsymbol{Y}$ に対する Σ_3 を $\Sigma^{*}_{3} = (\sigma^{*}_{i_1 i_2 i_3,\ j_1 j_2 j_3})$ と表わすとき、$\hat{\Sigma}^{*}_{3} = \Gamma \hat{\Sigma}_3 \Gamma'$ を示せば Q_3 の不変性が示される。実際に、$\boldsymbol{Y}$ の母分散，共分散を σ^{*}_{ij}，その推定量を S^{*}_{ij} とかくとき、

$$\hat{\sigma}^{*}_{i_1 i_2 i_3, j_1 j_2 j_3} = \Sigma' S^{*}_{i_1 j_1} S^{*}_{i_2 j_2} S^{*}_{i_3 j_3}$$

$$= \sum^{p}_{i_1', i_2', i_3'=1} \sum^{p}_{j_1', j_2', j_3'=1} \gamma_{i_1 i_2 i_3,\ i_1' i_2' i_3'} \gamma_{j_1 j_2 j_3,\ j_1' j_2' j_3'} \Sigma'_3 S_{i_1' j_1'} S_{i_2' j_2'} S_{i_3' j_3'}$$

であるから、$\hat{\Sigma}^{*}_{3} = \Gamma \hat{\Sigma}_3 \Gamma'$ がなりたつ。

同じようにして Q^{*}_{3}，Q_4 の不変性が示される。

次に他の多変量正規性の検定法との相違を考える。Mardia の多次元歪度 β_{1p} は

$$\beta_{1p} = \sum^{p}_{i_1, i_2, i_3=1} \sum^{p}_{j_1, j_2, j_3=1} \sigma^{i_1 j_1} \sigma^{i_2 j_2} \sigma^{i_3 j_3} \kappa_{i_1 i_2 i_3} \kappa_{j_1 j_2 j_3}$$

で定義される。これは 1 次変換に対して不変であるから $\sigma_{ij} = \delta_{ij}$ と変換して考えると、

$$\beta_{1p} = \Sigma^{p}_{i=1} \kappa^{2}_{iii} + 3\Sigma_{i_1 < i_2} \kappa^{2}_{i_1 i_1 i_2} + 6\Sigma_{i_1 < i_2 < i_3} \kappa^{2}_{i_1 i_2 i_3}$$

であり、3 次キュミュラントの加重平方和である。したがって $\kappa_3 = 0$ と $\beta_{1p} = 0$ は同値であるが、検定においては検出力の相違が予想される。Mardia の多次元尖度 β_{2p} は

$$\beta_{2p} = \Sigma^{p}_{i_1, i_2=1} \Sigma^{p}_{j_1, j_2=1} \sigma^{i_1 i_2} \sigma^{j_1 j_2} \kappa_{i_1 i_2 j_1 j_2} + p(p+2)$$

で定義され、4 次キュミュラントの加重和で与えられる。したがって $\beta_{2p} = p(p+2)$ は $\kappa_4 = 0$ を意味しない。β_{2p} は分散共分散行列により基準化された観測ベクトル $\boldsymbol{X}$ の、平均からの長さの二乗の分布の分散に関係しているが、κ_4 は他の側面にも関係している。Q_4 検定の意味はこの点で不明確である。

Cox & Small〔2〕は回帰の 2 次項の存在を検定する方式を提案している。座標に依存する方法は意味のとらえやすさとともに回帰式の任意性も含んでおり、また検定方式も複雑にな

る。座標に依存しない、不変な方式は数値計算ならびに分布論上の難点をもっている。他方 Q_3 検定はあらゆる2次項の存在に対する同時検定であるともみなされる。

Malkovich & Afifi〔5〕の多次元歪度と尖度はそれぞれ3次と4次のキュミュラントを用いて定義できる。これらに基づく検定も計算と分布論上の難点をもつ。

Q_3 と Q_4 に基づくキュミュラント検定では、各キュミュラントを推定する必要があるので、これらを個々に検討することができる。しかし変量数の増加に伴い、評価すべきキュミュラントの個数は急速に増大する。また検定は総括的となり検出力の低下をまねくとも考えられること、同時キュミュラントの実際的意味が不明確なことなども難点である。これらの問題とともに、検定統計量の近似分布の精度や他の検定法との差異、検出力などの検討も残されている。

参 考 文 献

〔1〕 Asano, Ch., Wakimoto, K., Shohoji, T., Komazawa, T., Jojima, K., Goto, M., Tanaka, Y., Tarumi, T. and Ohsumi, N. (1978) The Statistical Principle and Methodology in NISAN System -an introduction to NISAN system-, Res. Rep. No. 88, Res, Instit. Fund. Infor. Sc., Kyushu University.

〔2〕 Cox, D.R. and Small, N. J. H. (1978). Testing multivariate normality. Biometrika, Vol. 65, 263-272.

〔3〕 Dahiya, R. C. & Gurland, J. (1973). A test of fit for bivariate distributions. J. Roy. Statist. Soc. B, Vol. 35, 452-465.

〔4〕 Goto, M., Uesaka, H. and Asano, Ch. (1978). Methodology of Data

Investigation in NISAN System : Multivariate Normality Tests. Res。Rep. No.90, Res. Instit. Fund. Infor. Sc., Kyushu University.

〔5〕 Kendall, M. G. and Stuart, A. (1969). Advanced Theory of Statistics, Griffin, London.

〔6〕 Malkovich, J. F. & Afifi, A. A. (1973). On tests for multivariate normality. J. Amer. Statist. Assoc., Vol. 68, 176-179.

〔7〕 Mardia, K. V. (1970). Measures of mulivariate skewness and kurtosis with applications. Biometrika, Vol. 57, 519-530.

〔8〕 Mardia, K. V. (1974). Applications of some measures of multivariate skewness and Kurtosis in testing normality and robustness studies. Sankhya B, Vol. 36, 115-128.

〔9〕 Takeuchi, K. (1974). A test for multivariate normality. Behaviometrika, 1, 59-64.