

Elliptical 母集団下での次元に関する検定基準に対する上界の導出

北大・工・院 吉田 清隆
北大・工 今井 英幸
北大・工 佐藤 義治

1. はじめに

観測値が多母集団から得られるとき，母集団間にどの程度違いがあるか興味がある．次元はその差異を表し，判別関数の立場では線形判別関数の数に相当する．多変量分散分析の次元に関する検定は，通常，多変量正規母集団の仮定のもとで考えられており，代表的な検定方式として尤度比 (LR) 型，Lawley-Hotelling (LH) 型，Bartlett-Nanda-Pillai (BNP) 型があげられる．しかし，これらの帰無分布は局外母数に依存しており，小標本のときその影響が大きくなる．Schott[1] は LR 型，LH 型について，帰無分布に対する上限を導出した．本報告では Schott の結果を elliptical 母集団において拡張し，上記 3 タイプについて帰無分布に対する上界を与える．さらに contaminated 正規母集団の下で，LR 型，LH 型における上限を与える．

2 Elliptical 分布

p 変量確率ベクトル x の特性関数が次式であたえられるとき， x の分布を elliptical 分布とよび， $x \sim EC_p(\mu, \Omega, \psi)$ と書く．

$$\phi_x(t) = \exp(it'\mu)\psi(t'\Omega t)$$

ここに， $t: p \times 1$ は変数ベクトル， $\mu: p \times 1$ ， $\Omega: p \times p$ はパラメータ， ψ は実関数で分布によって定まる．確率密度関数が存在するならば

$$f(x) = C_p |\Omega|^{-\frac{1}{2}} g\{(x - \mu)'\Omega^{-1}(x - \mu)\}$$

という形をしている．ここに C_p は正規化定数， g は非負の実関数であり ψ に依存する． x の平均ベクトル，分散共分散行列はそれぞれ $E(x) = \mu$ ， $V(x) = -2\psi'(0)\Omega$ である．正規分布以外の elliptical 分布の例として次の 2 つをあげておく．

(i) contaminated 正規分布

$$\begin{aligned} f(x) &= (1 - \varepsilon)(2\pi)^{-\frac{p}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \mu)'\Omega^{-1}(x - \mu)\right) \\ &+ \varepsilon(2\pi\sigma^2)^{-\frac{p}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)'\Omega^{-1}(x - \mu)\right) \quad (0 \leq \varepsilon \leq 1) \end{aligned}$$

(ii) 多変量 t 分布

$$f(x) = \frac{\nu^{\frac{p}{2}} \Gamma(\frac{\nu+p}{2})}{\pi^{\frac{p}{2}} \Gamma(\frac{\nu}{2})} |\Omega|^{-\frac{1}{2}} \{\nu + (x - \mu)'\Omega^{-1}(x - \mu)\}^{-\frac{\nu+p}{2}} \quad (\nu: \text{自然数})$$

3 多変量分散分析における次元検定

第 i 母集団の j 番目の個体の観測値を表わす p 変量確率ベクトルを y_{ij} ($i = 1, \dots, k$; $j = 1, \dots, n_i$; $n_1 + \dots + n_k = n$) とし，平均ベクトルを μ_i ，共分散行列を Σ とする．ただし μ_i ， Σ は未知である．

np 次ベクトル $y = (y'_{11}, \dots, y'_{kn_k})'$ が elliptical 分布 $EC_{np}(\mu, I_n \otimes \Omega, \psi)$ にしたがっていると仮定する．ここに， $\mu = (\mu'_1, \dots, \mu'_1, \dots, \mu'_k, \dots, \mu'_k)'$ である． y を適当な直交行列により正準形

$$\tilde{y} \sim EC_{np}(\tilde{m}, I_n \otimes \Omega, \psi)$$

に変換する．ここに， $\tilde{y} = (\tilde{y}'_1, \dots, \tilde{y}'_n)'$ ， $\tilde{m} = (\sqrt{n}\tilde{\mu}', \tilde{m}'_2, \dots, \tilde{m}'_k, \mathbf{o}', \dots, \mathbf{o}')'$ であり，このとき，母集団間の変動行列 Δ ，群内標本変動行列 E ，群間標本変動行列 H はそれぞれ $\Delta = \sum_{i=2}^k \tilde{m}_i \tilde{m}'_i$ ， $E = \sum_{i=k+1}^n \tilde{y}_i \tilde{y}'_i$ ， $H = \sum_{i=2}^k \tilde{y}_i \tilde{y}'_i$ となる． $\Delta \Sigma^{-1}$ の固有値を $\delta_1 \geq \dots \geq \delta_p$ とすると「次元＝ s 」という仮説は

$$H_s : \delta_1 \geq \dots \geq \delta_s > \delta_{s+1} = \dots = \delta_p = 0$$

と表現できる．帰無仮説 H_s の本質的な部分は $\delta_{s+1} = \dots = \delta_p = 0$ であり， $\delta_1, \dots, \delta_s$ は局外母数である． H_s の検定を行うための基準は HE^{-1} の固有値 $l_1 > \dots > l_p (> 0)$ の内の $p - s$ 個の小さい方の固有値 $l_{s+1}, \dots, l_p$ の関数として与えられ，

$$\begin{aligned} \text{LR 型 } T_1 &= \log \prod_{i=s+1}^p (1 + l_i), \\ \text{LH 型 } T_2 &= \sum_{i=s+1}^p l_i, \\ \text{BNP 型 } T_3 &= \sum_{i=s+1}^p l_i / (1 + l_i) \end{aligned} \quad (1)$$

が代表的であるが，これらの帰無分布は局外母数に依存する．そこで，局外母数の影響を受けない検定方式を求める必要がある．

4. Elliptical 母集団での上界

正規母集団の下における，Schott[1]の手法を用いることにより，上記3タイプに関して，帰無分布に対する上界を導出する．

定理 1

$z = (z'_1, \dots, z'_{n-1-s})'$ に対し $z \sim EC_{(n-1-s)(p-s)}(\mathbf{o}, \mathbf{I}_{n-1-s} \otimes \mathbf{I}_{p-s}, \psi)$ と仮定する．

$W_H = \sum_{i=1}^{k-s-1} z_i z'_i$ ， $W_E = \sum_{i=k-s}^{n-1-s} z_i z'_i$ とし， $W_H W_E^{-1}$ の固有値を $w_1 > \dots > w_{p-s} (> 0)$ とする．

このとき

$$T_1^* = \log \prod_{i=1}^{p-s} (1 + w_i), \quad T_2^* = \sum_{i=1}^{p-s} w_i, \quad T_3^* = \sum_{i=1}^{p-s} w_i / (1 + w_i) \quad (2)$$

とすると，任意の $\delta_1, \dots, \delta_s$ に対して

$$\Pr\{T_i > c\} \leq \Pr\{T_i^* > c\} \quad (i = 1, 2, 3) \quad (3)$$

が成り立つ．

5. Contaminated 正規分布の下での上界

Elliptical 母集団の内，第2節であげた contaminated 正規母集団に限定して議論をすると，次の定理2より，(3)の左辺の確率が右辺の確率に達するための条件が与えられる．

定理 2

母集団分布として contaminated 正規分布を仮定する．すなわち， $y : np \times 1 = (y'_1, \dots, y'_n)'$ の確率密度関数を

$$\begin{aligned} f(y) &= (1 - \varepsilon)(2\pi)^{-\frac{np}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \mu)' \Omega^{-1} (y - \mu)\right) \\ &+ \varepsilon(2\pi\sigma^2)^{-\frac{np}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)' \Omega^{-1} (y - \mu)\right) \quad (0 \leq \varepsilon \leq 1) \end{aligned}$$

とする．また定理 1 の $z : (n-1-s)(p-s) \times 1 = (z'_1, \dots, z'_{n-1-s})'$ の確率密度関数を

$$g(z) = (1-\varepsilon)(2\pi)^{-(n-1-s)(p-s)/2} \exp\left(-\frac{1}{2}z'z\right) + \varepsilon(2\pi\sigma^2)^{-(n-1-s)(p-s)/2} \exp\left(-\frac{1}{2\sigma^2}z'z\right) \quad (0 \leq \varepsilon \leq 1)$$

とする．このとき， T_i, T_i^* ($i = 1, 2$) に対して

$$\lim_{\delta_1, \dots, \delta_s \rightarrow \infty} \Pr\{T_i > c\} = \Pr\{T_i^* > c\} \quad (i = 1, 2) \quad (4)$$

が成り立つ．ここに， T_i, T_i^* ($i = 1, 2$) は (1) (2) で与えられるものである．

定理 1, 2 より次の結果が得られる．

定理 3

母集団分布として contaminated 正規分布を仮定したとき，

$$\sup \Pr\{T_i > c\} = \Pr\{T_i^* > c\} \quad (i = 1, 2) \quad (5)$$

が成り立つ．

6. まとめ

Elliptical 母集団下において，Schott[1] の手法を用いることにより， T_i の帰無分布に対する上界 T_i^* ($i = 1, 2, 3$) を導出した． T_i^* ($i = 1, 2, 3$) の帰無分布は局外母数の影響を受けない．特に contaminated 正規母集団の下では，LR 型，LH 型の上界は上限であることが言えた．他の elliptical 分布に関する上限の導出に関しては，違う手法を検討する必要があるようである．

参考文献

1. Schott, J. R. (1984). Optimal bounds for the distributions of some test criteria for tests of dimensionality. *Biometrika*. **71**, 561-567.
2. Suzukawa, A., Sato, Y. (1994). Comparison of powers and exact Bahadur slopes of some tests for dimensionality in MANOVA models. *Communications in Statistics : Theory and Method*. **23**, 3581-3591.
3. 吉田清隆, 今井英幸, 宮腰政明, 伊達惇. (1994). 非正規母集団における次元に関する検定統計量の性質. 情報処理北海道シンポジウム '94 講演論文集, 84.
4. 吉田清隆, 今井英幸, 佐藤義治. (2001). Elliptical 分布における多変量分散分析の次元検定について．第 69 回日本統計学会講演報告集, 328-329.