高次元大標本枠組みにおける一般化分散の同等性検定

東京理科大・理・院 杉山 高聖　大阪府立大・工 兵頭 昌
大分県立看護科学大・看護 渡邉 弘己　明星大・教育 塚田　真一　東京理科大・理　瀬尾　隆

k を 2 以上の整数とし, $\mathbf{x}_{i1},\ldots,\mathbf{x}_{iN_i}$ $(i \in \{1,\ldots,k\})$ を, 平均ベクトル $\boldsymbol{\mu}_i$, 分散共分散行列 $\boldsymbol{\Sigma}_i$ である正規母集団 $\mathcal{N}_p(\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ からの大きさ N_i の無作為標本とする. このとき, 仮説検定問題

$$\mathcal{H}: |\boldsymbol{\Sigma}_1| = |\boldsymbol{\Sigma}_2| = \cdots = |\boldsymbol{\Sigma}_k| \quad \text{vs.} \quad \mathcal{A}: \neg\mathcal{H}$$

について考える. Najarzadeh [2] は, この検定に対する尤度比検定統計量

$$T_L = Np\left\{\ln\left(\sum_{i=1}^{k}\frac{N_i}{N}|\mathbf{S}_i|^{1/p}\right) - \sum_{i=1}^{k}\frac{N_i}{N}\ln|\mathbf{S}_i|^{1/p}\right\}$$

を導出した. ただし, $\mathbf{S}_i$ は不偏分散共分散行列である. $c_1,\ldots,c_k > 0$, $\sum_{i=1}^{k} c_i = 1$ を満たすノンランダムな k-次ベクトル $\mathbf{c} = (c_1,\ldots,c_k)^\top$ に対して, Jensen の不等式に関する基準

$$\Lambda(\mathbf{c}) = Np\left\{\ln\left(\sum_{i=1}^{k} c_i|\boldsymbol{\Sigma}_i|^{1/p}\right) - \sum_{i=1}^{k} c_i\ln|\boldsymbol{\Sigma}_i|^{1/p}\right\}$$

を定義する. $\Lambda(\mathbf{c}) \geq 0$ であり, $\Lambda(\mathbf{c}) = 0$ と $\mathcal{H}$ は同値であることから, $\Lambda(\mathbf{c})$ は仮説検定のための合理的な基準といえる. 特に, $\mathbf{c} = (N_1/N, N_2/N, \cdots, N_k/N)^\top =: \mathbf{c}_L$ とすれば, 尤度比検定統計量 T_L は $\Lambda(\mathbf{c}_L)$ の自然な推定量と解釈することができる. しかしながら, 高次元大標本枠組みの下で, 尤度比検定 T_L は, 基準 $\Lambda(\mathbf{c}_L)$ に対して漸近バイアスを持つ.

本研究では, 高次元大標本枠組みの下で, $\Lambda(\mathbf{c})$ の一致推定量

$$L_H(\mathbf{c}) = Np\left\{\ln\left(\sum_{i=1}^{k} c_i\widehat{|\boldsymbol{\Sigma}_i|}^{1/p}\right) - \sum_{i=1}^{k} c_i\ln\widehat{|\boldsymbol{\Sigma}_i|}^{1/p}\right\}$$

を構成した. ただし, $\widehat{|\boldsymbol{\Sigma}_i|} = |\mathbf{S}_i|\prod_{\ell=1}^{p}\{n_i/(n_i-\ell)\}$ である. ここで, $n_i = N_i - 1$ である. さらに, Cai et al. [1] で与えられた $|\mathbf{S}_i|$ に関する漸近理論を応用し, $L_H(\mathbf{c})$ が正規確率変数の 2 次形式に分布収束することを示した. とくに, 正規確率変数の 2 次形式が自由度 $k-1$ のカイ 2 乗変数の定数倍となるように $\mathbf{c}$ を選ぶと, $\mathbf{c} = (\sum_{j=1}^{k}\hat{\psi}_j^2)^{-1}(\hat{\psi}_1^2,\hat{\psi}_2^2,\ldots,\hat{\psi}_k^2) =: \mathbf{c}_H$. ただし, $\hat{\psi}_i = \{-\ln(1-p/n_i)\}^{-1/2}$ である. $L_H(\mathbf{c}_H)$ に適当な定数を乗じた

$$T_H = \hat{q}\sum_{j=1}^{k}\hat{\psi}_j^2 L_H(\mathbf{c}_H)$$

の高次元大標本枠組みの下での極限分布は, 自由度 $k-1$ のカイ 2 乗分布となる. ただし, $\hat{q} = p/n, n = \sum_{i=1}^{k} n_i$ である. 以上の結果に基づき, 有意水準 $\alpha(\in(0,1))$ の近似検定

$$T_H > \chi_{k-1}^2(\alpha) \iff \mathcal{H} \text{ を棄却}$$

を提案する. ただし, $\chi_{k-1}^2(\alpha)$ は 自由度 $k-1$ のカイ 2 乗分の上側 100α 点である. 尚, 詳細な漸近理論や提案検定のサイズ・パワーに関するシミュレーション結果については, 当日報告する.

参考文献

[1] Cai, T. T., Liang, T., Zhou, H. H. (2015). Law of Log Determinant of Sample Covariance Matrix and Optimal Estimation of Differential Entropy for High-Dimensional Gaussian Distributions, *J. Multivariate Anal.* **137**, 161–172.

[2] Najarzadeh, D. (2017). Testing equality of generalized variances of k multivariate normal populations, *Comm. Statist. Simulation Comput.* **46**, 6414–6423.