

一般欠測データの下での2標本問題における 多変量正規母集団の同等性検定

東京理科大学大学院	理学研究科	野村 玲実
東京理科大学	理学部	八木 文香
東京理科大学	理学部	瀬尾 隆

1 はじめに

一般欠測データの下で, 2つの多変量正規母集団の同等性検定問題について考える. すなわち, $\boldsymbol{\mu}^{(\ell)}$ と $\boldsymbol{\Sigma}^{(\ell)}$, $\ell = 1, 2$ をそれぞれ第 ℓ 母集団の平均ベクトルと分散共分散行列とすると, この検定問題は

$$H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)}$$

である平均ベクトルと分散共分散行列の同時検定を考えることと同じであり, 完全データの下では, 尤度比検定統計量とその修正尤度比検定統計量が与えられている (Muirhead (1982), Srivastava (2002) など参照). また 1 変量の場合については, Zhang et al. (2012) で議論され, 正確な尤度比検定が与えられている. 一方, 欠測データの下での同時検定については, 単調型欠測の場合の議論があり, Hao and Krishnamoorthy (2001) や Hosoya and Seo (2015, 2016) など尤度比検定統計量の導出とその帰無分布に対する近似上側パーセント点などが与えられている. ただし, Hao and Krishnamoorthy (2001) や Hosoya and Seo (2015) では 2-step 単調欠測データの下での 1 標本問題における同時検定を議論しており, Hosoya and Seo (2016) では 2-step 単調欠測データの下での多標本問題について議論している. このように単調欠測データの下での同時検定についての議論はあるが, 単調でない一般欠測データの下での同時検定の議論はない. そこで本報告では一般欠測データの下での 2 標本問題に対する同時検定問題, すなわち, 2つの多変量正規母集団の同等性検定問題を考え, その尤度比検定統計量を導出し, 尤度比検定を与える.

2 最尤推定量と反復法

第 ℓ 母集団 ($\ell = 1, 2$) に対して互いに独立で欠測値を含む p 次元観測ベクトルがそれぞれ $n^{(\ell)}$ 個あるとする. このとき $H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)}$ である仮説検定問題を考える. 尤度比検定統計量を与えるため, 最尤推定量の導出について考える. 同じ欠測パターンをもつ観測ベクトルを 1 つのグループとし, そのグループの総数を $K^{(\ell)}$ とする. また, 第 k 番目のグループにおける観測ベクトルの数を $n_k^{(\ell)}$ ($k = 1, 2, \dots, K^{(\ell)}$) とする. ここに $\sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} = n^{(\ell)}$, $\ell = 1, 2$ である. このとき, k 番目のグループに属する観測ベクトル $\mathbf{x}_{kj}^{(\ell)}$ ($k = 1, 2, \dots, K^{(\ell)}, j = 1, 2, \dots, n_k^{(\ell)}, \ell = 1, 2$) の欠測部分を除

くような変換行列 $\mathbf{B}_k^{(\ell)}$ を $\mathbf{x}_{kj}^{(\ell)}$ にかけて、データの変換を行う。すなわち

$$\mathbf{z}_{kj}^{(\ell)} = \mathbf{B}_k^{(\ell)} \mathbf{x}_{kj}^{(\ell)}$$

と変換する (Srivastava and Carter (1986) 参照). 例えば, $\mathbf{x}_{kj}^{(\ell)} = (x_{kj1}^{(\ell)}, x_{kj2}^{(\ell)}, *, x_{kj4}^{(\ell)}, *, x_{kj6}^{(\ell)}, x_{kj7}^{(\ell)})'$: 7×1 とすると

$$\mathbf{B}_k^{(\ell)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

であり, $\mathbf{z}_{kj}^{(\ell)} = (x_{kj1}^{(\ell)}, x_{kj2}^{(\ell)}, x_{kj4}^{(\ell)}, x_{kj6}^{(\ell)}, x_{kj7}^{(\ell)})' : 5 \times 1$ となる. このような, $\mathbf{z}_{kj}^{(\ell)} (= \mathbf{B}_k^{(\ell)} \mathbf{x}_{kj}^{(\ell)})$, $k = 1, 2, \dots, K^{(\ell)}$, $j = 1, 2, \dots, n_k^{(\ell)}$, $\ell = 1, 2$ に対して, 正規性を仮定し, 対数尤度関数を微分することによって, $\boldsymbol{\mu}^{(\ell)}$ の最尤推定量は

$$\hat{\boldsymbol{\mu}}^{(\ell)} = \left[\sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \hat{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)} \right]^{-1} \left[\sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \hat{\boldsymbol{\Lambda}}_k^{(\ell)-1} \bar{\mathbf{z}}_k^{(\ell)} \right], \ell = 1, 2$$

となり, $\boldsymbol{\Sigma}^{(\ell)}$ の最尤推定量 $\hat{\boldsymbol{\Sigma}}^{(\ell)}$ は

$$\sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \hat{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)} = \sum_{k=1}^{K^{(\ell)}} \mathbf{B}_k^{(\ell)'} \hat{\boldsymbol{\Lambda}}_k^{(\ell)-1} \hat{\mathbf{U}}_k^{(\ell)} \hat{\mathbf{U}}_k^{(\ell)'} \hat{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)}$$

を満たす解となる. ただし

$$\begin{aligned} \hat{\boldsymbol{\Lambda}}_k^{(\ell)} &= \mathbf{B}_k^{(\ell)} \hat{\boldsymbol{\Sigma}}^{(\ell)} \mathbf{B}_k^{(\ell)'}, \quad \hat{\mathbf{U}}_k^{(\ell)} = \mathbf{Z}_k^{(\ell)} - \mathbf{B}_k^{(\ell)} \hat{\boldsymbol{\mu}}^{(\ell)} \mathbf{e}_{n_k^{(\ell)}}', \quad \bar{\mathbf{z}}_k^{(\ell)} = \frac{1}{n_k^{(\ell)}} \sum_{\alpha=1}^{n_k^{(\ell)}} \mathbf{z}_{k\alpha}^{(\ell)}, \\ \mathbf{Z}_k^{(\ell)} &= (\mathbf{z}_{k1}^{(\ell)}, \mathbf{z}_{k2}^{(\ell)}, \dots, \mathbf{z}_{kn_k^{(\ell)}}^{(\ell)}), \quad \mathbf{e}_{n_k^{(\ell)}} = (1, 1, \dots, 1)'. \end{aligned}$$

同様に帰無仮説 ($H_0 : \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)}$) の下での $\boldsymbol{\mu} (= \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)})$ の最尤推定量は

$$\tilde{\boldsymbol{\mu}} = \left[\sum_{\ell=1}^2 \sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \tilde{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)} \right]^{-1} \left[\sum_{\ell=1}^2 \sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \tilde{\boldsymbol{\Lambda}}_k^{(\ell)-1} \bar{\mathbf{z}}_k^{(\ell)} \right]$$

となり, $\tilde{\boldsymbol{\Sigma}} (= \boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)})$ の最尤推定量) は

$$\sum_{\ell=1}^2 \sum_{k=1}^{K^{(\ell)}} n_k^{(\ell)} \mathbf{B}_k^{(\ell)'} \tilde{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)} = \sum_{\ell=1}^2 \sum_{k=1}^{K^{(\ell)}} \mathbf{B}_k^{(\ell)'} \tilde{\boldsymbol{\Lambda}}_k^{(\ell)-1} \tilde{\mathbf{U}}_k^{(\ell)} \tilde{\mathbf{U}}_k^{(\ell)'} \tilde{\boldsymbol{\Lambda}}_k^{(\ell)-1} \mathbf{B}_k^{(\ell)}$$

を満たす解となる. ただし

$$\tilde{\boldsymbol{\Lambda}}_k^{(\ell)} = \mathbf{B}_k^{(\ell)} \tilde{\boldsymbol{\Sigma}} \mathbf{B}_k^{(\ell)'}, \quad \tilde{\mathbf{U}}_k^{(\ell)} = \mathbf{Z}_k^{(\ell)} - \mathbf{B}_k^{(\ell)} \tilde{\boldsymbol{\mu}} \mathbf{e}_{n_k^{(\ell)}}'.$$

$\hat{\boldsymbol{\Sigma}}^{(\ell)}$ や $\tilde{\boldsymbol{\Sigma}}$ の正確な値は求めることが困難なため, Srivastava and Carter (1986) のアイデアによる反復法を用いて数値的に求める. $\hat{\boldsymbol{\Sigma}}^{(\ell)}$ を例として以下に手順を示す. まず, $\boldsymbol{\Sigma}_i^{(\ell)}, \boldsymbol{\mu}_i^{(\ell)}$ を第 i 近似値とし, $\boldsymbol{\Delta}_i^{(\ell)}$ を分散共分散行列の第 i 更新量とする. 初期値 $\boldsymbol{\Sigma}_0^{(\ell)}$ は完全データ部分から計算した分散共分散行列の最尤推定量とする. このとき, 反復式は

$$\boldsymbol{\Sigma}_{i+1}^{(\ell)} = \boldsymbol{\Sigma}_i^{(\ell)} + \boldsymbol{\Delta}_{i+1}^{(\ell)}$$

と書くことができる. ここに $\text{vec}(\mathbf{\Delta}_{i+1}^{(\ell)}) = (\mathbf{Q}_i^{(\ell)})^{-1} \text{vec}(\mathbf{E}_i^{(\ell)})$ で

$$\mathbf{Q}_i^{(\ell)} = \sum_{k=1}^{K^{(\ell)}} (n_k^{(\ell)} \mathbf{D}_{ik}^{(\ell)} \otimes \mathbf{D}_{ik}^{(\ell)} - \mathbf{D}_{ik}^{(\ell)} \otimes \mathbf{F}_{ik}^{(\ell)} - \mathbf{F}_{ik}^{(\ell)} \otimes \mathbf{D}_{ik}^{(\ell)}), \quad \mathbf{E}_i^{(\ell)} = \sum_{k=1}^{K^{(\ell)}} (n_k^{(\ell)} \mathbf{D}_{ik}^{(\ell)} - \mathbf{F}_{ik}^{(\ell)})$$

であり, $\mathbf{D}_{ik}^{(\ell)}, \mathbf{F}_{ik}^{(\ell)}$ は

$$\begin{aligned} \mathbf{D}_{ik}^{(\ell)} &= \mathbf{B}_k^{(\ell)'} (\mathbf{B}_k^{(\ell)} \mathbf{\Sigma}_i^{(\ell)} \mathbf{B}_k^{(\ell)'})^{-1} \mathbf{B}_k^{(\ell)}, \\ \mathbf{F}_{ik}^{(\ell)} &= \mathbf{B}_k^{(\ell)'} (\mathbf{B}_k^{(\ell)} \mathbf{\Sigma}_i^{(\ell)} \mathbf{B}_k^{(\ell)'})^{-1} \mathbf{U}_{ik}^{(\ell)} \mathbf{U}_{ik}^{(\ell)'} (\mathbf{B}_k^{(\ell)} \mathbf{\Sigma}_i^{(\ell)} \mathbf{B}_k^{(\ell)'})^{-1} \mathbf{B}_k^{(\ell)} \end{aligned}$$

である. ただし, $\mathbf{U}_{ik}^{(\ell)} = \mathbf{Z}_k^{(\ell)} - \mathbf{B}_k^{(\ell)} \boldsymbol{\mu}_i^{(\ell)} \mathbf{e}_{n_k^{(\ell)}}'$ である. よって尤度比は

$$\lambda = \frac{\prod_{k=1}^{K^{(1)}} |\widehat{\Lambda}_k^{(1)}|^{\frac{1}{2} n_k^{(1)}} \prod_{k=1}^{K^{(2)}} |\widehat{\Lambda}_k^{(2)}|^{\frac{1}{2} n_k^{(2)}}}{\prod_{k=1}^{K^{(1)}} |\widetilde{\Lambda}_k^{(1)}|^{\frac{1}{2} n_k^{(1)}} \prod_{k=1}^{K^{(2)}} |\widetilde{\Lambda}_k^{(2)}|^{\frac{1}{2} n_k^{(2)}}}$$

で与えられる. 尤度比検定統計量 $-2 \log \lambda$ の極限分布は自由度 $p(p+3)/2$ のカイ二乗分布に従う. 数値実験の結果については当日報告する.

参考文献

- [1] Hao, J. and Krishnamoorthy, K. (2001). Inferences on a normal covariance matrix and generalized variance with monotone missing data. *Journal of Multivariate Analysis*, **78**, 62-82.
- [2] Hosoya, M. and Seo, T. (2015). Simultaneous testing of the mean vector and the covariance matrix with two-step monotone missing data. *SUT Journal of Mathematics*, **51**, 83-98.
- [3] Hosoya, M. and Seo, T. (2016). On the likelihood ratio test for the equality of multivariate normal populations with two-step monotone missing data. *Journal of Statistical Theory and Practice*, **10**, 673-692.
- [4] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- [5] Shutoh, N., Kasumi, M., Morinaga, W., Yamada, S. and Seo, T. (2010). Testing equality of mean vectors in two sample problem with missing data. *Communications in Statistics-Simulation and Computation*, **39**, 487-500.
- [6] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*. Wiley, New York.
- [7] Srivastava, M. S. and Carter, E. M. (1986). The maximum likelihood method for non-response in sample surveys. *Survey Methodology*, **12**, 61-72.
- [8] Zhang, L., Xu, X. and Chen, G. (2012). The exact likelihood ration test for equality of two normal populations. *The American Statistician*, **66**, 180-184.