2-step 単調欠測データのもとでの
部分平均ベクトルの検定に対する修正尤度比検定統計量

東京理科大・理　川崎　玉恵
東京理科大・理　瀬尾　隆

1　はじめに

1 標本，2 標本問題における部分平均ベクトルの仮説検定問題を 2-step 単調欠測データのもとで考える．部分平均ベクトルの仮説検定問題について，欠測値を含んでいない完全データの場合は，尤度比を用いた検定統計量が 1 標本問題では Rao (1949) や Giri (1964) などで議論されており，これらは Rao の U 統計量と呼ばれている．また 2 標本問題になると，部分平均ベクトルの仮説検定問題は判別分析における追加情報と関連しており，Eaton and Kariya (1975) や Provost (1990) などで議論されている．多標本問題では，母集団分布が一般分布のもとでの Rao の U 統計量の帰無分布に対する漸近展開を Gupta et al. (2006) が与えており，Naito, Kawasaki and Seo (2018) では，部分平均ベクトルの仮説検定問題に対する T^2 型検定統計量を提案し，その近似分布と同時信頼区間を与えている．欠測値を含むデータに対しては，Kawasaki and Seo (2016) が 1 標本問題における 2-step 単調欠測データのもとでの最尤推定量と尤度比検定統計量を導出している．しかし，欠測値を含む場合のこの尤度比を用いると，完全データのときのような Rao の U 統計量が導出できないため，Kawasaki and Seo (2016) では尤度比検定統計量に対して，欠測率を用いた近似を提案している．

本報告では，2-step 単調欠測データのもとでの 1 標本，2 標本問題における部分平均ベクトルの仮説検定問題について，1 標本問題では Kawasaki and Seo (2016) で与えた尤度比検定統計量を，2 標本問題では最尤推定量を導出し，尤度比検定統計量を与えたのち，検定統計量の帰無分布を大標本漸近枠組みの下で漸近展開することで，それぞれの修正尤度比検定統計量を与える．さらにこれらの精度について，モンテカルロ・シミュレーションにより数値的評価を行う．

2　1 標本問題における部分平均ベクトルの仮説検定問題

2.1　完全データの場合

ここではまず，1 標本問題のもとでの完全データにおける部分平均ベクトルの仮説検定問題について紹介する．N 個の確率ベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ がそれぞれ独立に $N_p(\boldsymbol{\mu}, \Sigma)$ に従うとする．ただし，Σ は未知とする．さらに，$\boldsymbol{\mu}$ について次の分割を考える．

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1',\ \boldsymbol{\mu}_{(23)}')' \tag{1}$$

ただし

$$\boldsymbol{\mu}_1 = (\mu_1, \mu_2, \ldots, \mu_{p_1})',\ \boldsymbol{\mu}_{(23)} = (\mu_{p_1+1}, \mu_{p_1+2}, \ldots, \mu_p)'. \tag{2}$$

このとき以下の仮説検定問題を考える．

$$H_0 : \boldsymbol{\mu}_{(23)} = \boldsymbol{\mu}_{(23)0} \text{ given } \boldsymbol{\mu}_1 = \boldsymbol{\mu}_{10} \text{ vs. } H_1 : \boldsymbol{\mu}_{(23)} \neq \boldsymbol{\mu}_{(23)0} \text{ given } \boldsymbol{\mu}_1 = \boldsymbol{\mu}_{10} \tag{3}$$

ただし，$\boldsymbol{\mu}_{10}$ と $\boldsymbol{\mu}_{(23)0}$ は既知とする．

この仮説検定問題は尤度比 λ を用いて，以下のように与えられている．

$$U = \lambda^{-2/N} - 1 = \frac{T_p^2 - T_{p_1}^2}{N - 1 + T_{p_1}^2}$$

ただし

$$T_p^2 = N(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)' S^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0),\ T_{p_1}^2 = N(\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_{10})' S_{11}^{-1} (\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_{10})$$

であり

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i = \begin{pmatrix} \overline{\boldsymbol{x}}_1 \\ \overline{\boldsymbol{x}}_{(23)} \end{pmatrix},\ S = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' = \begin{pmatrix} S_{11} & S_{1(23)} \\ S_{(23)1} & S_{(23)(23)} \end{pmatrix}$$

である．また，この検定統計量 U について，$(N-p)U/(p-p_1)$ は H_0 の下で自由度 $p-p_1, N-p$ の F 分布に従うことが知られている．

2.2 2-step 単調欠測データの場合

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_1}$ が $N_p(\boldsymbol{\mu}, \Sigma)$ に，$\boldsymbol{x}_{N_1+1}, \boldsymbol{x}_{N_1+2}, \ldots, \boldsymbol{x}_N$ が $N_{p_1+p_2}(\boldsymbol{\mu}_{(12)}, \Sigma_{(12)(12)})$ に従う互いに独立な確率ベクトルとする．ただし，$\boldsymbol{\mu}$ と Σ はそれぞれ次のよう分割される．

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{10} \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_{(12)} \\ \boldsymbol{\mu}_3 \end{pmatrix},\quad \Sigma = \left(\begin{array}{cc|c} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \hline \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{array}\right) = \begin{pmatrix} \Sigma_{(12)(12)} & \Sigma_{(12)3} \\ \Sigma_{3(12)} & \Sigma_{33} \end{pmatrix}$$

また，$\boldsymbol{x}_j$ をそれぞれ $p_1 \times 1, p_2 \times 1, p_3 \times 1$ の確率ベクトルとし，$\boldsymbol{x}_j = (\boldsymbol{x}_{1j}', \boldsymbol{x}_{2j}', \boldsymbol{x}_{3j}')' = (\boldsymbol{x}_{(12)j}', \boldsymbol{x}_{3j}')'$ と表す．ただし，$\boldsymbol{x}_{ij}, i = 1, 2, 3, j = 1, 2, \ldots, N_1$ は $p_i \times 1$，$p = p_1 + p_2 + p_3$ であり，$N_2 = N - N_1$ である．このとき，2.1 節と同様に仮説検定問題 (3) を考える．

この問題における尤度比検定統計量 $-2\log\lambda$ は Kawasaki and Seo (2016) で与えられている．ただし

$$\lambda = \left(\frac{|\widetilde{\Phi}_{(12)(12)}|}{|\widehat{\Psi}_{11}| \cdot |\widehat{\Psi}_{22}|}\right)^{-N/2} \left(\frac{|\widetilde{\Phi}_{33}|}{|\widehat{\Psi}_{33}|}\right)^{-N_1/2}$$

であり，記号の説明については Kawasaki and Seo (2016) を参照されたい．この尤度比検定統計量における帰無分布の漸近展開を以下の大標本漸近枠組み

$$\delta_i = \frac{N_i}{N} \to \text{ positive constants as } N_1, N_2 \to \infty,\ i = 1, 2$$

の下で行い，修正尤度比検定統計量を導出していく．ただし摂動展開は以下を用いる．

$$\boldsymbol{z}_F = \begin{pmatrix} \boldsymbol{z}_{F1} \\ \boldsymbol{z}_{F2} \\ \boldsymbol{z}_{F3} \end{pmatrix} = \sqrt{N_1}\begin{pmatrix} \overline{x}_{F1} \\ \overline{x}_{F2} \\ \overline{x}_{F3} \end{pmatrix},\ \boldsymbol{z}_L = \begin{pmatrix} \boldsymbol{z}_{L1} \\ \boldsymbol{z}_{L2} \end{pmatrix} = \sqrt{N_2}\begin{pmatrix} \overline{x}_{L1} \\ \overline{x}_{L2} \end{pmatrix},$$

$$V_F = \begin{pmatrix} V_{F11} & V_{F12} & V_{F13} \\ V_{F21} & V_{F22} & V_{F23} \\ V_{F31} & V_{F32} & V_{F33} \end{pmatrix} = \sqrt{N_1 - 1}(S_F - I_p),$$

$$V_L = \begin{pmatrix} V_{F11} & V_{F12} \\ V_{F21} & V_{F22} \end{pmatrix} = \sqrt{N_2 - 1}(S_L - I_{p_{(12)}})$$

また

$$\overline{\boldsymbol{x}}_F = (\overline{\boldsymbol{x}}'_{F(12)}, \overline{\boldsymbol{x}}'_{F3})' = (\overline{\boldsymbol{x}}'_{F1}, \overline{\boldsymbol{x}}'_{F2}, \overline{\boldsymbol{x}}'_{F3})' = \left(\frac{1}{N_1}\sum_{j=1}^{N_1} \boldsymbol{x}'_{1j},\ \frac{1}{N_1}\sum_{j=1}^{N_1} \boldsymbol{x}'_{2j},\ \frac{1}{N_1}\sum_{j=1}^{N_1} \boldsymbol{x}'_{3j} \right)',$$

$$\overline{\boldsymbol{x}}_L = (\overline{\boldsymbol{x}}'_{L1}, \overline{\boldsymbol{x}}'_{L2})' = \left(\frac{1}{N_2}\sum_{j=N_1+1}^{N} \boldsymbol{x}'_{1j},\ \frac{1}{N_2}\sum_{j=N_1+1}^{N} \boldsymbol{x}'_{2j} \right)'$$

である．

以上から，尤度比検定統計量 $-2\log\lambda$ の特性関数は

$$\mathrm{E}[e^{it(-2\log\lambda)}] = (1-2it)^{-(p_2+p_3)/2}\left\{1 + \frac{(p_2+p_3)c}{2N}\{(1-2it)^{-1}-1\}\right\} + o(N^{-1})$$

と展開することができる．ただし

$$c = \frac{1}{p_2+p_3}\left\{\frac{1}{2}p_2(2p_1+p_2+2) + \frac{1}{\delta_1}p_3(p+3) - \frac{1}{\delta_1}p_3(p_3+2)\right\}.$$

よって，$-2\log\lambda$ の帰無分布に対する漸近展開は

$$\Pr(-2\log\lambda \le x) = G_{p_2+p_3}(x) + \frac{(p_2+p_3)c}{2N}\{G_{p_2+p_3+2}(x) - G_{p_2+p_3}(x)\} + o(N^{-1})$$

のように与えられる．ただし，$G_f(x)$ は自由度 f の χ^2 分布の分布関数とする．さらに，バートレット修正係数は $\rho = 1 - c/N$ で与えられる．

よって，修正尤度比検定統計量 $-2\rho\log\lambda$ を与えることができ，その分布関数は次のように表すことができる．

$$\Pr(-2\rho\log\lambda \le x) = G_{p_2+p_3}(x) + O(N^{-2})$$

3 2標本問題における部分平均ベクトルの仮説検定問題

2-step 単調欠測データのもとでの 2 標本問題における部分平均ベクトルの仮説検定問題

$$H_0 : \boldsymbol{\mu}^{(1)}_{(23)} = \boldsymbol{\mu}^{(2)}_{(23)} \text{ given } \boldsymbol{\mu}^{(1)}_1 = \boldsymbol{\mu}^{(2)}_1 \text{ vs. } H_1 : \boldsymbol{\mu}^{(1)}_{(23)} \ne \boldsymbol{\mu}^{(2)}_{(23)} \text{ given } \boldsymbol{\mu}^{(1)}_1 = \boldsymbol{\mu}^{(2)}_1 \qquad (4)$$

を考える．ただし，$\boldsymbol{\mu}^{(i)}_1, \boldsymbol{\mu}^{(i)}_{(23)}$ の分割は 1 標本問題のときの (1) 式，(2) 式と同様であり，$i = 1, 2$ は母集団を表す添え字とする．

仮説検定問題 (4) における尤度比検定統計量を導出し，1 標本問題と同様に大標本漸近枠組みのもとで漸近展開を行い，修正尤度比検定統計量の導出を行う．詳しい結果については，当日報告する．

4 数値実験

100 万回のモンテカルロ・シミュレーションにより本報告で与えた修正尤度比検定統計量についての数値的評価を行う．数値実験の結果は当日報告する．

参考文献

[1] Eaton, M. L. and Kariya, T. (1975). Tests on means with additional information. Technical Report #243, *University of Minnesota.*

[2] Giri, N. C. (1964). On the likelihood ratio test of a normal multivariate testing problem. *Ann. Math. Stat.*, **35**, 181–189.

[3] Gupta, A. K., Xu, J. and Fujikoshi, Y. (2006), An asymptotic expansion of the distribution of Rao's U-statistic under a general condition, *J. Multivariate Anal.*, **97**, 492-513.

[4] Kawasaki, T. and Seo, T. (2016). A test for subvector of mean vector with two-step monotone missing data. *SUT J. Math.*, **52**, 21–39.

[5] Naito, T., Kawasaki, T. and Seo, T. (2018). T^2 type test statistic and simultaneous confidence intervals for sub-mean vectors in k-sample problem. *Technical Report No.18-03, Statistical Research Group, Hiroshima University.*

[6] Provost, S. B. (1990). Estimators for the parameters of a multivariate normal random vector with incomplete data on two subvectors and test of independence. *Comput. Statist. Data Anal.*, **9**, 37–46.

[7] Rao, C. R. (1949). On some problems arising out of discrimination with multiple characters. *Sankhyā*, **9**, 343–364.