欠測値データのもとで標本数および分散共分散行列が異なる場合の平均ベクトル間の同等性検定

首藤　信通 (東京理科大学大学院 理学研究科)
久住麻希子 (ホトジェニック株式会社)
森永　　亘 (ホトジェニック株式会社)
山田　春一 (ホトジェニック株式会社)
瀬尾　　隆 (東京理科大学 理学部)

本報告においては欠測値を含むデータのもとで, 標本数および分散共分散行列が異なる場合の 2 標本問題に対する平均ベクトル間の同等性検定, および同時信頼区間の構成について, 反復法によって与えられる最尤推定量を用いた漸近的な議論を行った.

不完全データのもとで等しい分散共分散行列をもつ 1 標本問題, 2 標本問題に対する平均ベクトル間の同等性検定, 同時信頼区間の構成については, Srivastava (1985), Srivastava and Carter (1986) によって与えられている. また, 浜本ら (2007) によって 2 標本問題に対する平均ベクトル間の同等性検定における尤度比検定統計量の漸近的性質に対し, 数値的評価が与えられている. また, 首藤, 瀬尾 (2007) においては, Anderson (2003) のアイデアに対して, 欠測値を含む場合への適用を考えることにより, 異なる分散共分散行列をもち, 標本数, 欠測パターンが共に等しい 2 標本問題に対する平均ベクトル間の同等性検定が考えられている.

以上の先行研究を基に, 本報告では首藤, 瀬尾 (2007) の理論的結果に対して, 標本数が異なる場合への拡張を考え, モンテカルロ・シミュレーションを行うことにより, 与えられた平均ベクトル間の同等性検定における T^2 型統計量, 尤度比検定統計量 $-2\log\lambda$ に対して, 漸近的性質について数値的評価を与えた. さらに, k 標本問題への拡張についても議論した.

具体的には第 i 母集団 $(i=1,2)$ の観測ベクトル $\boldsymbol{x}_j^{(i)}$ $(j=1,2,\ldots,N^{(i)})$ は欠測値をもつ観測ベクトルを含み, 平均ベクトル $\boldsymbol{\mu}^{(i)}$, 分散共分散行列 $\boldsymbol{\Sigma}^{(i)}$ の p 変量正規分布に従うものとし, 2 標本問題に対する平均ベクトル間の同等性検定問題, すなわち

$$H_0\colon \boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)} \quad \text{vs.} \quad H_1\colon \boldsymbol{\mu}^{(1)}\neq\boldsymbol{\mu}^{(2)}$$

に対する尤度比検定を与えた. 尤度比検定統計量を導出するため, 欠測パターンによる観測ベクトルのグループ分けを行い, そのグループの総数を s, 第 ℓ 番目のグループにおける観測ベクトル数を $N_\ell^{(i)}$ とする. ただし, $\sum_{\ell=1}^{s} N_\ell^{(i)} = N^{(i)}$ である. このとき, 観測ベクトルから欠測値を取り除くため, 第 ℓ 番目のグループに属するそれぞれの観測ベクトルに対して, Srivastava (1985), Srivastava and Carter (1986) において用いられている変換行列 $\boldsymbol{B}_\ell$ をかける変換を行う. 例えば, 第 2 成分が欠測している観測ベクトル $\boldsymbol{x}_{\ell j}^{(i)} = (x_{\ell j,1}^{(i)}, *, x_{\ell j,3}^{(i)}, \ldots, x_{\ell j,p}^{(i)})'$ $(i=1,2,\ j=1,2,\ldots,N_\ell^{(i)})$ に対して, この変換を施した観測ベクトルは $\boldsymbol{z}_{\ell j}^{(i)} \equiv \boldsymbol{B}_\ell \boldsymbol{x}_{\ell j}^{(i)} = (x_{\ell j,1}^{(i)}, x_{\ell j,3}^{(i)}, \ldots, x_{\ell j,p}^{(i)})'$ となる. 従って, $\boldsymbol{z}_{\ell j}^{(i)} \sim \mathrm{N}_{p_\ell}(\boldsymbol{B}_\ell\boldsymbol{\mu}^{(i)},\ \boldsymbol{B}_\ell\boldsymbol{\Sigma}^{(i)}\boldsymbol{B}_\ell')$ である. ここに, p_ℓ は第 ℓ 番目のグループに属する観測ベクトル $\boldsymbol{x}_{\ell j}^{(i)}$ に含まれる欠測していない成分数である. 変換後の観測ベクトル $\boldsymbol{z}_{\ell j}^{(i)}$ を基に尤度関数を構成し, 尤度方程式を反復法 (Newton – Raphson Method) を用いて数値的に解くことにより, 最尤推定量を得る. こ

の結果を用いて T^2 型統計量, 尤度比検定統計量 $-2\log\lambda$ を構成した. 特に, 標本数・欠測パターンが等しく, 異なる分散共分散行列をもつ 2 標本問題に対する平均ベクトル間の同等性検定問題においては, Anderson (2003) の考えを用いることにより, 1 標本問題に帰着させせることができるので

$$H_0\colon \boldsymbol{\mu} = \mathbf{0} \quad \text{vs.} \quad H_1\colon \boldsymbol{\mu} \neq \mathbf{0}$$

に対する尤度比検定を考え, 検定統計量を構成した. ここに $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ である. また, 異なる分散共分散行列をもつ k 標本問題に対する平均ベクトル間の同等性検定問題

$$H_0\colon \boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(k)} \quad \text{vs.} \quad H_1\colon \boldsymbol{\mu}^{(i)} \neq \boldsymbol{\mu}^{(j)} \ \text{for some } i, j (i \neq j)$$

への拡張を行い, この問題に対する尤度比検定を与えた.

数値実験

欠測値を含む正規乱数を用いたモンテカルロ・シミュレーションを行った. 具体的には, 等しい分散共分散行列をもつと仮定したときの各母集団のデータセットを決めて, 等しい分散共分散行列をもつ 2 標本問題, 標本数・欠測パターンが等しく, 異なる分散共分散行列をもつ 2 標本問題, 異なる標本数・分散共分散行列をもつ 2 標本問題それぞれに対する平均ベクトル間の同等性検定について, 完全データ部分の観測ベクトル数を固定し, 欠測値を含む観測ベクトル数の割合を変動させシミュレーションを行った. このシミュレーションにより, 検定統計量の上側 100α%点と漸近的結果である χ^2 分布の上側 100α%点を比較した.

次に, 異なる標本数・分散共分散行列をもつ 2 標本問題に対する平均ベクトル間の同等性検定に対し, 分散共分散行列が異なるもとで同様のシミュレーションを行い, 検定統計量の漸近的性質について, 分散共分散行列が等しい場合の結果との比較を行った.

さらに, 異なる標本数・分散共分散行列をもつ平均ベクトル間の同等性検定について, 分散共分散行列が等しい仮定のもとで第 2 母集団からの観測ベクトル数を変動させ, シミュレーションを行い, 検定統計量の漸近的性質について考察した.

参考文献

[1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis.* (3rd ed.). John Wiley & Sons, New York.

[2] 浜本 功司, 久住 麻希子, 森永 亘, 山田 春一, 瀬尾 隆 (2007). 欠測値データがある場合の平均ベクトル間の同等性検定と同時信頼区間. 2007 年度統計関連学会連合大会 講演報告集, 259.

[3] 首藤 信通, 瀬尾 隆 (2007). 欠測値データのもとで異なる分散共分散行列をもつ 2 標本問題に対する平均ベクトルの同等性検定. 2007 年度統計関連学会連合大会 講演報告集, 27.

[4] Srivastava, M.S. (1985). Multivariate Data with Missing Observations. *Comm. Statist. Theory Methods*, **14**, 775–792.

[5] Srivastava, M.S. and Carter, E.M. (1986). The Maximum Likelihood Method for Non-Response in Sample Survey. *Survey Methodology*, **12**, 61–72.