データに欠測がある場合の分散共分散行列に関する尤度比検定について

千代岡 那王（東京理科大学大学院 理学研究科）
瀬尾　　隆（東京理科大学 理学部）

観測データがランダムに欠測している場合の分散共分散行列に関する仮説検定問題を考える．本報告では，特に，(I) Sphericity 検定と，(II) 一様構造モデルに対する検定問題の2つを考える．Sphericity とは，分散共分散行列が $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_p$ で表される構造で，一様構造とは，分散共分散行列が $\boldsymbol{\Sigma} = \sigma^2[(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p']$ で表される構造である．ここで，σ^2，ρ は未知パラメータで，$\boldsymbol{I}_p$ は $p \times p$ の単位行列，$\boldsymbol{1}_p = (1,\ldots,1)'$ はすべての成分が1の $p \times 1$ ベクトルである．

まず，データ行列を $\{x_{ij}\}$, $i = 1,\ldots,p$, $j = 1,\ldots,N$ とし，$\boldsymbol{x}_j = (x_{1j}, x_{2j}, \ldots, x_{pj})'$ は平均ベクトル $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$ の p 変量正規分布に従う互いに独立な観測ベクトルとし，それぞれランダムに欠測値を含むものとする．観測データが欠測している場合を考えているので，観測ベクトルを欠測パターンが同じものごとにグループ分割し，欠測を取り除くような変換を行う．すなわち，p 次元観測ベクトルにおいて欠測していない成分の数を $p_\ell(1 \le p_\ell \le p)$ とし，p 次元観測ベクトルを欠測のない p_ℓ 次元観測ベクトルにするために，Srivastava (1985) などで用いられている方法を用いて，欠測パターン ℓ ごとに変換行列 $\boldsymbol{B}_\ell$ を作成し，次のような変換を考える．

$$\boldsymbol{z}_{\ell j} = \boldsymbol{B}_\ell \boldsymbol{x}_{\ell j},\ \ j = 1,\ldots,N,\ \ \ell = 1,\ldots,L,$$

ここで，N は観測ベクトルの数，L は欠測パターンが同じ観測ベクトルをもつグループの総数，n_ℓ は ℓ 番目のグループの観測ベクトルの数 $(N = \sum_{\ell=1}^{L} n_\ell)$ とする．また，$\boldsymbol{B}_\ell$ は $p_\ell \times p$ の変換行列，$\boldsymbol{z}_{\ell j}$ は $p_\ell \times 1$ の変換された観測ベクトルである．

例えば，$*$ を欠測値として，$\boldsymbol{x}_j = (x_{1j}, *, x_{3j}, *, x_{5j}, \ldots, x_{pj})'$ とすると

$$\boldsymbol{B}_\ell = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

であり，$\boldsymbol{z}_{\ell j} = \boldsymbol{B}_\ell \boldsymbol{x}_j = (x_{1j}, x_{3j}, x_{5j}, \ldots, x_{pj})'$ となる．

このとき，$\boldsymbol{z}_{\ell j} \sim \mathrm{N}_{p\ell}(\boldsymbol{B}_\ell\boldsymbol{\mu}, \boldsymbol{B}_\ell\boldsymbol{\Sigma}\boldsymbol{B}_\ell')$ より，変換後の観測ベクトル $\boldsymbol{z}_{\ell j}$ の尤度関数は

$$\mathrm{L}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \mathrm{Const.} \prod_{\ell=1}^{L} |\boldsymbol{\Lambda}_\ell|^{-\frac{1}{2}N_\ell} \times \mathrm{etr}\left[-\frac{1}{2}\boldsymbol{\Lambda}_\ell^{-1} \sum_{\ell=1}^{L} \boldsymbol{V}_\ell\right]$$

となる. ただし, $\boldsymbol{\Lambda}_\ell = \boldsymbol{B}_\ell \boldsymbol{\Sigma} \boldsymbol{B}'_\ell$, $\boldsymbol{V}_\ell = \sum_{j=1}^{N_\ell}(\boldsymbol{z}_{\ell j} - \boldsymbol{B}_\ell \boldsymbol{\mu})(\boldsymbol{z}_{\ell j} - \boldsymbol{B}_\ell \boldsymbol{\mu})'$ である. 上記の尤度関数から対数尤度関数および尤度方程式を考える. そして, 尤度方程式より最尤推定量 $\widehat{\boldsymbol{\mu}}$, $\widetilde{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$, $\widetilde{\boldsymbol{\Sigma}}$ ($\widetilde{\boldsymbol{\mu}}$, $\widetilde{\boldsymbol{\Sigma}}$ は帰無仮説の下での最尤推定量) を求める.

$\boldsymbol{\Sigma}$ が既知の場合の $\boldsymbol{\mu}$ の最尤推定量 $\widehat{\boldsymbol{\mu}}$ は次のように与えられる.

$$\widehat{\boldsymbol{\mu}} = \left[\sum_{\ell=1}^{L} N_\ell \boldsymbol{B}'_\ell (\boldsymbol{B}_\ell \boldsymbol{\Sigma} \boldsymbol{B}'_\ell)^{-1} \boldsymbol{B}_\ell\right]^{-1} \left[\sum_{\ell=1}^{L}\sum_{j=1}^{N_\ell} \boldsymbol{B}'_\ell (\boldsymbol{B}_\ell \boldsymbol{\Sigma} \boldsymbol{B}'_\ell)^{-1} \boldsymbol{z}_{\ell j}\right]$$

また, $\boldsymbol{\Sigma}$ の最尤推定量 $\widehat{\boldsymbol{\Sigma}}$ は, 尤度方程式

$$\sum_{\ell=1}^{L} N_\ell \boldsymbol{B}'_\ell (\boldsymbol{B}_\ell \widehat{\boldsymbol{\Sigma}} \boldsymbol{B}'_\ell)^{-1} \boldsymbol{B}_\ell = \sum_{\ell=1}^{L}\sum_{j=1}^{N_\ell} N_\ell \boldsymbol{B}'_\ell (\boldsymbol{B}_\ell \widehat{\boldsymbol{\Sigma}} \boldsymbol{B}'_\ell)^{-1} (\boldsymbol{z}_{\ell j} - \boldsymbol{B}_\ell \widehat{\boldsymbol{\mu}})(\boldsymbol{z}_{\ell j} - \boldsymbol{B}_\ell \widehat{\boldsymbol{\mu}})' (\boldsymbol{B}_\ell \widehat{\boldsymbol{\Sigma}} \boldsymbol{B}'_\ell)^{-1} \boldsymbol{B}_\ell$$

から正確な値を求めることが困難なので, Srivastava and Carter (1986) による反復法を用いて数値的に導出する. また, $\widetilde{\boldsymbol{\Sigma}}$ については, 検定 (I) の場合は Sphericity 構造である分散共分散行列の性質を取り入れた反復法を用い, 検定 (II) の場合は, 一様構造の性質を取り入れた反復法を用いる.

以上より, 尤度比 λ が

$$\lambda = \prod_{\ell=1}^{L} \frac{|\boldsymbol{B}_\ell \widehat{\boldsymbol{\Sigma}} \boldsymbol{B}'_\ell|^{N_\ell/2}}{|\boldsymbol{B}_\ell \widetilde{\boldsymbol{\Sigma}} \boldsymbol{B}'_\ell|^{N_\ell/2}}$$

となる. 尤度比検定統計量 $-2\log\lambda$ は (I) の場合, 漸近的に自由度 $p(p+1)/2-1$ の χ^2 分布に従い, (II) の場合は漸近的に自由度 $p(p+1)/2-2$ の χ^2 分布に従うので, それぞれの検定に対する尤度比検定を与えることができる.

また, 完全データの場合には, Box (1949) を利用した修正尤度比検定統計量が Srivastava (2002) などによって与えられているが, データに欠測がある場合にも同じ修正尤度比検定統計量を用いて, モンテカルロ・シミュレーションによる数値的評価を行う.

参考文献

[1] Box, G. E. P. (1949). A General Distribution Theory for a Class of Likelihood Criteria, *Biometrika*, **36**, 317-346.

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

[3] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*, John Wiley & Sons, New York.

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