分散行列が異なる2群に対するDempster検定統計量の高次元漸近展開

広島大・理・院　姫野 哲人

1　はじめに

本研究では、MANOVA モデルにおいて分散行列が異なる 2 群のデータがある場合を考える。つまり、

$$Y_i = X\mathcal{B}_i + \mathcal{E}_i, \quad \mathcal{E}_i \sim N_{N\times p}(\boldsymbol{O}, I_N \otimes \Sigma_i), \quad \mathcal{E}_1 \amalg \mathcal{E}_2 \quad (i=1,2)$$

($Y_i : N \times p$ 観測行列, $X : N \times k$ 計画行列, $\mathcal{B}_i : k \times p$ 回帰係数行列) という状況を考える。そして、

$$H_0 : \mathcal{B}_1 = \mathcal{B}_2, \qquad H_1 : \mathcal{B}_1 \neq \mathcal{B}_2$$

という帰無仮説 H_0 と対立仮説 H_1 に対する検定に Dempster 検定統計量を用いて、その統計量の高次元漸近展開を用いる場合に、$\Sigma_1 \neq \Sigma_2$ という状況が統計量にどのような影響を及ぼすのかを考察していく。

2　Dempster 検定統計量

まず、上のモデルは

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X & \boldsymbol{O} \\ \boldsymbol{O} & X \end{pmatrix} \begin{pmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{pmatrix} + \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix}$$

と変形でき、帰無仮説 H_0 は

$$\begin{pmatrix} 1 & & 0 & -1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & -1 \end{pmatrix} \begin{pmatrix} \mathcal{B}_1 \\ \mathcal{B}_2 \end{pmatrix} = \boldsymbol{O}$$

とあらわすことができる。そして、

$$\begin{aligned} S_e &= (Y_1 - \hat{Y}_1)'(Y_1 - \hat{Y}_1) + (Y_2 - \hat{Y}_2)'(Y_2 - \hat{Y}_2) \\ S_h &= \frac{1}{2}(\hat{Y}_1 - \hat{Y}_2)'(\hat{Y}_1 - \hat{Y}_2) \\ \hat{Y}_i &= X(X'X)^{-1}X'Y_i \quad (i=1,2) \end{aligned}$$

となり、$\mathrm{tr}S_h/\mathrm{tr}S_e$ が Dempster 検定統計量と言われるものである。そして、

$$\mathrm{tr}S_e = \mathrm{tr}(\Sigma_1\mathcal{E}_{(1)}'\mathcal{E}_{(1)} + \Sigma_2\mathcal{E}_{(2)}'\mathcal{E}_{(2)}), \quad S_h = \frac{1}{2}\mathrm{tr}(\Sigma\mathcal{E}_{(3)}'\mathcal{E}_{(3)})$$

$$\mathcal{E}_{(i)} \overset{i.d.}{\sim} N_{n\times p}(\boldsymbol{O}, I_n \otimes I_p) \quad (n = N-k, i=1,2), \quad \mathcal{E}_{(3)} \overset{i.d.}{\sim} N_{k\times 2p}(\boldsymbol{O}, I_k \otimes I_{2p})$$

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_1^{1/2}\Sigma_2^{1/2} \\ \Sigma_2^{1/2}\Sigma_1^{1/2} & \Sigma_2 \end{pmatrix}$$

と変形することができ、$T = \sqrt{p}(2n\mathrm{tr}S_h/\mathrm{tr}S_e - k)$ の高次元漸近展開を求める。

3 高次元漸近展開

本研究では、高次元漸近展開を

$$n \to \infty, p \to \infty, \frac{n}{p} \to c \in (0, \infty)$$

という状況下で考える。また、

$$\frac{1}{p}\mathrm{tr}(\Sigma_1 + \Sigma_2)^i = O(1) \quad (i = 1, 2, \cdots)$$

という仮定を置き、

$$\frac{1}{p}\mathrm{tr}(\Sigma_1 + \Sigma_2)^i = c_i$$

とする。すると、Σ_1, Σ_2 が既知であれば、

$$P(\frac{T}{\sigma} \leq z) = \Phi(z) - \phi(z)\{\frac{1}{\sqrt{p}}\frac{a_3}{\sigma^3}h_2(z) + \frac{1}{p}(\frac{a_4}{\sigma^4}h_3(z) + \frac{a_6}{\sigma^6}h_5(z)) + \frac{1}{2n}\frac{a_2}{\sigma^2}z + o(n^{-1})\}$$

$$\sigma = \sqrt{\frac{2kc_2}{c_1^2}}, \quad a_3 = \frac{4kc_3}{3c_1^3}, \quad a_4 = \frac{2kc_4}{c_1^4}, \quad a_6 = \frac{8k^2c_3^2}{9c_1^6}, \quad a_2 = \frac{k^2}{c_1^2}\frac{2}{p}\mathrm{tr}(\Sigma_1^2 + \Sigma_2^2)$$

となる。ここで、$\Sigma_1 = \Sigma_2$ だった場合はこの漸近展開は

$$c_i = \frac{1}{p}\mathrm{tr}\Sigma_1^i (= \frac{1}{p}\mathrm{tr}\Sigma_2^i)$$

となり、また a_2 の部分が

$$a_2 = \frac{k^2c_2}{c_1^2}$$

と変わるだけである。ここで、$\Sigma_1 \neq \Sigma_2$ であるにも関わらず、$\Sigma_1 = \Sigma_2$ と判断し、その共通の分散を $\frac{1}{2}(\Sigma_1 + \Sigma_2)$ であると判断したとする。このとき、$\Sigma_1 \neq \Sigma_2$ と正しく判断した場合の漸近展開との差は

$$\phi(z)\frac{z}{2n}(\frac{2k^2\mathrm{tr}(\Sigma_1^2 + \Sigma_2^2)/p}{(\mathrm{tr}\Sigma_1 + \mathrm{tr}\Sigma_1)^2/p^2} - \frac{k^2\mathrm{tr}(\Sigma_1 + \Sigma_2)^2/4p}{(\mathrm{tr}\Sigma_1 + \mathrm{tr}\Sigma_1)^2/4p^2}) = \phi(z)\frac{z}{2n}\frac{k^2\mathrm{tr}(\Sigma_1 - \Sigma_2)^2/4p}{(\mathrm{tr}\Sigma_1 + \mathrm{tr}\Sigma_1)^2/4p^2}$$

であるので、

$$\frac{p\mathrm{tr}(\Sigma_1 - \Sigma_2)^2}{n(\mathrm{tr}\Sigma_1 + \mathrm{tr}\Sigma_1)^2} = \frac{\mathrm{tr}(\Sigma_1 - \Sigma_2)^2}{npc_2}$$

が大きくなればなるほど誤差が大きくなっていく。つまり、Σ_1 と Σ_2 の各成分の差の 2 乗和がそのまま誤差に影響する。このことは、Σ_1, Σ_2 が未知の場合の漸近展開にも同様の結果が得られることが予想できる。この Σ_1, Σ_2 が未知の場合の漸近展開とその数値シュミレーションについては当日報告する。