多変量多重比較法の開発と検討

東海大学・理学部　　道家暎幸

1. はじめに

多重比較法は、Tukey(1953), Scheffé(1953), Dunnett(1955) 等によって提案されている。Hochberg-Tamhane(1987) は Union-Intersection 法に基づく多変量多重比較法について論じている。多変量多重比較法では、最大固有値の分布関数を求める事になるが、この分布関数は母集団の数と変量の数に依存し、それらが大きくなると、その分布関数を求める事は容易でなく、漸化式の形で与えられる。

この研究の目的は、k 個の p 変量正規母集団における母平均ベクトル間の差に関する多変量多重比較法を構築することである。次に本研究での多変量多重比較法を、分散共分散行列が既知の場合と未知の場合の分けてその方法を議論する。k 個の p 変量正規母集団の分散共分散行列が既知の場合、任意に選ばれた 2 母集団の母平均ベクトルの差の仮説を検定するため、 χ^2 統計量を用いる。更に全てのペアについての同時検定を考える時、これらの χ^2 統計量の中の最大値の分布関数を導出する。分散共分散行列が未知の場合、統計量として Hotelling の T^2 統計量を用て、同様の手順で T^2 統計量の最大値の分布関数を求める。この方法に関しては、現在、理論的な検討を行っている。

2. 多変量多重比較法（分散共分散行列既知）

k 個の p 変量正規母集団があり、 i 番目の母集団の反応ベクトル $\boldsymbol{x}_i (i=1,2,\cdots,k)$ は p-変量正規分布 $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ に従い、$\boldsymbol{\Sigma}_i$ は既知とする。いま、仮説

$$H_{\{i,j\}}: \boldsymbol{\mu}_i = \boldsymbol{\mu}_j, \qquad K_{\{i,j\}}: \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j$$

を与える。全ての i,j の組み合わせに対する同時仮説 $\{H_{\{1,2\}}, H_{\{1,3\}}, \cdots, H_{\{k-1,k\}}\}$ を考える。検定統計量として、自由度 p のカイ 2 乗統計量

$$\chi^2_{i,j} = (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{j}})' \left(\frac{\boldsymbol{\Sigma}_i}{N_i} + \frac{\boldsymbol{\Sigma}_j}{N_j} \right)^{-1} (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{j}})$$

を用いる。同時仮説の下で、$\chi^2_{1,2}, \chi^2_{1,3}, \cdots, \chi^2_{k-1,k}$ の同時信頼領域を次に与える。

$$\bigcap_{i<j} D_{i,j} = \bigcap_{i<j} [\chi^2_{i,j} \le t] = [\max_{i<j} (\chi^2_{i,j}) \le t].$$

3. 多変量正規変数への変数変換

いま $\boldsymbol{y}_{i,j} = \bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j$ とおくと、$H_{\{i,j\}}$ の下で $\boldsymbol{y}_{i,j} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{i,j})$ 。ここで $\boldsymbol{\Sigma}_{i,j} = \dfrac{\boldsymbol{\Sigma}_i}{N_i} + \dfrac{\boldsymbol{\Sigma}_j}{N_j}$ である。いま $\boldsymbol{z}_{i,j}(p \times 1) = \boldsymbol{H}\boldsymbol{y}_{i,j}$ とし、ここで $\boldsymbol{H}(p \times p)$ は

$$\boldsymbol{H}\boldsymbol{\Sigma}_{i,j}\boldsymbol{H}' = \boldsymbol{I}(p \times p),$$

を満たす正則行列とする。$H_{\{i,j\}}$ の下で、 $\chi^2_{i,j}$ 統計量は

$$\chi^2_{i,j} = (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{j}})'\boldsymbol{\Sigma}^{-1}_{i,j}(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{j}}) = \boldsymbol{y}'_{i,j}\boldsymbol{\Sigma}^{-1}_{i,j}\boldsymbol{y}_{i,j} = \boldsymbol{z}'_{i,j}\boldsymbol{z}_{i,j}$$

であり、$\boldsymbol{z}_{i,j}$ をもとにした確率ベクトルは

$$\boldsymbol{z}(mp \times 1) = \left(\boldsymbol{z}'_{1,2}, \boldsymbol{z}'_{1,3}, \ldots, \boldsymbol{z}'_{k-1,k}\right)' \sim N_{mp}(\boldsymbol{0}, \boldsymbol{\Sigma})$$

であり、ここで ${}_kC_2 = m$ である。

4. $\max\limits_{i<j}(\chi^2_{i,j})$ **の分布**

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{\frac{mp}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\Bigl(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\Bigr)$$

ここで $\max\limits_{i<j}(\chi^2_{i,j}) = \chi^2_{k-1,k}$ と仮定すると、$\chi^2_{k-1,k}$ に対応する $\boldsymbol{z}_{k-1,k}$ の周辺密度関数を導くため、変数変換 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}$ を行う。ここで

$$\boldsymbol{A}(mp \times mp) = \begin{pmatrix} \boldsymbol{I}_{(m-1)p} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{-1}_{22}(\overline{(m-1)p} \times p) \\ \boldsymbol{0}(p \times \overline{(m-1)p}) & \boldsymbol{I}_p \end{pmatrix}$$

である。これから $\boldsymbol{y} \sim N_{mp}(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$ である。$\boldsymbol{y}$ の同時密度関数は

$$f(\boldsymbol{y}) \ = \frac{1}{(2\pi)^{\frac{(m-1)p}{2}}|\boldsymbol{\Sigma}_{11.2}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{y}'_1\boldsymbol{\Sigma}^{-1}_{11.2}\boldsymbol{y}_1} \times \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\boldsymbol{y}'_2\boldsymbol{y}_2}$$

であり、$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}^{-1}_{22}\boldsymbol{\Sigma}_{21}, \boldsymbol{y}'_2\boldsymbol{y}_2 = \chi^2_{k-1,k}$ である。ここで、$\boldsymbol{y}'_1\boldsymbol{\Sigma}^{-1}_{11.2}\boldsymbol{y}_1 = \chi^2_1$ とおくと、χ^2_1 と $\chi^2_{k-1,k}$ は、それぞれ自由度 $(m-1)p$ と p の χ^2 分布に従う。従って、最大 χ^2 統計量の分布関数は

$$\begin{aligned} F(x) &= P(0 \leq \chi^2_{k-1,k} \leq x) \\ &= c_{k,p}\int_0^x \left[\int_0^{\chi^2_{k-1,k}} \frac{1}{\Gamma\left(\frac{(m-1)p}{2}\right)2^{\frac{(m-1)p}{2}}}(\chi^2_1)^{\frac{(m-1)p}{2}-1}e^{-\frac{\chi^2_1}{2}}d\chi^2_1\right] \\ &\qquad \frac{1}{\Gamma\left(\frac{p}{2}\right)2^{\frac{p}{2}}}(\chi^2_{k-1,k})^{\frac{p}{2}-1}e^{-\frac{\chi^2_{k-1,k}}{2}}d\chi^2_{k-1,k} \end{aligned}$$

で与えられる。

参考文献

[1] Dunnett, C. W.(1955) : A multiple comparison procedure for comparing several treatments with a control., Journal of the American Statistical Association 50, 1096-1121.

[2] Hochberg, Y. and Tamhane, A. C.(1987) : Multiple Comparison Procedures, John Wiley & Sons.

[3] Scheffé, H.(1953) : A method for judging all contrasts in the analysis of variance., *Biometrika* 40, 87-104.

[4] Tukey, J. W.(1953) : The Problem of Multiple comparisons, Mimeographed monograph.