高次元非正規母集団でのMANOVA検定におけるDempster統計量の漸近分布の頑健性について

中央大学理工学研究科　博士課程後期課程
山田　隆行

ここでは，観測項目の数 p が標本サイズの数 N が共に大きい場合のMANOVA検定における Dempster 統計量の漸近分布について考察を行う．q 個の p 変量母集団 $\Pi_1,\dots,\Pi_q$ があり，Π_i の平均ベクトルを $\boldsymbol{\mu}_i$ とおく．第 i 母集団 Π_i から得られた $j(j=1,\dots,N_i)$ 個目の標本観測値ベクトル $\boldsymbol{x}_j^{(i)}$ に次のような線形構造を仮定する．

$$\boldsymbol{x}_j^{(i)}=\boldsymbol{\mu}_i+C\boldsymbol{\varepsilon}_j^{(i)}.$$

ここに，C は $p\times p$ の正則行列，$\boldsymbol{\varepsilon}_j^{(i)}=(\varepsilon_{j1}^{(i)},\dots,\varepsilon_{jp}^{(i)})'$ は $\boldsymbol{x}_j^{(i)}$ に対応する誤差変数ベクトルとする．誤差変数ベクトルは各成分が互いに独立に4次までの積率が

$$\mathrm{E}(\varepsilon_{jk}^{(i)})=0,\quad \mathrm{E}({\varepsilon_{jk}^{(i)}}^2)=1,\quad \mathrm{E}({\varepsilon_{jk}^{(i)}}^3)=0,\quad \mathrm{E}({\varepsilon_{jk}^{(i)}}^4)=\gamma,$$

である分布に従うものする．この構造を仮定したときに，各母集団の共分散行列 Σ は共通であり，CC' で与えられる．帰無仮説

$$H_0:\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_q$$

に対して p と N が共に大きい場合の検定法を考える．Fujikoshi et al. (2004) では各母集団に正規分布が仮定できる場合に統計量

$$\tilde{T}_D=\sqrt{p}\{n(\operatorname{tr}S_h/\operatorname{tr}S_e)-q\}$$

を用いた検定法を提案している．ここに，S_h と S_e はそれぞれ仮説と誤差に伴う平方和行列であり，

$$S_h=\sum_{i=1}^{q}N_i(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}})',\quad \bar{\boldsymbol{x}}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{x}_j^{(i)},\quad \bar{\boldsymbol{x}}=\frac{1}{N}\sum_{i=1}^{q}N_i\bar{\boldsymbol{x}}_i,$$

$$N=N_1+\cdots+N_q,\quad S_e=\sum_{i=1}^{q}S_i,\quad S_i=\sum_{j=1}^{N_i}(\boldsymbol{x}_j^{(i)}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_j^{(i)}-\bar{\boldsymbol{x}}_i)'$$

で与えられる．報告者は Fujikoshi et al. (2004) にある $\tilde{T}_D$ の漸近正規性がいくつかの仮定を加えれば観測値が非正規母集団より得られた場合でも成り立つことを示した．以下が概略である．今，

$$X=(\boldsymbol{x}_1^{(1)},\dots,\boldsymbol{x}_{N_1}^{(1)},\boldsymbol{x}_1^{(2)},\dots,\boldsymbol{x}_{N_q}^{(q)})'=A\Theta+\mathcal{E}C'$$

となる．ここに，

$$
\begin{aligned}
&A=(\boldsymbol{e}_1,\ldots,\boldsymbol{e}_1,\boldsymbol{e}_2,\ldots,\boldsymbol{e}_q)',\\
&\boldsymbol{e}_i=(\delta_{i1},\ldots,\delta_{iq})',\quad \delta_{ij}\text{はクロネッカーのデルタ},\\
&\Theta=(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_q)',\quad \mathcal{E}=(\boldsymbol{\varepsilon}_1^{(1)},\ldots,\boldsymbol{\varepsilon}_{N_1}^{(1)},\boldsymbol{\varepsilon}_1^{(2)},\ldots,\boldsymbol{\varepsilon}_{N_q}^{(q)})'
\end{aligned}
$$

である．このとき仮説 H_0 は $B\Theta=O$ と表せる．ただし，

$$
B=(I_{q-1},-\mathbf{1}_{q-1}),\quad \mathbf{1}_{q-1}=(1,\ldots,1)'
$$

とする．多変量一般化線形モデルの下で

$$
\operatorname{tr}S_h=\operatorname{tr}(\tilde{C}\mathcal{E}'\tilde{A}\mathcal{E}),\quad \operatorname{tr}S_e=\operatorname{tr}(\tilde{C}\mathcal{E}'\tilde{B}\mathcal{E})
$$

と表せる．ただし $\tilde{C}=C'C$ とし，

$$
\tilde{A}=A(A'A)^{-1}B'(B(A'A)^{-1}B')^{-1}B(A'A)^{-1}A',\quad \tilde{B}=I_N-A(A'A)^{-1}A'
$$

とする．行列 $\tilde{C}$ の (i,j) 成分を $\tilde{c}_{ij}$ とし，$\mathcal{E}=(\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n)'$ と表すと

$$
\frac{1}{\sqrt{p}}\begin{pmatrix}\frac{\operatorname{tr}S_h-(q-1)\operatorname{tr}\Sigma}{\sqrt{q-1}}\\ \frac{\operatorname{tr}S_e-n\operatorname{tr}\Sigma}{\sqrt{n}}\end{pmatrix}=\sum_{i=1}^{p}\tilde{c}_{ii}\begin{pmatrix}\frac{\boldsymbol{\varepsilon}_i'\tilde{A}\boldsymbol{\varepsilon}_i-(q-1)}{\sqrt{q-1}}\\ \frac{\boldsymbol{\varepsilon}_i'\tilde{B}\boldsymbol{\varepsilon}_i-n}{\sqrt{n}}\end{pmatrix}+2\sum_{i<j}^{p}\tilde{c}_{ij}\begin{pmatrix}\frac{\boldsymbol{\varepsilon}_i'\tilde{A}\boldsymbol{\varepsilon}_j}{\sqrt{q-1}}\\ \frac{\boldsymbol{\varepsilon}_i'\tilde{B}\boldsymbol{\varepsilon}_j}{\sqrt{n}}\end{pmatrix}\tag{1}
$$

と表すことができる．(1) の前半の和と後半の和が無相関であるので，それぞれの漸近正規性を示すことにより (1) の漸近正規性が示せる．前半の方は 2 変量ベクトルの中心極限定理より，後半の方はマルチンゲール差に対する中心極限定理を用いる．これらの漸近正規性は高次元枠組み

$$
A_1:\quad q;\text{fix},\quad p\to\infty,\quad N=O(p^{\delta}),\quad 0<\delta\le 1
$$

及び仮定

$$
\begin{aligned}
&A_2:\quad C=(c_{ij}),\quad c_{ij}\ge 0,\\
&A_3:\quad \tau_i=\operatorname{tr}\Sigma^i/p=O(1),\quad (i=1,\ldots,4),\\
&A_4:\quad N_i/N=O(1)\quad \text{as } p\to\infty.
\end{aligned}
$$

で成り立つ．まとめると，この 2 変量ベクトルは漸近独立であり，各成分は平均 0 の正規分布に漸近的に従う．(1) の漸近正規性を利用することにより，統計量 $\tilde{T}_D$ の漸近正規性が示せる．