確率統計基礎講義 A

若木宏文

wakaki@hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/index.shtml

2019.6.7

多変量分散分析

問題設定

モデル化: 要因 A について k 個の水準 $A_1, \ldots, A_k$ を考え，それぞれの水準における次の標本が得られるとする.

$$\begin{array}{lcl} \text{第 } 1 \text{ 水準} & : & \boldsymbol{Y}_{11}, \ldots, \boldsymbol{Y}_{1n_1} \\ \vdots & & \vdots \\ \text{第 } k \text{ 水準} & : & \boldsymbol{Y}_{k1}, \ldots, \boldsymbol{Y}_{kn_k} \end{array},$$

統計モデル:

$$\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{in_i} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_i, \Sigma) \quad (i = 1, \ldots, k)$$

(水準が異なれば独立)

仮説検定問題:

$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ (帰無仮説)

$H_1 : \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ のうちどれかは異なる (対立仮説)

検定統計量

郡内平方和積和行列：$W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_i)(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}}_i)'$,

群間平方和積和行列：$B = \sum_{i=1}^{k} n_i (\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{Y}})(\bar{\boldsymbol{Y}}_i - \bar{\boldsymbol{Y}})'$,

$$\bar{\boldsymbol{Y}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{Y}_{ij}, \quad \bar{\boldsymbol{Y}} = n^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{Y}_{ij} \quad \left(n = \sum_{i=1}^{k} n_i\right)$$

$$\Rightarrow \quad S_0^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_{ij} - \bar{\boldsymbol{Y}})' = B + W$$

検定統計量（続き）

定理：B と W は独立であり $W \sim W_p(n-k,\Sigma)$．H_0 が正しいときは $B \sim W_p(k-1,\Sigma)$ が成り立つ．従って H_0 の下で

$$\mathsf{E}\left[\frac{1}{n-k}W\right] = \mathsf{E}\left[\frac{1}{k-1}B\right] = \Sigma$$

$f_1 > f_2 > \cdots > f_{k-1}$ を $W^{-1}B$ の固有値とすると，次の３つの検定が代表的な検定として知られている．

Wilks の検定: $\displaystyle\prod_{j=1}^{k-1}\frac{1}{1+f_j} > c_1 \Rightarrow$ 棄却

Lawley-Hotelling の検定: $\displaystyle\sum_{j=1}^{k-1} f_j > c_2 \Rightarrow$ 棄却

Bartlette-Nanda-Pillai の検定: $\displaystyle\sum_{j=1}^{k-1}\frac{f_j}{1+f_j} > c_3 \Rightarrow$ 棄却

平方和の２次形式による表現

$$\mathbf{Y}_i = (\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{in_i})' \ (i = 1, \ldots, k)\ ,\ \mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_k')'$$ とし

$$D = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{1}_{n_k} \end{pmatrix}$$

とすると

$$(D'D)^{-1}D'\mathbf{Y} = (\bar{\boldsymbol{Y}}_1, \ldots, \bar{\boldsymbol{Y}}_k)'$$
$$D(D'D)^{-1}D'\mathbf{Y} = (\bar{\boldsymbol{Y}}_1\mathbf{1}_{n_1}', \ldots, \bar{\boldsymbol{Y}}_k\mathbf{1}_{n_k}')'$$

より

$$W = \mathbf{Y}'(I_n - D(D'D)^{-1}D')\mathbf{Y}$$

B と W の独立性と分布

$$Q = D(D'D)^{-1}D', P = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n',$$
$$\boldsymbol{\varepsilon}_{ij} = \mathbf{Y}_{ij} - \boldsymbol{\mu}_i \ (i = 1, \ldots, k;\ j = 1, \ldots, n_i)\ ,$$
$$\mathbf{Z} = (\boldsymbol{\varepsilon}_{11}, \ldots, \boldsymbol{\varepsilon}_{1n_1}, \ldots, \boldsymbol{\varepsilon}_{kn_k})',\ \mathbf{M} = (\boldsymbol{\mu}_1\mathbf{1}_{n_1}', \ldots, \boldsymbol{\mu}_k\mathbf{1}_{n_k}')'$$

とおくと $\boldsymbol{\varepsilon}_{11}, \ldots, \boldsymbol{\varepsilon}_{kn_k} \overset{i.i.d.}{\sim} N_p(\mathbf{0}, \Sigma)$ で,

$$B = (\mathbf{Z} + \mathbf{M})'(Q - P)(\mathbf{Z} + \mathbf{M})$$
$$W = (\mathbf{Z} + \mathbf{M})'(I_n - Q)(\mathbf{Z} + \mathbf{M}) = \mathbf{Z}'(I_n - Q)\mathbf{Z}$$

となり, $(I_n - Q)(Q - P) = O_{n\times n}$ となるから,
B と W は独立で, $W \sim W_p(n-k, \Sigma)$, $\mathbf{M} = \mathbf{O}$ ならば
$B \sim W_p(k-1, \Sigma)$