

# 共変量のある判別問題について

大塚大 基礎工 山口光代

共変量をもつ2つの正規母集団  $\pi_i: N_{p+q} \left[ \begin{bmatrix} \mu_i \\ \eta \end{bmatrix}, \Sigma \right]$   
 $i=1, 2$  からの  $N_i$  個の観測値  $\begin{bmatrix} x_{i1} \\ y_{i1} \end{bmatrix}, \dots, \begin{bmatrix} x_{iN_i} \\ y_{iN_i} \end{bmatrix}$  に基  
 いて、他の観測値  $\begin{bmatrix} x \\ y \end{bmatrix}$  が、 $\pi_1$  または  $\pi_2$  のいづれに属するかを  
 判別する問題を考える。ここに

$\mu_i (p \times 1), i=1, 2; \eta (q \times 1), \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} (p+q \times p+q)$   
 は未知パラメータで、 $\Delta^2 = (\mu_1 - \mu_2)' \Sigma_{11.2}^{-1} (\mu_1 - \mu_2) > 0$  と  
 する。  $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ 。 この判別問題に対して、  
 次の3つの判別統計量を扱う。

[1]. Cochran and Bliss (AMS 1948) に於て2与2に於て  
 に線形判別統計量

$$W^* = \left[ x^* - \frac{1}{2} (\bar{x}_1^* + \bar{x}_2^*) \right]' S_{11.2}^{-1} (\bar{x}_1^* - \bar{x}_2^*)$$

[2] Fujikoshi and Kanazawa に於て2与2に於て最も  
 判別統計量

$$\begin{aligned} Z^* = & \frac{N_1}{N_1+1+l_1/k_1} (x^* - \bar{x}_1^*)' S_{11 \cdot 2}^{-1} (x^* - \bar{x}_1^*) \\ & - \frac{N_2}{N_2+1+l_2/k_2} (x^* - \bar{x}_2^*)' S_{11 \cdot 2}^{-1} (x^* - \bar{x}_2^*) \end{aligned}$$

[3] 同様に  $Z^*$  において  $l_1 = l_2 = 0$  とし

$$\begin{aligned} R^* = & \frac{N_1}{N_1+1} (x^* - \bar{x}_1^*)' S_{11 \cdot 2}^{-1} (x^* - \bar{x}_1^*) \\ & - \frac{N_2}{N_2+1} (x^* - \bar{x}_2^*)' S_{11 \cdot 2}^{-1} (x^* - \bar{x}_2^*) \end{aligned}$$

ただし

$$x^* = x - S_{12} S_{22}^{-1} y, \quad \bar{x}_i^* = \bar{x}_i - S_{12} S_{22}^{-1} \bar{y}_i \quad i=1, 2,$$

$\bar{x}_i, \bar{y}_i$  : 標本平均  $S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$  は  $\Sigma$  の最良不偏推定量

$$S_{11 \cdot 2} = S_{11} - S_{12} S_{22}^{-1} S_{21}$$

$$l_i = (y - \bar{y}_i)' S_{22}^{-1} (y - \bar{y}_i), \quad k_i = n/N_i, \quad n = N_1 + N_2 - 2$$

である。

上の3つの統計量は、次の様な仮説検定問題と考えるときの検定統計量である。即ち。

$$\begin{aligned} H_1 : & \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x_{11} \\ y_{11} \end{bmatrix}, \dots, \begin{bmatrix} x_{1N_1} \\ y_{1N_1} \end{bmatrix} \text{ は } \pi_1 \text{ に属し} \\ & \begin{bmatrix} x_{21} \\ y_{21} \end{bmatrix}, \dots, \begin{bmatrix} x_{2N_2} \\ y_{2N_2} \end{bmatrix} \text{ は } \pi_2 \text{ に属する。} \end{aligned}$$

$$\begin{aligned} H_2 : & \begin{bmatrix} x_{11} \\ y_{11} \end{bmatrix}, \dots, \begin{bmatrix} x_{1N_1} \\ y_{1N_1} \end{bmatrix} \text{ は } \pi_1 \text{ に属し} \\ & \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x_{21} \\ y_{21} \end{bmatrix}, \dots, \begin{bmatrix} x_{2N_2} \\ y_{2N_2} \end{bmatrix} \text{ は } \pi_2 \text{ に属する。} \end{aligned}$$

$H_i$  ( $i=1, 2$ ) のFでの尤度関数を  $L_i$  とするとき。

- [1]  $\mu_1, \mu_2, \eta, \Sigma$  が既知として、 $L_1, L_2$  に対して  $\mu_1, \mu_2, \eta, \Sigma$  の最大推定量を代入したものが  $W^*$  である。
- [2]  $\max_{\mu_1, \mu_2, \eta, \Sigma} L_1 / \max_{\mu_1, \mu_2, \eta, \Sigma} L_2$  が  $Z^*$  である。
- [3]  $\Sigma$  が既知として、 $\max_{\mu_1, \mu_2, \eta} L_1, \max_{\mu_1, \mu_2, \eta} L_2$  を求め、それによって  $\Sigma$  の最大推定量を代入したものが  $R^*$  になっている。

判別方程式は

$$W^* > c \quad \text{又は} \quad W^* \leq c$$

$$Z^* < k \quad \text{又は} \quad Z^* \geq k$$

$$R^* < e \quad \text{又は} \quad R^* \geq e$$

に従って、 $\begin{bmatrix} x \\ y \end{bmatrix}$  は  $\pi_1$  又は  $\pi_2$  に属すると判別される。分離点  $c, k, e$  は、0 又は正数、又は  $\bar{x}, \bar{y}, \bar{s}$  の関数である。又、最大判別方程式は

$$Z^* < 0 \quad \text{又は} \quad Z^* \geq 0$$

に従って、 $\begin{bmatrix} x \\ y \end{bmatrix}$  は  $\pi_1$  又は  $\pi_2$  に属すると判別する式である。

Covariate 変数がない場合、即ち  $q=0$  の場合は、 $R^*$  と  $Z^*$  は同一の統計量になり、 $W^*$  は Anderson の  $W$  に、 $R^*, Z^*$  は、John-Kudo の  $Z$  に対応している。

$k_1, k_2$  が、それぞれ 定数に収束するよう  $N_1, N_2$  が  
大きくなる時の、 $W^*, Z^*, R^*$  の極限分布は、

$$[\tilde{y}] \text{ が } \pi_1 \text{ に属しているなら, } W^* \leadsto N[\tfrac{1}{2}\Delta^2, \Delta^2]$$

$$Z^*, R^* \leadsto N[-\Delta^2, 4\Delta^2]$$

$$[\tilde{y}] \text{ が } \pi_2 \text{ に属しているなら, } W^* \leadsto N[-\tfrac{1}{2}\Delta^2, \Delta^2]$$

$$Z^*, R^* \leadsto N[\Delta^2, 4\Delta^2]$$

に従う。

本稿では、 $W^*, R^*, Z^*$  の分布の漸近展開を求め、それらの漸近展開式から得られる、ある種の誤判別の確率の大小を比較することによって、3つの統計量の“良さ”といったものを比較する。

### §1. 判別統計量の分布の漸近展開式。

#### 0]. $W^*, R^*, Z^*$ の分布の漸近展開

$$\begin{aligned} &Pr\{(W^* - \tfrac{1}{2}\Delta^2)\Delta^{-1} \leq u \mid [\tilde{y}] \in \pi_1\} \\ &= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ \frac{p}{2\Delta} H_1(u) + \frac{p}{2\Delta^2} H_2(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \right. \\ &\quad + \frac{1}{N_2} \left\{ -\frac{p}{2\Delta} H_1(u) + \left( \frac{p}{2\Delta^2} + \frac{1}{2} \right) H_2(u) - \frac{1}{\Delta} H_3(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \\ &\quad \left. + \frac{1}{n} \left\{ -\frac{1}{2}(p+q+1)\Delta H_1(u) + \left( \frac{1}{4}\Delta^2 + \frac{3}{2}(p+q+1) \right) H_2(u) - \Delta H_3(u) + H_4(u) \right\} \right] \\ &+ O_2 \quad : \text{Memon and Okamoto (AMS 1970)} \end{aligned}$$

$$\Pr\{(R^* + \Delta^2)(2\Delta)^{-1} \leq u \mid [\frac{x}{y}] \in \pi_1\}$$

$$= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ \frac{p}{2\Delta^2} H_2(u) + \frac{1}{2\Delta} H_3(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \right. \\ \left. + \frac{1}{N_2} \left\{ -\frac{\Delta}{2} H_1(u) + \left( \frac{p}{2\Delta^2} - \frac{1}{2} \right) H_2(u) + \frac{1}{2\Delta} H_3(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \right. \\ \left. + \frac{1}{m} \left\{ \frac{1}{2}(p+g+1)\Delta H_1(u) + \left( \frac{3}{2}(p+g+1) + \frac{\Delta^2}{4} \right) H_2(u) + \Delta H_3(u) \right. \right. \\ \left. \left. + H_4(u) \right\} \right] + O_2 \quad : \text{Fujikoshi and Kanazawa}$$

$$\Pr\{(Z^* + \Delta^2)(2\Delta)^{-1} \leq u \mid [\frac{x}{y}] \in \pi_1\}$$

$$= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ \frac{p}{2\Delta^2} H_2(u) + \frac{1}{2\Delta} H_3(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \right. \\ \left. + \frac{1}{N_2} \left\{ -\frac{\Delta}{2} H_1(u) + \left( \frac{p}{2\Delta^2} - \frac{1}{2} \right) H_2(u) + \frac{1}{2\Delta} H_3(u) + \frac{1}{2\Delta^2} H_4(u) \right\} \right. \\ \left. + \frac{1}{m} \left\{ \frac{1}{2}(p+1)\Delta H_1(u) + \left( \frac{1}{2}(3p+g+3) + \frac{\Delta^2}{4} \right) H_2(u) + \Delta H_3(u) \right. \right. \\ \left. \left. + H_4(u) \right\} \right] + O_2 \quad : \text{Fujikoshi and Kanazawa}$$

$[\frac{x}{y}] \in \pi_2$  の場合も、上の各式の  $N_1$  と  $N_2$  に  $\lambda$  を代入して、得られる。又、

$$H_1(u) = 1, \quad H_2(u) = -u, \quad H_3(u) = u^2 - 1, \quad H_4(u) = -u^3 + 3u$$

である。

[2] Studentized  $W^*, R^*, Z^*$  の分布の漸近展開。

$$D^2 = (\bar{x}_1^* - \bar{x}_2^*)' S_{11,2}^{-1} (\bar{x}_1^* - \bar{x}_2^*) \text{ は } O_p(1) \text{ と仮定可い。}$$

$$Pr\{(W^* - \frac{1}{2}D^2)D^{-1} \leq u \mid [\frac{x}{y}] \in \pi_1\}$$

$$= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ \frac{1}{2\Delta} (p-1) H_1(u) + \frac{1}{2} H_2(u) \right\} \right. \\ \left. + \frac{1}{n} \left\{ (p+q) H_2(u) + \frac{1}{4} H_4(u) \right\} \right] + O_2$$

: Kanazawa and Fujikoshi

$$Pr\{(R^* + D^2)(2D)^{-1} \leq u \mid [\frac{x}{y}] \in \pi_1\}$$

$$= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ -\frac{1}{2\Delta} (p-2) H_1(u) + \frac{1}{2} H_2(u) + \frac{1}{2\Delta} H_3(u) \right\} \right. \\ \left. + \frac{1}{N_2} \left\{ -\left(\frac{p}{2\Delta} + \frac{\Delta}{2}\right) H_1(u) - H_2(u) - \frac{1}{2\Delta} H_3(u) \right\} \right. \\ \left. + \frac{1}{n} \left\{ (p+q) H_2(u) + \frac{1}{4} H_4(u) \right\} \right] + O_2$$

: Fujikoshi and Kanazawa

$$Pr\{(Z^* + D^2)(2D)^{-1} \leq u \mid [\frac{x}{y}] \in \pi_1\}$$

$$= \Phi(u) + \phi(u) \left[ \frac{1}{N_1} \left\{ -\frac{1}{2\Delta} (p-2) H_1(u) + \frac{1}{2} H_2(u) + \frac{1}{2\Delta} H_3(u) \right\} \right. \\ \left. + \frac{1}{N_2} \left\{ -\left(\frac{p}{2\Delta} + \frac{\Delta}{2}\right) H_1(u) - H_2(u) - \frac{1}{2\Delta} H_3(u) \right\} \right. \\ \left. + \frac{1}{n} \left\{ -\frac{3\Delta}{2} H_1(u) + p H_2(u) + \frac{1}{4} H_4(u) \right\} \right] + O_2$$

: Fujikoshi and Kanazawa

$[\frac{x}{y}] \in \pi_2$  の場合 V. 8. 上の各式の  $N_1$  と  $N_2$  に  $\lambda$  を換之れず  
す。

52. 第一種の誤判別の確率を一定にしたいときの  
第二種の誤判別の確率

ここでは、Studentized  $W^*$ ,  $R^*$ ,  $Z^*$  の各々の漸近展開を用いて、 $[\tilde{y}]$  が  $\pi_1$  に属しているのに、 $\pi_2$  に属していると誤って判別する確率が与えられた値:  $\alpha$  になるように、それぞれの分離点を決めるとき、 $[\tilde{y}]$  が  $\pi_2$  に属しているとき、誤って  $\pi_1$  に属していると判別する確率を求める。

$$\Pr\{W^* \leq C_\alpha \mid [\tilde{y}] \in \pi_1\} = \alpha, \quad \Pr\{R^* \geq e_\alpha \mid [\tilde{y}] \in \pi_1\} = \alpha,$$

$$\Pr\{Z^* \geq k_\alpha \mid [\tilde{y}] \in \pi_1\} = \alpha$$

となる分離点、 $C_\alpha$ ,  $e_\alpha$ ,  $k_\alpha$  は、次の漸近展開式によって与えられる。

$$C_\alpha = \xi_W D + \frac{1}{2} D^2, \quad e_\alpha = 2\xi_R D - D^2, \quad k_\alpha = 2\xi_Z D - D^2$$

ただし

$$\xi_W = u_0 - \frac{1}{2N_1\Delta} \left\{ 2(p-1)\Delta H_2(u_0) \right\} - \frac{1}{4n} \left\{ 4(p+g)H_2(u_0) + H_4(u_0) \right\} + O_2$$

$$\begin{aligned} \xi_R &= -u_0 + \frac{1}{2N_1\Delta} \left\{ (p-2)H_1(u_0) + \Delta H_2(u_0) - H_3(u_0) \right\} \\ &\quad + \frac{1}{2N_2\Delta} \left\{ (p+\Delta^2)H_1(u_0) - 2\Delta H_2(u_0) + H_3(u_0) \right\} \\ &\quad + \frac{1}{4n} \left\{ 4(p+g)H_2(u_0) + H_4(u_0) \right\} + O_2 \end{aligned}$$

$$\begin{aligned}\xi_2 = & -u_0 + \frac{1}{2N_1\Delta} \{ (p-2)H_1(u_0) + \Delta H_2(u_0) - H_3(u_0) \} \\ & + \frac{1}{2N_2\Delta} \{ (p+\Delta^2)H_1(u_0) - 2\Delta H_2(u_0) + H_3(u_0) \} \\ & + \frac{1}{4n} \{ 2g\Delta H_1(u_0) + 4pH_2(u_0) + H_4(u_0) \} + O_2\end{aligned}$$

ただし,  $u_0$  は  $N[0,1]$  の  $100\alpha\%$  点である. これらの分  
離点を用いるとき.

$$\begin{aligned}Pr\{W^* > c_\alpha \mid [\frac{x}{y}] \in \pi_2\} \\ = 1 - \Phi(u_0 + \Delta) - \phi(u_0 + \Delta) \left[ \frac{1}{2N_1\Delta} \{ \Delta^2 - (p-1) \} - \frac{1}{2N_2\Delta} \{ \right. \\ \left. 4(u_0 + \Delta)\Delta - 3(p-1) \} - \frac{\Delta}{4n} \{ u_0^2 + 2(p+g-1) \} \right] + O_2\end{aligned}$$

$$\begin{aligned}Pr\{R^* < c_\alpha \mid [\frac{x}{y}] \in \pi_2\} \\ = 1 - \Phi(u_0 + \Delta) - \phi(u_0 + \Delta) \left[ \frac{1}{2N_1\Delta} \{ 2u_0^2 + 4\Delta u_0 + 3\Delta^2 + (p-1) \} \right. \\ \left. - \frac{1}{2N_2\Delta} \{ 2(u_0 + \Delta)(u_0 + 3\Delta) - (p-1) \} - \frac{\Delta}{4n} \{ u_0^2 + 2(p+g-1) \} \right] \\ + O_2\end{aligned}$$

$$\begin{aligned}Pr\{Z^* < k_\alpha \mid [\frac{x}{y}] \in \pi_2\} \\ = 1 - \Phi(u_0 + \Delta) - \phi(u_0 + \Delta) \left[ \frac{1}{2N_1\Delta} \{ 2u_0^2 + 4\Delta u_0 + 3\Delta^2 - (p-1) \} \right. \\ \left. - \frac{1}{2N_2\Delta} \{ 2(u_0 + \Delta)(u_0 + 3\Delta) - (p-1) \} - \frac{\Delta}{4n} \{ u_0^2 + 2(p-g-1) \} \right] \\ + O_2\end{aligned}$$



これより、すべし、言え、このとき、共変量の次元が小さく  
 なる。  $R^*$  と  $Z^*$  を比べると、  $Z^*$  の方が、第二種の誤確率が小さ  
 くなるということである。又、この三つの判別誤確率の中、  
 一番小さい確率を与えるのは、次の表の統計量である。三つ  
 の統計量の大きさの表にすると、もっと細かく場合分けしな  
 ければならぬので、次表にとどめる。

| $\Delta \backslash R_2$ | $R_2 \leq R_1$ | $R_1 < R_2 < 2R_1 + g$ | $2R_1 + g = R_2$ | $2R_1 + g < R_2$ |                |
|-------------------------|----------------|------------------------|------------------|------------------|----------------|
| 0                       | $Z^*$          | $W^*$                  | $W^*$            | $u_0 < u^*$      | $u^* \leq u_0$ |
|                         |                |                        |                  | $W^*$            | $W^*$          |
|                         |                | $Z^*$                  | $Z^*$            | $\Delta_2$       |                |
|                         |                |                        |                  | $\Delta_1$       |                |
| 大                       |                |                        |                  | $W^*$            |                |
|                         |                |                        |                  |                  |                |

$$R_1 = R_2 \quad u^* = -\{(R_2 - 2R_1 - g)(p-1)/(R_1 + g)\}^{1/2}, \quad \Delta_0 = -(u_0^2 + p-1)/2u_0,$$

$$\Delta_1 = \{-(R_2 - R_1)u_0 + R_1\}/(R_2 - 2R_1 - g),$$

$$\Delta_2 = \{-(R_2 - R_1)u_0 - R_1\}/(R_2 - 2R_1 - g),$$

$$R^2 = (R_2 - R_1)\{(R_1 + g)u_0^2 - (R_2 - 2R_1 - g)(p-1)\} \quad \text{である。}$$

判別問題において、  $\Delta$  が小さいときの方が判別しにくい  
 わけであるが、  $\Delta$  が小さいとき、片方の誤確率と一定になる  
 という条件の下では、線形判別統計量の方が、より好ましい  
 という結論になる。なお、この表は、  $W^*$ ,  $R^*$ ,  $Z^*$  の漸近展

開式 §1. [1] かう計算しても、同じ結果が得られる。又、共変量を無視して、この判別問題を、普通の  $p$ -次元判別問題として、取り扱ふと、

$$\Delta_0^2 = (\mu_1 - \mu_2)' \Sigma_{11}^{-1} (\mu_1 - \mu_2) < \Delta^2$$

であるから、第 2 種の誤判別の確率は、一般に大きくなる。

### 参考文献

- [1] Cochran, W. G. and Bliss, C. I. (1948). Discriminant functions with covariance. *Ann. Math. Stat.* 19 151-176.
- [2] Cochran, W. G. (1964). Comparison of two methods of handling covariates in discriminatory analysis. *Ann. Inst. Statist. Math.* 16 43-53.
- [3] Memon, A. Z. and Okamoto, M. (1970). The classification statistic  $W^*$  in covariate discriminant analysis. *Ann. Math. Stat.* 41 1491-1499.
- [4] Kanazawa, M. and Fujikoshi, Y. (1975). An asymptotic expansion of the distribution of the studentized classification  $W^*$  and its applications.
- [5] Fujikoshi, Y. and Kanazawa, M. (1974). The maximum likelihood classification statistic in covariate discriminant analysis and its applications.