

# 判別関数のクラスにおける高次元漸近展開とその応用

一橋大学経済研究所

松本智恵子

2004年8月4~7日

統計 SummerSeminar2004 発表資料

## 1. 導入

観測された  $p$  次元の値  $\boldsymbol{x}$  が二つの母集団  $\Pi_1: N_p(\boldsymbol{\mu}_1, \Sigma)$  と  $\Pi_2: N_p(\boldsymbol{\mu}_2, \Sigma)$  のどちらに属するかを判別する問題を考える. 母集団パラメーターである平均ベクトル  $\boldsymbol{\mu}_i$  と分散共分散行列  $\Sigma$  が未知の場合, これらのパラメーターは母集団からの  $N_1 + N_2$  個の初期標本  $\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{1N_1} \in \Pi_1$  と  $\boldsymbol{x}_{21}, \dots, \boldsymbol{x}_{2N_2} \in \Pi_2$  によって推定された標本平均ベクトル  $\bar{\boldsymbol{x}}_i$  とプールされた標本分散共分散行列  $S$  で代用される. 但し,

$$\bar{\boldsymbol{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}, \quad S = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{N_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)', \quad (n = N_1 + N_2 - 2)$$

であるとする.

このような仮定の下で, 一般的な判別関数のクラス

$$\begin{aligned} T &= T(\boldsymbol{x}) \\ &= \frac{1}{2} \left\{ (1+a)(\boldsymbol{x} - \bar{\boldsymbol{x}}_1)' S^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_1) - (1-a)(\boldsymbol{x} - \bar{\boldsymbol{x}}_2)' S^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_2) \right\} - b \end{aligned} \quad (1.1)$$

を考える. 但し,  $a$  と  $b$  は  $m \rightarrow \infty$  で  $a = O(m^{-1})$ ,  $b = O(1)$  を満たすとある定数であるとし,  $m$  は標本数  $N_i$  と次元  $p$  のどちらかを取るパラメーターであるとする. この統計量より次の判別ルール

$$T < 0 \Leftrightarrow \boldsymbol{x} \in \Pi_1, \quad T \geq 0 \Leftrightarrow \boldsymbol{x} \in \Pi_2$$

が導き出される. 又, この判別関数 (1.1) は, 定数  $a$  と  $b$  を色々と変えることにより判別関数の一つのクラスを構成することに注意する. 実際,  $a = 0$ ,  $b = -\log c$  と置くと,  $T$  は

$$\left\{ \boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \right\}' S^{-1} (\bar{\boldsymbol{x}}_2 - \bar{\boldsymbol{x}}_1) + \log c$$

と表され, これは W-rule (A.14) を示している. 又,

$$a = (a_1 - a_2)/(a_1 + a_2), \quad a_1 = N_1/(N_1 + 1), \quad a_2 = N_2/(N_2 + 1),$$

$b = 0$  と置くと,  $T$  は

$$\frac{N_1}{N_1 + 1} (\mathbf{x} - \bar{\mathbf{x}}_1)' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) - \frac{N_2}{N_2 + 1} (\mathbf{x} - \bar{\mathbf{x}}_2)' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2)$$

と表され, これは **Z-rule** (A.17) を示している (判別分析と 2 母集団での判別関数については, Anderson (1984) や付録 A.1 を参照のこと). 今発表では, 次元  $p$  と標本数  $N_i$  がともに大きいと仮定したときの判別関数のクラス (1.1) についての誤判別確率についての漸近展開を引き出し, その漸近展開式を用いた応用について考察する. 第 2 節では, 誤判別確率の漸近展開式について, これまでの研究結果を振り返り, 高次元大標本における誤判別確率の漸近近似式を導出する. 第 3 節では, 第 2 節で導出した漸近展開式を用いて, 誤判別確率の不偏推定量や判別関数のクラス内で最適な判別関数の導出, そして, Student 化した判別関数の分布の漸近展開についての考察を行う. なお, 第 4 節では第 2 節と第 3 節で導出された結果についての数値実験を行っており, 付録にて判別関数と漸近展開についての説明を行っている.

## 2. 誤判別確率の漸近展開

### 2.1 これまでの研究

標本線形判別関数 **W-rule** (A.14) や最大尤度法 **Z-rule** (A.17), そして標本 2 次判別関数 (A.13) はそのシンプルな形と正規性の下での最適性の為, 判別分析において広く使われている. しかしながら, その分布の正確な表現は扱うにはとても複雑なものなので, その代わりに漸近的な近似がよく使われている.

母集団からの標本数  $N_i$  が十分大きい場合について, Okamoto (1963, 1968) では, **W-rule** に関して  $O_2(n)$  の項までの漸近展開が

$$P(2|1) = \Phi\left(-\frac{\Delta}{2}\right) + \frac{a_1}{N_1} + \frac{a_2}{N_2} + \frac{a_3}{n} + O_2(n), \quad (2.1)$$

という形で求められている. 但し,

$$\begin{aligned} a_1 &= (2\Delta^2)^{-1}(d_0^{(4)} + 3pd_0^{(2)}), \quad a_2 = (2\Delta^2)^{-1}\{d_0^{(4)} - (p-4)d_0^{(2)}\}, \\ a_3 &= \frac{1}{2}(p-1)d_0^{(2)}, \quad d_0^{(i)} = \frac{d^i}{dc^i}\Phi(c)|_{c=-\Delta/2} \quad (i = 2, 4), \\ \Delta^2 &= (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2), \quad (\text{Mahalanobis Distance}) \end{aligned}$$

であるとする. 又, Memon and Okamoto (1971) では, **Z-rule** に関して  $O_2(n)$  の項までの漸近展開が

$$P(2|1) = \Phi\left(-\frac{\Delta}{2}\right) + \frac{\tilde{a}_1}{N_1} + \frac{\tilde{a}_2}{N_2} + \frac{\tilde{a}_3}{n} + O_2(n), \quad (2.2)$$

という形で求められている. 但し,

$$\tilde{a}_1 = (2\Delta^2)^{-1}(-d_0^{(4)} + (p-4)d_0^{(2)}),$$

$$\tilde{a}_2 = (2\Delta^2)^{-1}\{3d_0^{(4)} + (p+8)d_0^{(2)}\}, \quad \tilde{a}_3 = \frac{1}{2}(p-1)d_0^{(2)},$$

であるとする. 更に, Siotani and Wang (1977) では, W-rule と Z-rule に関しての  $O_3(n)$  の項までの誤判別確率の漸近展開が求められている. 但し,  $O_j(n)$  は  $(N_1^{-1}, N_2^{-1}, n^{-1})$  に依存する  $j$  次の order を表しているものとする. これらの式の正確さは, 母集団の次元  $p$  と母集団間の Mahalanobis 距離  $\Delta$  ( $\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$ ) に依存している為, 高次元の仮定の下では近似が悪くなってしまうことが知られている. それゆえに, 高次元大標本における漸近展開式についての考察が必要になる.

次元と標本数が共に大きくなると仮定した時の判別関数の分布や誤判別確率の漸近近似についてもこれまでに様々な研究がなされている. W-rule に関しては, Deep(1970) により  $N_1 = N_2$  を仮定した時の誤判別確率についての漸近展開が導出されている. 又, Wyman et al (1990) では,  $N_1 = N_2$  の時の W-rule についての幾つかの近似の精度を比較し, Raudys(1972) の近似が彼らの論文で調べたパラメータの組み合わせではもっとも精度が良いことが示されている. Saradanasa (1993) では, Z-rule における誤判別確率の極限分布を, 独立だが同一分布に従わない確率変数の和に関する形式的な Edgeworth 展開を用いて

$$P(2|1) \simeq \Phi\left(-\frac{\Delta}{2}\sqrt{1-y}\right), \quad (2.3)$$

という形で導出している. 但し,  $n = N_1 + N_2 - 2$  であり,  $n \rightarrow \infty$  で  $p/n \rightarrow y \in (0, 1)$  あるとする. Fujikoshi and Seo (1998) では, W-rule や Z-rule を含むクラス

$$\tilde{T} = \frac{1}{2} \{(\mathbf{X} - \bar{\mathbf{X}}_2)' S^{-1} (\mathbf{X} - \bar{\mathbf{X}}_2) - d(\mathbf{X} - \bar{\mathbf{X}}_1)' S^{-1} (\mathbf{X} - \bar{\mathbf{X}}_1)\}, \quad (2.4)$$

について, 分布や誤判別確率の極限分布を求めている. この  $\tilde{T}$  は, (1.1) 式の  $T$  で  $a = -(d-1)/(d+1)$ ,  $b = 0$  としたものであることに注意する. 彼らにより導出された誤判別確率の極限分布は,

$$P(2|1) \simeq \Phi\left(-\frac{\zeta}{\sigma}\right), \quad (2.5)$$

という形で表されている. 但し,

$$\zeta = \frac{1}{2} \left(\frac{N}{N-p}\right) \left\{ \Delta^2 + \frac{p}{N_1 N_2} (N_1 - N_2) + p(1-d)(1 + N_1^{-1}) \right\},$$

$$\sigma^2 = \left(\frac{N}{N-p}\right)^3 \left( \Delta^2 + \frac{pN}{N_1 N_2} \right), \quad N = N_1 + N_2,$$

であるとする. 数値実験によって, この近似 (2.5) が次元の小さいところでも良い近似をしていることが示されている. ここで, (2.3) と Z-rule における (2.5) の結果が同じではないことに注意する.

現在, Tonda and Wakaki (2003) において, W-rule に関する誤判別確率の漸近展開が

$$P(2|1) = \Phi(\nu) + \phi(\nu)f_1(\Delta) + O_{3/2}, \quad (2.6)$$

という形で導出されている. 但し,

$$\nu = \nu(\Delta)$$

$$= -\frac{1}{2} \left( \frac{N-p}{N-1} \right)^{1/2} \left\{ \Delta^2 + \frac{(p-1)(N_1 - N_2)}{N_1 N_2} \right\} \left( \Delta^2 + \frac{N(p-1)}{N_1 N_2} \right)^{-1/2},$$

であり,  $f_1(\Delta)$  は  $O_1$  の項であるとする. ここで,  $O_j$  は  $(N_1^{-1}, N_2^{-1}, p^{-1})$  に依存する  $j$  次の order を表している. 又, この結果を用いて,  $O_1$  までの不偏推定量も求められている. なお, これまでに述べた高次元大標本における極限分布や漸近展開の近似の精度は全て母集団間の Mahalanobis 距離に依存している.

ところで, ここでは特に触れないが, 標本 2 次判別関数 (A.13) についても現在いくつかの研究がなされている. 分散  $\Sigma_i$  ( $i = 1, 2$ ) が  $\Sigma_1 = c\Sigma_2$  であるという条件の下では, Wakaki (1990) において, 極限分布に非心  $\chi^2$ -分布を用いた漸近展開式が大標本の仮定の下で得られている. そして, 分散に全く条件を仮定しないときの分布の漸近展開については, Matsumoto and Wakaki (2003) により,  $O(p^{-1})$  の項までの誤判別確率の漸近展開式が, 高次元大標本の仮定の下で, 独立だが同一分布に従わない確率変数の和に関する形式的な Edgeworth 展開を用いて導出されている.

## 2.2 判別関数のクラスにおける誤判別確率の漸近展開

(1.1) 式の統計量  $T$  の誤判別確率の漸近展開式を求める手順は以下の通りである. まず始めに, 統計量  $T$  を標準正規分布と  $\chi^2$ -分布に従ういくつかの確率変数の関数の形に分解する. 次に, 分解した統計量  $T$  を標準化し, その標準化した統計量  $T^*$  の特性関数を求める. 求めた特性関数を反転すれば分布関数が得られるので, そこから統計量  $T$  の誤判別確率の漸近展開式を求めることができる.

観測量  $\mathbf{x}$  が  $\Pi_2$  に属している時には,  $\mathbf{x} \in \Pi_1$  のときに導出される結果において  $(a, b, N_1, N_2)$  を  $(-a, -b, N_2, N_1)$  と入れ替えればよいので, 以下,  $\mathbf{x} \in \Pi_1$  の場合のみを考えることにする.

Fujikoshi and Seo (1998) での手法と同様に, 統計量  $T$  を標準正規分布と  $\chi^2$ -分布に従ういくつかの確率変数の関数の形に分解すると, 統計量  $T$  は

$$T = \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_3 - b, \quad (2.7)$$

という形で表すことが出来る. 但し,

$$\begin{aligned} \alpha_1 &= \frac{1}{2} \frac{nN_1}{N_1 + 1} \left\{ (1+a) \left( \frac{N_1 + 1}{N_1} \right)^2 - (1-a) \right\}, \\ \alpha_2 &= -(1-a) \frac{n}{N_1 + 1} \sqrt{\frac{N+1}{N_2}}, \quad \alpha_3 = -\frac{1}{2} (1-a) \frac{n(N+1)}{(N_1 + 1)N_2}, \\ T_1 &= g_1 \cdot g_2^2 + g_3, \quad T_2 = \sqrt{N_1} g_1 \cdot g_2 \cdot g_4^{1/2}, \quad T_3 = g_1 \cdot g_4, \\ g_1 &= \frac{1}{y_2}, \quad g_2 = z_1 + z_2 \left\{ \frac{y_3 y_5}{y_4 (y_6 + z_2^2)} \right\}^{1/2}, \quad g_3 = \frac{y_3}{y_4}, \quad g_4 = (z_3 + \xi)^2 + y_1, \\ z_i &\sim N(0, 1), \quad y_i \sim \chi_{f_i}^2, \\ f_1 &= f_3 = f_5 = p - 1, \quad f_2 = n - p + 1, \quad f_4 = n - p + 2, \quad f_6 = p - 2, \\ \xi &= \xi(\Delta) = \left\{ \frac{N_2(N_1 + 1)}{N + 1} \right\}^{1/2} \Delta, \quad \Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2), \end{aligned}$$

$$N = N_1 + N_2, \quad n = N - 2,$$

であるとする。ここで、 $z_i$  と  $y_i$  はそれぞれ独立であることに留意する。

今回、統計量  $T$  の分布の漸近展開を高次元大標本の仮定の下で導出したいので、次元  $p$  と標本数  $N_i$  が同じ order であると仮定する。漸近展開の計算を分かりやすくする為に、上のパラメーターについて

$$f_i = \rho_i \cdot m, \quad N_i = \lambda_i \cdot m, \quad a = \alpha/m, \quad \xi = \sqrt{m} \cdot \zeta,$$

と置く。但し、 $m$  は次元  $p$  と標本数  $N_i$  のどちらかを取るパラメーターであるとし、 $m \rightarrow \infty$  で  $\rho_i$ ,  $\lambda_i$ ,  $\alpha$ ,  $\zeta = O(1)$  であることに注意する。この時、

$$u_i = \frac{1}{\sqrt{2}\sqrt{f_i}}(y_i - f_i),$$

と置くと、 $u_i$  は  $m \rightarrow \infty$  で漸近的に標準正規分布に従う。この  $u_i$  を用いることにより、(2.7) の  $T_1$ ,  $T_2$ ,  $T_3$  はそれぞれ、

$$\begin{aligned} T_1 &= \frac{\rho_3}{\rho_4} + \frac{1}{\sqrt{m}} \left( \frac{\sqrt{2}\sqrt{\rho_3}}{\rho_4} u_3 - \frac{\sqrt{2}\rho_3}{\rho_4^{3/2}} u_4 \right) \\ &\quad + \frac{1}{m} \left( \frac{2\rho_3}{\rho_4^2} u_4^2 - \frac{2\sqrt{\rho_3}}{\rho_4^{3/2}} u_3 u_4 + \frac{z_1^2}{\rho_2} + \frac{2\rho}{\rho_2} z_1 z_2 + \frac{\rho^2}{\rho_2} z_2^2 \right) + O(m^{-3/2}), \\ T_2 &= \frac{\sqrt{\lambda_1 \eta_1}}{\rho_2} (z_1 + \rho z_2) \\ &\quad + \frac{1}{\sqrt{m}} \left\{ \frac{\rho\sqrt{\lambda_1 \eta_1}}{\sqrt{2}\rho_2} \left( \frac{1}{\sqrt{\rho_3}} u_3 z_2 - \frac{1}{\sqrt{\rho_4}} u_4 z_2 + \frac{1}{\sqrt{\rho_5}} u_5 z_2 - \frac{1}{\sqrt{\rho_6}} u_6 z_2 \right) \right. \\ &\quad \left. + \frac{\sqrt{2\lambda_1 \rho_1}}{2\rho_2 \sqrt{\eta_1}} u_1 (z_1 + \rho z_2) + \frac{\sqrt{\lambda_1} \zeta}{\rho_2 \sqrt{\eta_1}} z_3 (z_1 + \rho z_2) - \frac{\sqrt{2\lambda_1 \eta_1}}{\rho_2^{3/2}} u_2 (z_1 + \rho z_2) \right\} \\ &\quad + \frac{1}{m} \left\{ \frac{\rho\sqrt{\lambda_1 \eta_1}}{4\rho_2} \left( -\frac{1}{\rho_3} u_3^2 z_2 + \frac{3}{\rho_4} u_4^2 z_2 - \frac{1}{\rho_5} u_5^2 z_2 + \frac{3}{\rho_6} u_6^2 z_2 - \frac{2}{\rho_6} z_2^3 \right) \right. \\ &\quad \left. + \frac{2\sqrt{\lambda_1 \eta_1}}{\rho_2^2} u_2^2 (z_1 + \rho z_2) + \frac{\sqrt{\lambda_1} \rho_1}{2\rho_2 \sqrt{\eta_1}^3} z_3^2 (z_1 + \rho z_2) - \frac{\sqrt{\lambda_1} \rho_1}{4\rho_2 \sqrt{\eta_1}^3} u_1^2 (z_1 + \rho z_2) \right\} + O(m^{-3/2}), \\ T_3 &= \frac{\eta_1}{\rho_2} + \frac{1}{\sqrt{m}} \left( \frac{\sqrt{2}\sqrt{\rho_1}}{\rho_2} u_1 + \frac{2\zeta}{\rho_2} z_3 - \frac{\sqrt{2}\eta_1}{\rho_2^{3/2}} u_2 \right) \\ &\quad + \frac{1}{m} \left( \frac{1}{\rho_2} z_3^2 - \frac{2\sqrt{\rho_1}}{\rho_2^{3/2}} u_1 u_2 - \frac{2\sqrt{2}\zeta}{\rho_2^{3/2}} z_3 u_2 + \frac{2\eta_1}{\rho_2^2} u_2^2 \right) + O(m^{-3/2}), \end{aligned}$$

のように、(一部漸近的に) 標準正規分布に従う確率変数で表すことができる。但し、 $\eta_1 = \rho_1 + \zeta^2$ ,  $\rho = \sqrt{\rho_3 \rho_5 / \rho_4 \rho_6}$  であるとする。従って、(2.7) 式の統計量  $T$  は

$$T = t_0 + t_1 + t_2, \tag{2.8}$$

のように分解することができる。但し、

$$t_0 = t_{00} + t_{01} z_1 + t_{02} z_2, \tag{2.9}$$

$$\begin{aligned}
t_{00} &= -b + \frac{n}{N_1 + 1} \left\{ (N_1 a + 1) \frac{f_3}{f_4} - \frac{\eta_2(N+1)}{2f_2 N_2} \right\}, \\
t_{01} &= -\frac{n}{N_1 + 1} \sqrt{\frac{N+1}{N_2}} \sqrt{N_1} \frac{\sqrt{\eta_2}}{f_2}, \quad t_{02} = -\frac{n}{N_1 + 1} \sqrt{\frac{N+1}{N_2}} \sqrt{N_1} \frac{\sqrt{\eta_2}}{f_2} \rho, \\
\rho &= \sqrt{\frac{p-1}{N-p}}, \quad \eta_2 = f_1 + \xi^2,
\end{aligned}$$

であり,  $t_1$  と  $t_2$  はそれぞれ  $O(m^{-1/2})$  と  $O(m^{-1})$  の項であるとする. (2.9) 式を用いると, 統計量  $T$  は

$$\begin{aligned}
T^* &= t_0^* + \frac{t_1}{\sigma} + \frac{t_2}{\sigma}, \tag{2.10} \\
t_0^* &= (t_0 - t_{00})/\sigma, \quad \sigma = \sqrt{t_{01}^2 + t_{02}^2},
\end{aligned}$$

と標準化できる. これより,  $T^*$  の特性関数は

$$\begin{aligned}
C_{T^*}(s) &= E(e^{isT^*}) \tag{2.11} \\
&= E \left[ \exp \left\{ is \left( t_0^* + \frac{t_1}{\sigma} + \frac{t_2}{\sigma} \right) \right\} \right] + O(m^{-3/2}) \\
&= E(e^{ist_0^*}) + \frac{1}{\sigma} E(e^{ist_0^*} ist_1) + \frac{1}{\sigma} E \left[ e^{ist_0^*} \left\{ ist_2 + \frac{(is)^2 t_1^2}{2\sigma} \right\} \right] + O(m^{-3/2}),
\end{aligned}$$

と展開できる. この式の期待値を  $z_i$  の密度関数

$$\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$$

と  $u_i$  の密度関数

$$f_{u_i}(x) = \phi(x) \left\{ 1 + \frac{1}{\sqrt{m}} \frac{\sqrt{2}}{3\sqrt{\rho_i}} (x^3 - 3x) \right\} + O(m^{-1})$$

を用いて計算すると, 結局,  $T^*$  の特性関数は

$$C_{T^*}(s) = e^{-s^2/2} \left[ 1 + \left\{ \frac{1}{\sigma} \tau_1(is) + \frac{1}{\sigma^2} \tau_2(is)^2 + \frac{1}{\sigma^3} \tau_3(is)^3 + \frac{1}{\sigma^4} \tau_4(is)^4 \right\} \right] + O(m^{-3/2}), \tag{2.12}$$

と表される. 但し,

$$\begin{aligned}
\tau_1 &= \alpha_{11} \left( \frac{2f_3}{f_4^2} + \frac{1+\rho^2}{f_2} \right) + \alpha_{31} \left( \frac{2\eta_2}{f_2^2} + \frac{1}{f_2} \right) + \alpha_{12} \frac{f_3}{f_4} + \alpha_{32} \frac{\eta_2}{f_2}, \\
\tau_2 &= \alpha_{22}(t_{01} + \rho t_{02}) \frac{\sqrt{N_1 \eta_2}}{f_2} + \alpha_{11}^2 \left( \frac{f_3}{f_4^2} + \frac{f_3^2}{f_4^3} \right) + \alpha_{31}^2 \left( \frac{\eta_2^2}{f_2^3} + \frac{2\eta_2}{f_2^2} - \frac{f_1}{f_2^2} \right) \\
&\quad + \alpha_{21} \left\{ (t_{01} + \rho t_{02}) \left( \frac{2\sqrt{N_1 \eta_2}}{f_2^2} + \frac{\sqrt{N_1} f_1}{4f_2 \eta_2^{3/2}} \right) + \frac{\sqrt{N_1} \eta_2}{4f_2} \rho t_{02} \left( -\frac{1}{f_3} + \frac{3}{f_4} - \frac{1}{f_5} - \frac{3}{f_6} \right) \right\} \\
&\quad + \alpha_{21}^2 \left\{ \frac{1+\rho^2}{4} \left( \frac{2N_1 \eta_2}{f_2^3} + \frac{N_1}{f_2^2} - \frac{N_1 f_1}{2f_2^2 \eta_2} \right) + \frac{\rho^2 \eta_2 N_1}{8f_2^2} \left( \frac{1}{f_3} + \frac{1}{f_4} + \frac{1}{f_5} + \frac{1}{f_6} \right) \right\}, \\
\tau_3 &= \alpha_{11} \frac{1}{f_2} (t_{01} + \rho t_{02})^2 + \alpha_{11} \alpha_{21} t_{02} \left( \frac{\rho \sqrt{N_1} \eta_2}{f_2 f_4} + \frac{f_3 \rho \sqrt{N_1} \eta_2}{f_2 f_4^2} \right) \\
&\quad + \alpha_{31} \alpha_{21} (t_{01} + \rho t_{02}) \left( \frac{2\sqrt{N_1} \eta_2^{3/2}}{f_2^3} + \frac{2\sqrt{N_1} \eta_2}{f_2^2} - \frac{f_1 \sqrt{N_1}}{f_2^2 \sqrt{\eta_2}} \right),
\end{aligned}$$

$$\begin{aligned}
\tau_4 &= -\alpha_{21} \frac{\rho\sqrt{N_1\eta_2}}{2f_2f_6} t_{02}^3 \\
&\quad + \alpha_{21}^2 \left\{ \frac{(t_{01} + \rho t_{02})^2}{4} \left( \frac{2N_1\eta_2}{f_2^3} + \frac{N_1}{f_2^2} - \frac{N_1f_1}{2f_2^2\eta_2} \right) + \frac{\rho^2\eta_2N_1}{8f_2^2} t_{02}^2 \left( \frac{1}{f_3} + \frac{1}{f_4} + \frac{1}{f_5} + \frac{1}{f_6} \right) \right\}, \\
\alpha_{11} &= \frac{N_1n}{N_1+1}a + \frac{n}{N_1+1}, \quad \alpha_{12} = \frac{n}{N_1+1}a + \frac{n}{2N_1(N_1+1)}, \\
\alpha_{21} &= -\frac{n}{N_1+1}\sqrt{\frac{N+1}{N_2}}, \quad \alpha_{22} = \frac{n}{N_1+1}\sqrt{\frac{N+1}{N_2}}a, \quad \alpha_{31} = -\frac{n(N+1)}{2N_2(N_1+1)}, \quad \alpha_{32} = \frac{n(N+1)}{2N_2(N_1+1)}a,
\end{aligned}$$

であるとする。これを反転することにより以下の定理が導出される。

### 定理 2.1

統計量  $T^*$  の分布関数は漸近的に

$$F_{T^*}(x) = \Phi(x) - \phi(x) \left\{ \frac{1}{\sigma}\tau_1 + \frac{1}{\sigma^2}\tau_2h_1(x) + \frac{1}{\sigma^3}\tau_3h_2(x) + \frac{1}{\sigma^4}\tau_4h_3(x) \right\} + O_{3/2}, \quad (2.13)$$

と展開できる。但し、 $h_r(x)$  は  $r$  次の Hermite 多項式

$$h_1(x) = x, \quad h_2(x) = x^2 - 1, \quad h_3(x) = x^3 - 3x,$$

であり、 $O_j$  は  $(N_1^{-1}, N_2^{-1}, p^{-1})$  に依存する  $j$  次の order であるとする。

定理 2.1 より、判別関数のクラス (1.1) についての誤判別確率は次のように表すことができる。

### 定理 2.2

判別関数 (1.1) の誤判別確率は漸近的に

$$\begin{aligned}
P(2|1) &= P(T > 0 | \mathbf{x} \in \Pi_1) = P\left(T^* > -\frac{t_{00}}{\sigma} | \mathbf{x} \in \Pi_1\right) \\
&= 1 - F_{T^*}\left(-\frac{t_{00}}{\sigma}\right) \\
&= \Phi\left(\frac{t_{00}}{\sigma}\right) + \phi\left(-\frac{t_{00}}{\sigma}\right) \left\{ \frac{1}{\sigma}\tau_1 + \frac{1}{\sigma^2}\tau_2h_1\left(-\frac{t_{00}}{\sigma}\right) + \frac{1}{\sigma^3}\tau_3h_2\left(-\frac{t_{00}}{\sigma}\right) + \frac{1}{\sigma^4}\tau_4h_3\left(-\frac{t_{00}}{\sigma}\right) \right\} + O_{3/2},
\end{aligned} \quad (2.14)$$

で与えられる。

(2.14) の第 1 項は、 $(a, b) = (-(d-1)/(d+1), 0)$  としたとき、Fujikoshi and Seo (1998) の結果と同じものとなる。又、W-rule で考えると、定理 2.2 の結果は Tonda and Wakaki (2003) における結果と同じものである。

なお、展開の妥当性に関しては、統計量  $T$  が正規分布と  $\chi^2$ -分布を用いた形で表されているので、Bhattacharya and Ghosh (1978) より問題なく導出することができる。

### 3. 漸近展開式を用いた応用

第3節では、第2節で導出した判別関数のクラスにおける誤判別確率の漸近展開式 (2.14) を用いた応用について議論する。3.1節では誤判別確率の推定と Bias 修正について、3.2節では最適な判別関数について、そして3.3節では Student 化した判別関数の分布の漸近展開について議論している。

#### 3.1 誤判別確率の推定と Bias 修正

定理 2.2 の誤判別確率の展開式 (2.14) には、母集団パラメーター  $\mu_i$  と  $\Sigma$  から成るパラメーター  $\Delta$  が含まれているが、多くの場合  $\Delta$  は未知である為、 $\Delta$  の代わりにその推定量を用いて誤判別確率の計算を行っている。 $\Delta^2$  に含まれる母集団パラメーターをその不偏推定量で置き換えたパラメーター  $D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  について、その期待値の計算結果より、 $\Delta^2$  を

$$D_u^2 = \frac{n-p+1}{n} D^2 - \frac{(p-1)n}{N_1 N_2}, \quad (3.1)$$

で推定することにする（ここで、高次元大標本の仮定を置いていることに注意する）。しかしながら、(3.1) 式をそのまま誤判別確率の漸近展開式 (2.14) に代入すると、結果として導出される誤判別確率の推定値には  $O_{1/2}$  のバイアスが含まれてしまう。そこで、幾つかの修正が必要になってくる。

定理 2.2 より、誤判別確率の漸近展開式は

$$P(2|1) = \Phi(\nu) + \phi(\nu)f(\Delta),$$

と表すことができる。但し、 $\nu = \nu(\Delta^2) = t_{00}/\sigma$  であるとする。これより、Tonda and Wakaki (2003) での計算方法と同様に、

$$Q_{MW} = \Phi(\tilde{\nu}) + Q_1,$$

という誤判別確率の推定値を考える。但し、 $\tilde{\nu} = \nu(D_u^2)$  であり、 $Q_1$  は  $O_1$  の order を持つ項であるとする。 $O_1$  の order までの漸近不偏推定量を構築する為には、 $Q_{MW}$  のバイアスの order が  $O_{3/2}$  になるように  $Q_1$  を定義する必要がある。 $Q_{MW}$  のバイアスは、 $P_{\mathbf{x}_1, \mathbf{x}_2, S}(2|1)$  を正確な誤判別確率を表すものとして、

$$\begin{aligned} \text{Bias}(Q_{MW}) &= E_{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S}[P_{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S}(2|1) - Q_{MW}] \\ &= P(2|1) - E_{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S}[\Phi(\tilde{\nu}) - Q_1], \end{aligned} \quad (3.2)$$

と計算することができる。2.2節で行った統計量  $T$  の特性関数を計算するときの手法と同様に計算することにより次の補題を得る。

#### 補題 3.1

$D_u^2$  の分解と  $\Phi(\tilde{\nu})$  の Taylor 展開により、 $E_{\mathbf{x}_1, \mathbf{x}_2, S}[\Phi(\tilde{\nu})]$  は

$$E_{\mathbf{x}_1, \mathbf{x}_2, S}[\Phi(\tilde{\nu})] = \Phi(\nu) + \phi(\nu)g(\Delta) + O_{3/2}, \quad (3.3)$$

と展開される。但し、

$$\begin{aligned}
g(\Delta) &= \nu'(\Delta^2)g_2(\Delta^2) + \frac{g_1(\Delta^2)}{2} \{ \nu''(\Delta^2) - \nu(\Delta)[\nu'(\Delta^2)]^2 \}, \\
g_1(\Delta^2) &= \left( \frac{N+1}{N_1 N_2} \right)^2 \left\{ 2(p-1) + 4\xi^2 + \frac{2\eta_2}{f_2^2} \right\} + \frac{4(N+1)}{(N_1+1)^2 N_2} \eta_2(1+\rho), \\
g_2(\Delta^2) &= \frac{N+1}{(N_1+1)N_2} \left\{ 1 + \frac{2\eta_2}{f_2} \right\} + \frac{n}{N_1(N_1+1)} \frac{f_3}{f_4} - \frac{N}{N_1 N_2}, \\
\nu'(\Delta^2) &= \left. \frac{d}{dx} \nu(x) \right|_{x=\Delta^2}, \quad \nu''(\Delta^2) = \left. \frac{d^2}{dx^2} \nu(x) \right|_{x=\Delta^2},
\end{aligned} \tag{3.4}$$

であるとする。

定理 2.2 と補題 3.1 より、(3.2) 式は

$$\text{Bias}(Q_{MW}) = \phi(\nu) \{ f(\Delta) - g(\Delta) \} - E[Q_1] + O_{3/2},$$

と計算することができる。もし  $E[Q_1] = \phi(\nu) \{ f(\Delta) - g(\Delta) \}$  であれば、 $Q_{MW}$  のバイアスの order は  $O_{3/2}$  となり、

$$Q_{MW} = \Phi(\tilde{\nu}) + \hat{Q}_1, \quad \hat{Q}_1 = \phi(\tilde{\nu}) \{ f(D_u) - g(D_u) \}, \tag{3.5}$$

と定義される誤判別確率の推定量は  $O_1$  の項まで漸近的に不偏な推定量となる。

### 3.2 最適な判別関数

$c_i$  を、観測値  $\mathbf{x}$  が母集団  $\Pi_i$  に属するときの事前確率と誤判別による損失との積を表しているものとする。この時、判別ルールにおける risk は、 $c_i$  を誤判別時の損失  $c_{i|j}$  と事前確率  $\pi_i$  との積として

$$c_1 P(2|1) + c_2 P(1|2) \tag{3.6}$$

と定義される。一般に、(3.6) 式の値が小さければ小さいほど『良い』判別ルールであると考えられている。この節の目的は、誤判別確率の漸近展開式 (2.14) を用いて (3.6) 式を最小にする係数  $a$  と  $b$  を見つけることにより、(1.1) 式で与えられるクラス内での最適な判別ルールを見つけることである。

大標本のみを仮定した漸近展開においては、Wakaki and Aoshima (2003) により同様の議論が行われ、判別ルールに関する risk を最小にする cut-off-point の形と、幾つかの判別基準についての最適なルールが導出されている。その結果では、次元  $p$  を固定し、 $N_i \rightarrow \infty$  の場合、 $c_2/c_1 = 1$  では Z-rule が最適なものと一致している。

(2.14) 式で表された誤判別確率の漸近展開式を係数  $a$  と  $b$  の関数の形で表すと

$$\begin{aligned}
P(2|1) &= \Phi(-s_1 b + s_2 a + s_{31}) + \phi(-s_1 b + s_2 a + s_{31}) \\
&\times \left\{ \frac{1}{\sigma} (\tau_{10} + \tau_{11} a) + \frac{1}{\sigma^2} (\tau_{20} + \tau_{21} a + \tau_{22} a^2) h_1(s_1 b - s_2 a - s_{31}) \right. \\
&\left. + \frac{1}{\sigma^3} (\tau_{30} + \tau_{31} a) h_2(s_1 b - s_2 a - s_{31}) + \frac{1}{\sigma^4} \tau_4 h_3(s_1 b - s_2 a - s_{31}) \right\} + O_{3/2},
\end{aligned} \tag{3.7}$$

となる。但し,

$$\begin{aligned} s_1 &= \left(\frac{N-1}{N-p}\right)^{-3/2} \left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{-1/2}, & s_2 &= \left(\frac{N-1}{N-p}\right)^{-1/2} \left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{-1/2} p, \\ s_{31} &= -\frac{1}{2} \left(\frac{N-1}{N-p}\right)^{-1/2} \left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{-1/2} \left\{ \Delta^2 + \frac{p(N_1-N_2)}{N_1N_2} \right\}, \\ s_{32} &= -\frac{1}{2} \left(\frac{N-1}{N-p}\right)^{-1/2} \left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{-1/2} \left\{ \Delta^2 - \frac{p(N_1-N_2)}{N_1N_2} \right\}, \\ \sigma &= \sqrt{t_{01}^2 + t_{02}^2}, & \Delta &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \end{aligned}$$

であるとする。(3.7) 式の  $(a, b, N_1, N_2)$  を  $(-a, -b, N_2, N_1)$  と置き換えることにより,  $P(1|2)$  も

$$\begin{aligned} P(1|2) &= \Phi(s_1b - s_2a + s_{32}) + \phi(s_1b - s_2a + s_{32}) \\ &\times \left\{ \frac{1}{\sigma} (\tau'_{10} - \tau_{11}a) + \frac{1}{\sigma^2} (\tau_{20} - \tau'_{21}a + \tau_{22}a^2) h_1(-s_1b + s_2a - s_{32}) \right. \\ &\left. + \frac{1}{\sigma^3} (\tau'_{30} - \tau_{31}a) h_2(-s_1b + s_2a - s_{32}) + \frac{1}{\sigma^4} \tau_4 h_3(-s_1b + s_2a - s_{32}) \right\} + O_{3/2}, \quad (3.8) \end{aligned}$$

と表すことができる。但し,  $\tau_{ij}$  は  $N_1, N_2, p, \Delta$  より構成されるパラメーターであり,  $\tau'_{ij}$  は  $\tau_{ij}$  の  $N_1$  と  $N_2$  を入れ替えたものであることに注意する。(3.7) 式と (3.8) 式を用いることにより, risk (3.6) 式を漸近的に最小にする係数  $a$  と  $b$  を導出することができる。しかしながら, (3.7) 式と (3.8) 式の複雑さより, 係数  $a$  と  $b$  を直接得るのは難しい。そこで, まず始めに (3.6) 式の先導項

$$c_1 \Phi(-s_1b + s_2a + s_{31}) + c_2 \Phi(s_1b - s_2a + s_{32}) \quad (3.9)$$

を最小にすることを考えたい。すると,

$$s_2a - s_1b + \frac{s_{31} - s_{32}}{2} = (s_{31} + s_{32})^{-1} \log \frac{c_1}{c_2} \quad (3.10)$$

という係数  $a$  と  $b$  の関係式を得ることができる。この関係式を基にして, (3.7) 式と (3.8) 式の  $O_{3/2}$  の項を無視して計算することにより, (1.1) 式のクラス内で最適な判別ルールを形成する係数  $a$  と  $b$  について以下の定理を得ることができる。

### 定理 3.1

(3.10) 式の条件の元で, (1.1) のクラス内での最適な判別ルールは係数  $(a, b) = (a_0(\Delta^2), b_0(\Delta^2))$  の下で形成される。但し,

$$\begin{aligned} a_0(\Delta^2) &= \left[ \frac{c_2 \phi(\gamma_2) \tau_{11}}{\theta^{1/2}} - \frac{c_1 \phi(\gamma_1) \tau_{11}}{\theta^{1/2}} - \frac{c_2 \phi(\gamma_2) \tau'_{21} \gamma_2}{\theta} + \frac{c_1 \phi(\gamma_1) \tau_{21} \gamma_1}{\theta} \right. \\ &\left. + \frac{c_2 \phi(\gamma_2) \tau_{31} (-1 + \gamma_2^2)}{\theta^{3/2}} - \frac{c_1 \phi(\gamma_1) \tau_{31} (-1 + \gamma_1^2)}{\theta^{3/2}} \right] \left[ -\frac{2c_2 \phi(\gamma_2) \tau_{22} \gamma_2}{\theta} - \frac{2c_1 \phi(\gamma_1) \tau_{22} \gamma_1}{\theta} \right]^{-1}, \quad (3.11) \end{aligned}$$

$$b_0(\Delta^2) = \frac{s_2}{s_1} a_0(\Delta^2) + \frac{s_{31} - s_{32}}{2s_1} - \frac{\log \frac{c_1}{c_2}}{s_1(s_{31} + s_{32})}, \quad (3.12)$$

$$\gamma_1 = \frac{s_{31} + s_{32}}{2} + \frac{\log \frac{c_1}{c_2}}{s_{31} + s_{32}}, \quad \gamma_2 = \frac{s_{31} + s_{32}}{2} - \frac{\log \frac{c_1}{c_2}}{s_{31} + s_{32}},$$

$$\theta = \left(\frac{N-1}{N-p}\right)^{3/2} \left(\Delta^2 + \frac{pN}{N_1N_2}\right)^{1/2},$$

であり、 $\phi(x)$  は標準正規分布の密度関数を表しているものとする。

定理 3.1 内の係数  $a_0(\Delta^2)$  と  $b_0(\Delta^2)$  には未知パラメーター  $\Delta$  が含まれている。そこで、 $\Delta^2$  を (3.1) 式の  $D_u^2$  で推定することにする。

### 3.3 Student 化した判別関数の漸近展開

判別関数を、その極限分布の平均値と分散の推定量で『標準化』したとき、その統計量は“Student 化した判別関数”と呼ばれる。Anderson (1973a, b) は、大標本の仮定の下で、W-rule の Student 化した判別関数

$$W_{S1} = \frac{W - D^2/2}{D} \quad (\text{when } \mathbf{x} \in \Pi_1), \quad W_{S2} = \frac{W + D^2/2}{D} \quad (\text{when } \mathbf{x} \in \Pi_2), \quad (3.13)$$

の分布の漸近展開を  $O_1(n)$  の項まで求めている。又、Z-rule については、Fujikoshi and Kanazawa (1976) において、Student 化した判別関数

$$Z_{S1} = \frac{Z + D^2}{2D} \quad (\text{when } \mathbf{x} \in \Pi_1), \quad Z_{S2} = \frac{Z - D^2}{2D} \quad (\text{when } \mathbf{x} \in \Pi_2), \quad (3.14)$$

の分布の大標本における漸近展開式が  $O_1(n)$  の項まで求められている。但し、 $W$  と  $Z$  はそれぞれ W-rule と Z-rule の式を表し、 $D$  は標本 Mahalanobis 距離  $D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  を表しているものとする。今節では、2.2 節で導出した判別関数のクラスにおける極限分布での平均  $t_{00}$  と分散  $\sigma^2$  の、(3.1) 式で表される  $\Delta$  の推定量  $D_u$  を用いて導出した推定量で (1.1) 式の統計量  $T$  を Student 化し、誤判別確率の漸近展開式を導出した時に用いた計算方法を応用してその分布の漸近展開式を導出する。なお、Student 化した判別関数の分布の漸近展開を求める利点としては、指定された  $\alpha$  に対し、 $P[T(\mathbf{x}; \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S) \geq c | \Pi_1] = \alpha + O_j$  となるような cut-off-point  $c$  を未知母数  $\Delta$  ではなくその推定量を用いて与えることができるという点が挙げられる。

以下、2.2 節と同様に、 $\mathbf{x} \in \Pi_1$  の時のみについて考える。この時、統計量  $T$  の Student 化した判別関数は、極限分布の平均  $t_{00}$  と分散  $\sigma^2$  の、 $\Delta$  を  $D_u$  で推定した推定量を用いて

$$T_{S1} = \frac{T - t_{00}(D_u)}{\sigma(D_u)} \quad (3.15)$$

と表される。但し、

$$\begin{aligned} t_{00} &= -b + \frac{n}{N_1 + 1} \left\{ (N_1 a + 1) \frac{f_3}{f_4} - \frac{\eta_2(N + 1)}{2f_2 N_2} \right\}, \\ t_{01} &= -\frac{n}{N_1 + 1} \sqrt{\frac{N + 1}{N_2}} \sqrt{N_1} \frac{\sqrt{\eta_2}}{f_2}, \quad t_{02} = \rho t_{01}, \\ \sigma &= \sqrt{t_{01}^2 + t_{02}^2}, \quad \rho = \sqrt{\frac{p - 1}{N - p}}, \quad \eta_2 = f_1 + \xi^2, \quad \xi = \left\{ \frac{N_2(N_1 + 1)}{N + 1} \right\}^{1/2} \Delta, \\ D_u^2 &= \frac{n - p + 1}{n} D^2 - \frac{(p - 1)N}{N_1 N_2}, \quad D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \end{aligned}$$

であり,  $t_{00}(D_u)$  と  $\sigma(D_u)$  は  $t_{00}$  と  $\sigma$  の  $\Delta$  を  $D_u$  に置き換えたものであるとする.

統計量  $T_{S1}$  の分布の漸近展開についての計算方法は,  $D_u$  を分解して  $t_{00}(D_u)$  と  $\sigma(D_u)$  にあてはめる  
他は 2.2 節で行った  $T$  の分布の漸近展開を導出する方法と同様である. その結果,  $T_{S1}$  は

$$\begin{aligned} T_{S1} &= \frac{t_0 - t_{00}}{\sigma} + \frac{t_1 - \tau d_1}{\sigma} - \frac{(t_0 - t_{00})\varsigma d_1}{2\sigma^3} \\ &\quad + \frac{t_2 - \tau d_2}{\sigma} - \frac{(t_1 - \tau d_1)\varsigma d_1}{2\sigma^3} + (t_0 - t_{00}) \left( \frac{3\varsigma^2 d_1^2}{8\sigma^5} - \frac{\varsigma d_2}{2\sigma^3} \right) \\ &= t_0^* + t_{s1} + t_{s2} \end{aligned} \quad (3.16)$$

のように分解することができる. 但し,

$$\begin{aligned} d_1 &= \frac{(N+1)(n-p+1)}{N_1 N_2} \left( \frac{\sqrt{2}\sqrt{f_1}}{f_2} u_1 + \frac{2\xi}{f_2} z_3 - \frac{\sqrt{2}\eta_2}{f_2^{3/2}} u_2 \right) - \frac{2\sqrt{(N+1)\eta_2}}{N_1\sqrt{N_2}} (z_1 + \rho z_2), \\ d_2 &= \frac{(N+1)(n-p+1)}{N_1 N_2} \left( \frac{1}{f_2} z_3^2 - \frac{2\sqrt{f_1}}{f_2^{3/2}} u_1 u_2 - \frac{2\sqrt{2}\xi}{f_2^{3/2}} z_3 u_2 + \frac{2\eta_2}{f_2^2} u_2^2 \right) \\ &\quad - \frac{2(n-p+1)\sqrt{N+1}}{N_1^{3/2}\sqrt{N_2}} \left\{ \frac{\rho\sqrt{N_1}\eta_2}{\sqrt{2}f_2} \left( \frac{1}{\sqrt{f_3}} u_3 z_2 - \frac{1}{\sqrt{f_4}} u_4 z_2 + \frac{1}{\sqrt{f_5}} u_5 z_2 - \frac{1}{\sqrt{f_6}} u_6 z_2 \right) \right. \\ &\quad \left. + \frac{\sqrt{2N_1}f_1}{2f_2\sqrt{\eta_2}} u_1 (z_1 + \rho z_2) + \frac{\sqrt{N_1}\xi}{f_2\sqrt{\eta_2}} z_3 (z_1 + \rho z_2) - \frac{\sqrt{2N_1}\eta_2}{f_2^{3/2}} u_2 (z_1 + \rho z_2) \right\} + \frac{n-p+1}{N_1^2} \frac{f_3}{f_4}, \\ \tau &= -\frac{n}{2(n-p+1)}, \quad \varsigma = \frac{n^2}{(n-p+1)^2} (1 + \rho^2), \\ t_{s1} &= \frac{t_1 - \tau d_1}{\sigma} - \frac{(t_0 - t_{00})\varsigma d_1}{2\sigma^3}, \\ t_{s2} &= \frac{t_2 - \tau d_2}{\sigma} - \frac{(t_1 - \tau d_1)\varsigma d_1}{2\sigma^3} + (t_0 - t_{00}) \left( \frac{3\varsigma^2 d_1^2}{8\sigma^5} - \frac{\varsigma d_2}{2\sigma^3} \right), \end{aligned}$$

であるとする. 従って,  $T_{S1}$  の特性関数は

$$C_{T_{S1}} = E(e^{ist_0^*}) + E(e^{ist_0^*} i s t_{s1}) + E \left[ e^{ist_0^*} \left\{ i s t_{s2} + \frac{(is)^2}{2} t_{s1}^2 \right\} \right] + O(m^{-3/2}) \quad (3.17)$$

と表すことができるので, この期待値を 2.2 節と同様に計算し, 反転することにより以下の定理が導出される.

### 定理 3.2

統計量  $T_{S1}$  の分布は漸近的に

$$\begin{aligned} F_{T_{S1}}(x) &= \Phi(x) - \phi(x) [\{\omega_{21} + \omega_{22}h_1(x) + \omega_{23}h_2(x)\} \\ &\quad + \{\omega_{31} + \omega_{32}h_1(x) + \omega_{33}h_2(x) + \omega_{34}h_3(x) + \omega_{35}h_4(x) + \omega_{36}h_5(x)\}] + O_{3/2}, \end{aligned} \quad (3.18)$$

と表される. 但し,

$$\begin{aligned} h_1(x) &= x, \quad h_2(x) = x^2 - 1, \quad h_3(x) = x^3 - 3x, \\ h_4(x) &= x^4 - 6x^2 + 3, \quad h_5(x) = x^5 - 10x^3 + 15x, \end{aligned}$$

$$\begin{aligned}
\omega_{21} &= \omega_{23} = \frac{g_6 \varsigma}{t_{01}^2 (1 + \rho^2)^{1/2}}, & \omega_{22} &= \frac{2g_6 \tau}{t_{01}}, \\
\omega_{31} &= -\frac{\tau}{\sigma} g_1 - \frac{\alpha_{31}}{f_2} g_5 g_2 \frac{\varsigma}{2\sigma^3} + \frac{\tau \varsigma}{2\sigma^3} g_5^2 g_2 + \frac{2\tau \varsigma}{\sigma^3} g_6^2 (1 + \rho^2) \\
&+ \frac{\alpha_{11}}{\sigma} \left\{ \frac{2f_3}{f_4^2} + \frac{1}{f_2} (1 + \rho^2) \right\} + \frac{\alpha_{31}}{\sigma} \left( \frac{1}{f_2} + \frac{2\eta_2}{f_2^2} \right) + \frac{\alpha_{12}}{\sigma} \frac{f_3}{f_4} + \frac{\alpha_{32}}{\sigma} \frac{\eta_2}{f_2}, \\
\omega_{32} &= g_8 \frac{\alpha_{11}^2}{\sigma^2} \frac{f_3}{f_4} + \frac{\alpha_{31}^2}{2\sigma^2} \frac{g_2}{f_2^2} + \frac{\alpha_{21}^2}{2\sigma^2} \left\{ \frac{N_1 \eta_2}{2f_2^2} g_7 \rho^2 + (1 + \rho^2) g_4 \right\} \\
&- \alpha_{21} g_3 g_5 \frac{\varsigma}{\sigma^4} t_{01} (1 + \rho^2) + \frac{\alpha_{21}}{\sigma^2} t_{01} \left\{ \left( \frac{2\sqrt{N_1 \eta_2}}{f_2^2} + \frac{\sqrt{N_1} f_1}{4f_2 \sqrt{\eta_2}} \right) (1 + \rho^2) \right. \\
&+ \left. \frac{\sqrt{N_1 \eta_2}}{4f_2} \left( -\frac{1}{f_3} + \frac{3}{f_4} - \frac{1}{f_5} + \frac{3}{f_6} \right) \rho^2 - \frac{3\sqrt{N_1 \eta_2}}{2f_2 f_6} \rho^2 \right\} \\
&+ \frac{\alpha_{22}}{\sigma^2} \frac{\sqrt{N_1 \eta_2}}{f_2} t_{01} (1 + \rho^2) - \alpha_{31} g_2 g_5 \frac{\tau}{f_2 \sigma^2} - g_1 \frac{\varsigma}{2\sigma^4} t_{01}^2 (1 + \rho^2) \\
&+ g_2 g_5^2 \frac{\tau^2}{2\sigma^2} + g_6^2 \frac{2\tau^2}{\sigma^2} (1 + \rho^2) + g_2 g_5^2 \frac{\varsigma^2}{2\sigma^6} t_{01}^2 (1 + \rho^2) + g_6^2 \frac{6\varsigma^2}{\sigma^6} t_{01}^2 (1 + \rho^2)^2, \\
\omega_{33} &= \alpha_{11} \alpha_{21} g_8 \frac{\sqrt{N_1 \eta_2}}{2f_2 \sigma^3} t_{01} \rho^2 + \alpha_{21} \alpha_{31} \frac{g_3}{2f_2 \sigma^3} t_{01} (1 + \rho^2) - \alpha_{21} g_3 g_5 \frac{\tau}{\sigma^3} t_{01} (1 + \rho^2) \\
&+ \frac{\alpha_{11}}{f_2 \sigma^3} t_{01}^2 (1 + \rho^2)^2 - \alpha_{31} g_2 g_5 \frac{\varsigma}{2f_2 \sigma^5} t_{01}^2 (1 + \rho^2) + g_6^2 \frac{8\tau \varsigma}{\sigma^5} t_{01}^2 (1 + \rho^2)^2 + g_2 g_5^2 \frac{\tau \varsigma}{2\sigma^5} t_{01}^2 (1 + \rho^2), \\
\omega_{34} &= \frac{\alpha_{21}^2}{2\sigma^4} t_{01}^2 \left\{ \frac{N_1 \eta_2}{2f_2^2} g_7 \rho^4 + g_4 (1 + \rho^2)^2 \right\} - \alpha_{21} \frac{\sqrt{N_1 \eta_2}}{2f_2 f_6 \sigma^4} t_{01}^3 \rho^4 - g_3 g_5 \frac{\varsigma}{2\sigma^6} t_{01}^3 (1 + \rho^2)^2 \\
&+ g_6^2 \frac{2\tau^2}{\sigma^4} t_{01}^2 (1 + \rho^2)^2 + g_6^2 \frac{9\varsigma^2}{2\sigma^8} t_{01}^4 (1 + \rho^2)^3 + g_2 g_5^2 \frac{\varsigma^2}{8\sigma^8} t_{01}^4 (1 + \rho^2)^2, \\
\omega_{35} &= g_6^2 \frac{2\tau \varsigma}{\sigma^7} t_{01}^4 (1 + \rho^2)^3, & \omega_{36} &= g_6^2 \frac{\varsigma^2}{2\sigma^{10}} t_{01}^6 (1 + \rho^2)^4, \\
g_1 &= \frac{n - p + 1}{N_1^2} \frac{f_3}{f_4} + \frac{N + 1}{N_1 N_2} + \frac{2(N + 1)}{N_1 N_2} \frac{\eta_2}{f_2}, & g_2 &= 2f_1 + 4\xi^2 + \frac{2\eta_2^2}{f_2}, \\
g_3 &= \frac{1}{f_2} \frac{\sqrt{N_1}}{\sqrt{\eta_2}} \left( f_1 + 2\xi^2 + \frac{2\eta_2^2}{f_2} \right), & g_4 &= \frac{N_1 f_1}{2f_2^2 \eta_2} + \frac{N_1 \xi^2}{f_2^2 \eta_2} + \frac{2N_1 \eta_2}{f_2^3}, & g_5 &= \frac{N + 1}{N_1 N_2}, \\
g_6 &= \frac{\sqrt{N + 1} \sqrt{\eta_2}}{N_1 \sqrt{N_2}}, & g_7 &= \frac{1}{f_3} + \frac{1}{f_4} + \frac{1}{f_5} + \frac{1}{f_6}, & g_8 &= \frac{1}{f_4} + \frac{f_3}{f_4^2},
\end{aligned}$$

であるとする。

なお、 $\mathbf{x} \in \Pi_2$  の時の Student 化した統計量

$$T_{S2} = \frac{T - t'_{00}(D_u)}{\sigma'(D_u)}, \quad (3.19)$$

但し、 $t'_{00}(D_u)$  と  $\sigma'(D_u)$  は  $t_{00}(D_u)$  と  $\sigma(D_u)$  の  $(a, b, N_1, N_2)$  を  $(-a, -b, N_2, N_1)$  としたものであるが、この分布の漸近展開については、 $T_{S1}$  の結果で  $(a, b, N_1, N_2)$  を  $(-a, -b, N_2, N_1)$  とすればよい。

なお、 $T_{S1}$  と  $T_{S2}$  は判別関数のクラス (1.1) を Student 化したものなので、導出した漸近展開式 (3.18) を利用して、 $P(1|2)$  の値を指定した上で  $P(2|1)$  の値を最小にする係数  $a$  と  $b$  を求めることもできるだろう。

## 4. Simulation

この節では、これまでに導出した結果について数値的に考察する。4.1節では、2.2節で導出した漸近展開式 (2.14) について、次元と標本数が有限の場合にきちんと近似されているかどうかを、3.1節で導出した Bias 修正の結果と合わせて考察する。4.2節では、3.2節で導出した『最適な』判別ルールが本当に判別ルールを改良しているのかどうかを数値的に考察する。

### 4.1 誤判別確率の漸近展開に関する Simulation

パラメーターの値を

$$p; 3, 5, 10,$$

$$(N_1, N_2); (10, 10), (10, 20), (10, 30), (20, 10), (20, 20), (20, 40), (30, 10), (40, 20),$$

$$\Delta; 1.05, 1.68, 2.56, 3.29,$$

$$c = 1,$$

$$\mu_1 = -\mu_2 = (\Delta/2, 0, \dots, 0),$$

$$\Sigma = I_p = \text{diag}(1, \dots, 1).$$

と置く。この値を用いて初期標本を作製し、(1.1) 式を用いて W-rule と Z-rule の場合での誤判別確率のシミュレーション値を算出する。又、上記のパラメーターを導出した誤判別確率の漸近展開式 (2.14) に当てはめて誤判別確率の漸近近似値を算出し、シミュレーション値と比較した。更に、漸近展開式内にある母集団パラメーターに初期標本から作成した推定値を代入して得た結果と、漸近展開式を利用して導出した漸近不偏推定量における結果も算出している。それに加えて、大標本のみを仮定した場合の漸近展開式である Okamoto (1963, 1968) と Memon and Okamoto (1971) の結果での値も算出し、比較対象としている。

導出された結果は表 1 で示されている。但し、"Sim.", " $W_O$ ", " $Z_{MO}$ ", " $W_{AE}$ ", " $Z_{AE}$ ", " $W_{Est}$ ", " $Z_{Est}$ ", " $W_{Bias}$ " and " $Z_{Bias}$ " はそれぞれ次のような意味である。

- "Sim.": Monte-Carlo 法によって推定された誤判別確率.
- " $W_O$ ": Okamoto (1963, 1968) によって導出された漸近近似.
- " $Z_{MO}$ ": Memon and Okamoto (1971) によって導出された漸近近似.
- " $W_{AE}$ " and " $Z_{AE}$ ": (2.14) より導出された漸近展開による値.
- " $W_{Est}$ " and " $Z_{Est}$ ":  $\Delta$  を (3.1) で置き換えて作成された (2.14) の推定量の期待値.
- " $W_{Bias}$ " and " $Z_{Bias}$ ": 誤判別確率の漸近不偏推定量の期待値.

表 1:  $P(2|1)$  のシミュレーション値 ( $p = 10, c = 1$ )

$p = 10, c = 1$		W-rule					Z-rule				
$(N_1, N_2)$	$\Delta$	Sim.	$W_O$	$W_{AE}$	$W_{Est}$	$W_{Bias}$	Sim.	$Z_{MO}$	$Z_{AE}$	$Z_{Est}$	$Z_{Bias}$
(10, 10)	1.05	0.425	0.598	0.422	0.442	0.429	0.425	0.499	0.422	0.442	0.429
	1.68	0.338	0.390	0.337	0.360	0.303	0.338	0.340	0.337	0.360	0.303
	2.56	0.229	0.214	0.222	0.252	0.198	0.229	0.193	0.222	0.252	0.198
	3.29	0.159	0.120	0.144	0.171	0.123	0.159	0.111	0.144	0.171	0.123
(10, 20)	1.05	0.448	0.568	0.441	0.464	0.441	0.399	0.442	0.394	0.412	0.391
	1.68	0.348	0.362	0.327	0.352	0.299	0.297	0.296	0.292	0.322	0.238
	2.56	0.210	0.190	0.192	0.213	0.153	0.179	0.161	0.170	0.191	0.133
	3.29	0.124	0.103	0.111	0.133	0.076	0.110	0.089	0.097	0.117	0.069
(10, 30)	1.05	0.463	0.556	0.452	0.479	0.454	0.393	0.421	0.381	0.402	0.380
	1.68	0.335	0.349	0.324	0.344	0.240	0.285	0.278	0.274	0.290	0.200
	2.56	0.181	0.179	0.180	0.202	0.139	0.152	0.148	0.151	0.169	0.109
	3.29	0.111	0.094	0.099	0.102	0.068	0.093	0.080	0.082	0.085	0.051
(20, 10)	1.05	0.350	0.469	0.350	0.362	0.342	0.396	0.447	0.397	0.413	0.392
	1.68	0.263	0.312	0.262	0.277	0.199	0.297	0.302	0.297	0.313	0.229
	2.56	0.181	0.169	0.155	0.191	0.138	0.201	0.167	0.176	0.215	0.155
	3.29	0.100	0.093	0.090	0.106	0.057	0.112	0.093	0.103	0.115	0.063
(20, 20)	1.05	0.380	0.448	0.374	0.394	0.380	0.380	0.398	0.374	0.394	0.380
	1.68	0.274	0.294	0.267	0.288	0.197	0.274	0.269	0.267	0.288	0.197
	2.56	0.156	0.156	0.149	0.149	0.103	0.156	0.145	0.149	0.149	0.103
	3.29	0.091	0.084	0.082	0.084	0.046	0.091	0.079	0.082	0.084	0.046
(20, 40)	1.05	0.389	0.434	0.387	0.399	0.381	0.358	0.370	0.357	0.366	0.349
	1.68	0.267	0.280	0.266	0.276	0.199	0.247	0.248	0.247	0.255	0.186
	2.56	0.144	0.144	0.141	0.141	0.098	0.133	0.130	0.131	0.130	0.087
	3.29	0.080	0.076	0.074	0.075	0.032	0.075	0.069	0.068	0.069	0.026
(30, 10)	1.05	0.321	0.424	0.319	0.335	0.311	0.391	0.427	0.386	0.409	0.384
	1.68	0.238	0.282	0.234	0.252	0.173	0.285	0.286	0.281	0.303	0.213
	2.56	0.139	0.151	0.132	0.133	0.071	0.165	0.155	0.159	0.160	0.101
	3.29	0.083	0.082	0.073	0.075	0.032	0.098	0.085	0.089	0.092	0.040
(40, 20)	1.05	0.326	0.384	0.329	0.338	0.320	0.356	0.373	0.359	0.369	0.353
	1.68	0.230	0.255	0.230	0.238	0.136	0.249	0.251	0.250	0.259	0.148
	2.56	0.128	0.134	0.124	0.124	0.077	0.137	0.133	0.134	0.134	0.088
	3.29	0.070	0.071	0.065	0.067	0.019	0.075	0.071	0.071	0.073	0.025

Monte-Carlo によるシミュレーションの値について、標準偏差はどれも 0.01 未満であった。

表 1 より、 $p$  と  $N_i$  の比が小さい時には  $W_O$  や  $Z_{MO}$  の近似が良いが、 $p$  と  $N_i$  の比が 1 に近い時には、漸近近似の正確さを  $p$  に依存している  $W_O$  や  $Z_{MO}$  は近似が悪く、次元と標本数が共に大きくなると仮定した漸近展開の近似が良いことが見て取れる。又、漸近展開式における推定量に関しては、 $\Delta$  の値が小さいところでは漸近不偏推定量の近似が良いことが見て取れる。 $\Delta$  の値が大きくなるとその近似はあまり良くないが、その理由としては、導出した漸近展開式 (2.14) の近似の良さが  $\Delta$  に依存しているからであろうと推測される。(図 1 参照)

#### 4.2 最適な判別ルールに関する考察

4.1 節と同様に、平均ベクトルと共分散行列について、一般性の損失なしに  $\mu_1 = -\mu_2 = (\Delta/2, 0, \dots, 0)$ ,  $\Sigma = I_p = \text{diag}(1, \dots, 1)$  と仮定できる。この時、パラメーターを

$$p; 5, 10, 15, 20,$$

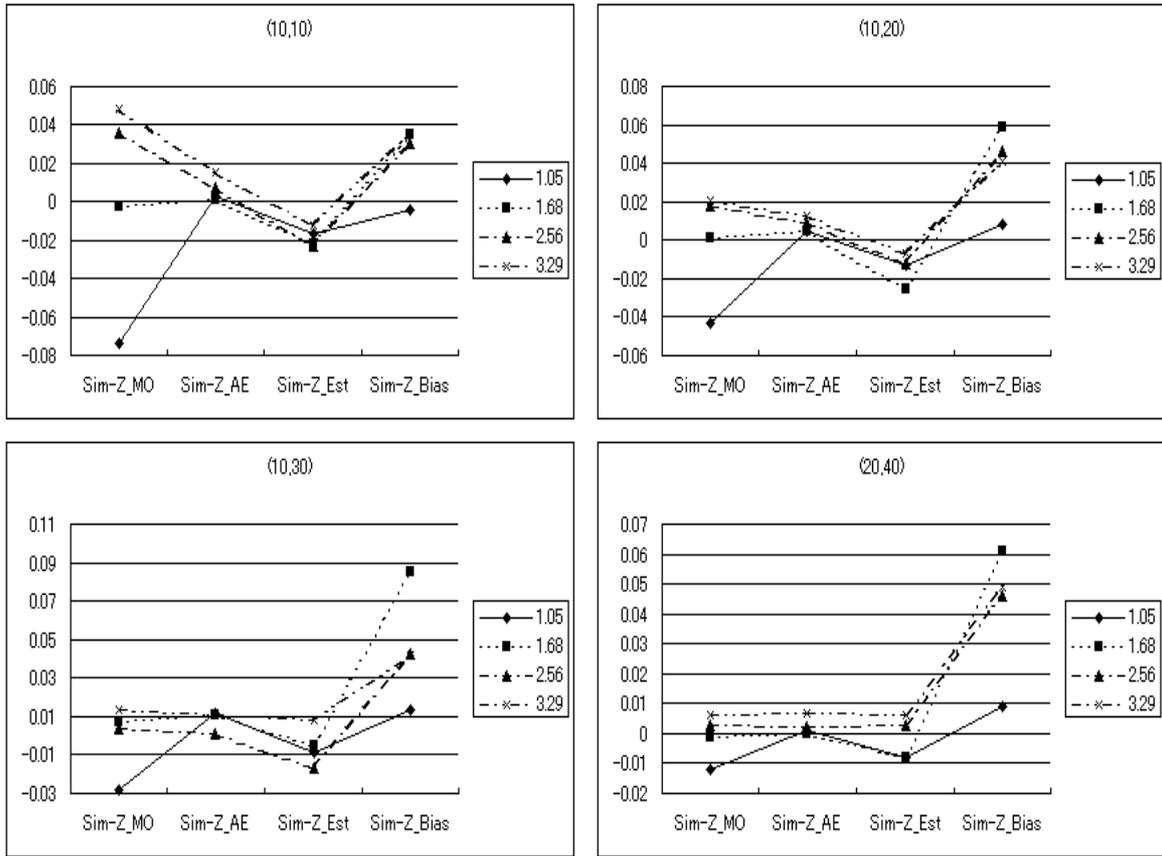
$$(N_1, N_2); (10, 20), (10, 30), (20, 10),$$

$$(20, 20), (20, 40), (30, 10), (40, 20),$$

$$\Delta; 1.05, 1.68, 2.56, 3.29,$$

$$(c_1, c_2); (1, 1), (1, 0.5), (0.75, 0.25).$$

図 1: Z-rule における数値実験結果の差による比較 ( $p = 10, c = 1$ )



と置き、それぞれのパラメーター毎に W-rule, Z-rule, そして 3.1 節で導出した最適化ルール (Min-rule と呼ぶことにする) の risk を Monte-Carlo 法を用いて計算した (但し,  $(N_1, N_2) = (10, 10)$  の場合は  $p = 15, 20$  のときには使わなかった)。

導出された結果は表 2, 3 で示されている。Monte-Carlo によるシミュレーションの値について、標準偏差はどれも 0.01 未満であった。

これらの表において、 $\Delta = 1.05$  のときには、Min-rule の risk は W-rule や Z-rule よりも低く、判別ルールが改良されていることが分かる。一方、 $\Delta = 1.68$  で  $p = 10$  のときには、W-rule や Z-rule の方が Min-rule よりも良い判別ルールであると示されている部分もある。 $c_1 \neq c_2$  のときも、 $c_1 = c_2$  のときと同様の傾向を示している。 $\Delta$  が大きいところで判別ルールが改良されない理由は、4.1 節で考察したように、導出した漸近展開式 (2.14) の近似の良さが  $\Delta$  に依存しているからであろう。なお、表 4 において、 $(a_0(\Delta^2), b_0(\Delta^2))$  の値と Z-rule における係数  $a$  の値を示しておく (W-rule についてはいつも  $a = 0$  である)。

表 2: 三つの判別ルールにおける risk の値 ( $\Delta = 1.05$ )

$\Delta = 1.05$		$p = 10$			$p = 20$		
$(c_1, c_2)$	$(N_1, N_2)$	W-rule	Z-rule	Min-rule	W-rule	Z-rule	Min-rule
(1, 1)	(10, 10)	0.867	0.850	0.718	–	–	–
	(10, 20)	0.808	0.792	0.696	0.900	0.890	0.701
	(10, 30)	0.792	0.779	0.699	0.881	0.858	0.703
	(20, 10)	0.797	0.783	0.693	0.890	0.879	0.704
	(20, 20)	0.769	0.763	0.719	0.853	0.844	0.738
	(20, 40)	0.715	0.710	0.694	0.791	0.781	0.728
	(30, 10)	0.799	0.784	0.699	0.881	0.866	0.702
	(40, 20)	0.714	0.708	0.691	0.792	0.786	0.746
(1, 0.5)	(10, 10)	0.612	0.499	0.493	–	–	–
	(10, 20)	0.580	0.500	0.463	0.690	0.500	0.486
	(10, 30)	0.573	0.500	0.459	0.683	0.500	0.448
	(20, 10)	0.538	0.500	0.453	0.630	0.501	0.496
	(20, 20)	0.523	0.500	0.467	0.615	0.500	0.462
	(20, 40)	0.486	0.500	0.461	0.571	0.500	0.467
	(30, 10)	0.512	0.500	0.446	0.592	0.500	0.433
	(40, 20)	0.470	0.500	0.458	0.534	0.500	0.472

表 3: 三つの判別ルールにおける risk の値 ( $\Delta = 1.68$ )

$\Delta = 1.68$		$p = 10$			$p = 20$		
$(c_1, c_2)$	$(N_1, N_2)$	W-rule	Z-rule	Min-rule	W-rule	Z-rule	Min-rule
(1, 1)	(10, 10)	0.676	0.663	0.634	–	–	–
	(10, 20)	0.603	0.592	0.580	0.766	0.756	0.704
	(10, 30)	0.573	0.561	0.558	0.708	0.690	0.666
	(20, 10)	0.592	0.586	0.578	0.758	0.744	0.697
	(20, 20)	0.544	0.541	0.541	0.681	0.674	0.667
	(20, 40)	0.499	0.496	0.496	0.589	0.583	0.582
	(30, 10)	0.569	0.554	0.551	0.716	0.691	0.663
	(40, 20)	0.496	0.494	0.494	0.595	0.586	0.586
(1, 0.5)	(10, 10)	0.490	0.487	0.547	–	–	–
	(10, 20)	0.439	0.500	0.433	0.588	0.494	0.568
	(10, 30)	0.419	0.500	0.404	0.550	0.500	0.481
	(20, 10)	0.409	0.499	0.435	0.539	0.496	0.575
	(20, 20)	0.380	0.500	0.393	0.493	0.500	0.478
	(20, 40)	0.346	0.500	0.350	0.428	0.500	0.409
	(30, 10)	0.383	0.500	0.403	0.482	0.500	0.472
	(40, 20)	0.337	0.500	0.349	0.405	0.500	0.410

## A. 付録

### A.1. 判別分析

#### A.1.1. 判別分析の基礎理論

判別分析とは、ある個体  $\mathbf{x}$  が  $K$  個の母集団  $\Pi_1, \dots, \Pi_K$  のうち、どれか 1 つに属する事は分かっているがどの母集団に属するかが分からないときに、その個体の特性値  $\mathbf{x} = (x_1, \dots, x_p)$  に基づいて、どの母集団に属するかをできるだけ高い成功率をもって判定する方法を見つけるという問題を扱っている。個体  $\mathbf{x}$  が  $\Pi_j$  ( $j = 1, \dots, K$ ) に所属するときの  $\mathbf{x}$  の確率分布 (測度) を  $P_j$  とするとき、 $P_1, \dots, P_K$  は互いに異なるものであるとする。

個体の判別方法は  $p$  次元空間  $\mathbf{R}^p$  を互いに素な  $K$  個の部分集合  $R_1, \dots, R_K$  に分割する分割の仕方によって決まる。  $\mathbf{R}^p$  の分割  $(R_1, \dots, R_K)$  が与えられたとき、個体  $\mathbf{x}$  は  $\mathbf{x} \in R_j$  ならば  $\mathbf{x} \in \Pi_j$  であると判別され

表 4: 係数  $(a_0(\Delta^2), b_0(\Delta^2))$  と Z-rule における  $a$  の値

$(c_1, c_2)$	$(N_1, N_2)$	Z-rule	$\Delta = 1.05$				$\Delta = 1.68$			
			$p = 10$		$p = 20$		$p = 10$		$p = 20$	
(1, 1)	(10, 10)	0.00	0.00	0.00	-	-	0.00	0.00	-	-
	(10, 20)	-0.02	-0.03	-0.05	-0.03	-0.18	-0.03	-0.09	-0.03	-0.27
	(10, 30)	-0.03	-0.04	-0.06	-0.04	-0.15	-0.04	-0.11	-0.04	-0.22
	(20, 10)	0.02	0.03	0.01	0.03	0.04	0.03	0.05	0.03	0.12
	(20, 20)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(20, 40)	-0.01	-0.01	-0.01	-0.01	-0.02	-0.01	-0.03	-0.01	-0.04
	(30, 10)	0.03	0.03	0.01	0.03	-0.02	0.04	0.05	0.03	0.04
	(40, 20)	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.01	0.02
(1, 0.5)	(10, 10)	0.00	-0.28	1.78	-	-	0.44	12.54	-	-
	(10, 20)	-0.02	-0.07	2.81	-0.80	-23.14	0.20	5.53	0.18	24.02
	(10, 30)	-0.03	-0.04	2.44	-0.26	0.21	0.13	3.87	0.14	11.88
	(20, 10)	0.02	-0.01	2.96	-0.77	-24.42	0.26	5.70	0.23	24.19
	(20, 20)	0.00	0.03	2.59	-0.14	1.77	0.17	3.79	0.18	11.54
	(20, 40)	-0.01	0.03	2.11	-0.04	2.74	0.12	2.75	0.09	5.34
	(30, 10)	0.03	0.04	2.67	-0.19	0.30	0.22	4.10	0.22	12.36
	(40, 20)	0.01	0.06	2.19	-0.01	2.90	0.15	2.82	0.12	5.50

る。以下、分割  $(R_1, \dots, R_K)$  とそれによって決まる判別方法  $r$  を同一視して  $r = (R_1, \dots, R_K)$  と表すことにする。個体  $\mathbf{x}$  が  $\Pi_k$  に属するのに、誤って  $\Pi_j$  に判別したときの損失を  $c_{j|k}$  と表す ( $c_{j|k} \geq 0$ ,  $c_{k|k} = 0$ )。個体が  $\Pi_k$  に属する事前確率 (先見確率) を  $\pi_k$  とするとき、判別方法  $r = (R_1, \dots, R_K)$  による判別の期待損失は、

$$L(r) = \sum_{k=1}^K \pi_k \sum_{j=1}^K c_{j|k} P_k(\mathbf{x} \in R_j) \quad (\text{A.1})$$

と表される。  $L(r)$  を最小にする判別方法  $r^0 = (R_1^0, \dots, R_K^0)$  はベイズルール (Bayes rule) と呼ばれている。  $f_j(\mathbf{x})$  を個体  $\mathbf{x}$  が  $\Pi_j$  に属するときの確率密度関数 (離散型の場合は確率関数) とすると、ベイズルールは、

$$\mathbf{x} \in R_i^0 \iff \sum_{k=1}^K c_{i|k} \pi_k f_k(\mathbf{x}) = \min \left\{ \sum_{k=1}^K c_{j|k} \pi_k f_k(\mathbf{x}) \mid j = 1, \dots, K \right\} \quad (\text{A.2})$$

によって与えられる。

事前確率  $\pi_j$  が未知の場合には期待損失  $L(r)$  は計算できない。この場合にはミニマックス判別法 (minimax discrimination rule) がよく用いられる。個体が  $\Pi_j$  に属するという条件の下での判別方法  $r = (R_1, \dots, R_K)$  による条件付期待損失は、

$$L_k(r) = \sum_{j=1}^K c_{j|k} P_k(\mathbf{x} \in R_j) \quad (k = 1, \dots, K) \quad (\text{A.3})$$

と表される。判別方法  $r = (R_1, \dots, R_K)$  と  $r^* = (R_1^*, \dots, R_K^*)$  に対して、

$$L_k(r^*) \leq L_k(r) \quad (k = 1, \dots, K) \quad (\text{A.4})$$

であり、少なくとも一つの不等号が「 $<$ 」であるとき、  $r^*$  は  $r$  より優れているという。ある判別方法  $r$  に対して  $r$  より優れた判別方法が存在しないとき、  $r$  は許容的であるという。この許容的判別法の集合の中から一つを選択する為の基準として、  $K$  個の条件付期待損失の最大値  $\max\{L_k(r) \mid k = 1, \dots, K\}$  があるが、この基準が最小である判別法をミニマックス判別法と呼ぶ。  $\pi_j > 0$  ( $j = 1, \dots, K$ ) に対するべ

イズルールは許容的であり、ベイズルールの中で  $L_1(r) = \dots = L_K(r)$  を満たすものはミニマックスである事が知られている。

又、多母集団の場合には、個体  $\mathbf{x} = (x_1, \dots, x_p)$  が  $K$  個の母集団  $\Pi_1, \dots, \Pi_K$  において観測されているとき、その母集団間の差異をできるだけ少数個の正準判別変量と呼ばれる  $\mathbf{x}$  の 1 次関数の組で説明する事を目的とする正準判別分析と呼ばれる方法もある。

### A.1.2. 2つの正規母集団に対する判別

$k = 2$  の時、期待損失 (A.1) は

$$L(r) = \pi_1 c_{2|1} P(2|1) + \pi_2 c_{1|2} P(1|2) \quad (\text{A.5})$$

という形で表される。  $f_i(\mathbf{x})$  を、個体  $\mathbf{x}$  が母集団  $\Pi_i$  に属するときの確率密度関数とすると、(A.5) 式を最小にする方法 (ベイズルール) は

$$f_1(\mathbf{x})/f_2(\mathbf{x}) > c \Leftrightarrow \text{assign } \mathbf{x} \text{ into } \Pi_1 \quad (\text{A.6})$$

と表される。但し、 $c$  は誤判別による損失との比率および事前確率によって決まる定数 ( $c = c_2 \pi_2 / c_1 \pi_1$ ) であるとする。母集団  $\Pi_i$  ( $i = 1, 2$ ) が平均ベクトル  $\mu_i$ 、分散共分散行列  $\Sigma_i$  の  $p$  次元正規分布  $N_p(\mu_i, \Sigma_i)$  であるとき、(A.6) の不等式は、

$$Q(\mathbf{x}; \mu_1, \mu_2, \Sigma_1, \Sigma_2) < -2 \log c \quad (\text{A.7})$$

と同等である。但し、

$$Q(\mathbf{x}; \mu_1, \mu_2, \Sigma_1, \Sigma_2) = (\mathbf{x} - \mu_1)' \Sigma_1^{-1} (\mathbf{x} - \mu_1) - (\mathbf{x} - \mu_2)' \Sigma_2^{-1} (\mathbf{x} - \mu_2) + \log |\Sigma_1 \Sigma_2^{-1}| \quad (\text{A.8})$$

であるとする。(A.8) の判別関数は 2 次判別関数 (quadratic discriminant function) と呼ばれている。更に、 $\Sigma_1 = \Sigma_2 (= \Sigma)$  である場合には、(A.6) は

$$L(\mathbf{x}; \mu_1, \mu_2, \Sigma) > \log c \quad (\text{A.9})$$

と同等なものになる。但し、

$$L(\mathbf{x}; \mu_1, \mu_2, \Sigma) = \mathbf{x}' \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \quad (\text{A.10})$$

であるとする。(A.10) の判別関数は線形判別関数 (linear discriminant function) と呼ばれている。

平均ベクトル  $\mu_i$  と分散共分散行列  $\Sigma_i$  が既知の場合はこれでよいが、大抵の一般的なデータではこれらは未知である。そこで、各母集団から得られた標本 (初期標本) に基づく推定量  $\hat{\mu}_1, \hat{\mu}_2$  及び  $\hat{\Sigma}_1, \hat{\Sigma}_2$  を用いた判別関数

$$Q(\mathbf{x}; \hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2) \quad (\text{A.11})$$

あるいは、 $\Sigma_1 = \Sigma_2 (= \Sigma)$  が予想される場合には  $\Sigma$  の推定量  $\hat{\Sigma}$  を用いた

$$L(\mathbf{x}; \hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}) \quad (\text{A.12})$$

を最適な判別関数の代用とする。これらの判別関数による判別方法はプラグインルール (plug-in rule) と呼ばれている。母集団  $\Pi_i (i = 1, 2)$  からの大きさ  $N_i$  の初期標本に基づいて、標本平均ベクトル  $\bar{\mathbf{x}}_i$  と標本分散共分散行列  $S_i$  を

$$\bar{\mathbf{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_{ij}, \quad S_i = \frac{1}{n_i} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \quad (n_i = N_i - 1),$$

のように定義する。未知母数の推定量としてこれらを用いると、(A.11) は

$$Q(\mathbf{x}; \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S_1, S_2) = (\mathbf{x} - \bar{\mathbf{x}}_1)' S_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) - (\mathbf{x} - \bar{\mathbf{x}}_2)' S_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) + \log |S_1 S_2^{-1}| \quad (\text{A.13})$$

と表される。これは標本2次判別関数 (sample quadratic discriminant function) と呼ばれている。又、 $\bar{\mathbf{x}}_i$  と、共通の分散共分散行列  $\Sigma$  の不偏推定量

$$S = n^{-1} (n_1 S_1 + n_2 S_2) \quad (n = n_1 + n_2)$$

を用いると、(A.12) は

$$L(\mathbf{x}; \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S) = \mathbf{x}' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{1}{2} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (\text{A.14})$$

と表され、W-統計量 (W-rule) あるいは標本線形判別関数 (sample linear discriminant function) と呼ばれている。

しかしながら、上の議論により導出されたプラグインルールは、母集団パラメーターが既知であるとして導出されたルールにその推定値を代入することにより推定を考えている為、ベイズルールの推定としては最適とはいえない。そこで、ベイズルールの別の推定方法を考えることにする。その方法として、仮説  $H: \mathbf{x} \in \Pi_1$  を対立仮説  $K: \mathbf{x} \in \Pi_2$  に対して検定を行うときの尤度比に基づいて判別関数を作成する方法が挙げられる。 $\Sigma_1 = \Sigma_2 (= \Sigma)$  と仮定したとき、この方法は、

$$d_i = \frac{N_i}{N_i + 1} (\mathbf{x} - \bar{\mathbf{x}}_i)' S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) \quad (i = 1, 2)$$

として

$$\frac{n + d_2}{n + d_1} \geq c \Rightarrow \mathbf{x} \in \Pi_1, \quad \frac{n + d_2}{n + d_1} < c \Rightarrow \mathbf{x} \in \Pi_2 \quad (\text{A.15})$$

という形で表される。但し、 $n = N_1 + N_2 - 2$  であるとする。このような判別法は尤度比法 (likelihood rule) と呼ばれている。特に、 $c = 1$  のときの判別法

$$Z = d_1 - d_2 \leq 0 \Rightarrow \mathbf{x} \in \Pi_1, \quad Z > 0 \Rightarrow \mathbf{x} \in \Pi_2 \quad (\text{A.16})$$

は最尤法 (maximum likelihood rule) 或いは Z-統計量 (Z-rule) と呼ばれている. この  $Z$  を具体的に表すと

$$\frac{N_1}{N_1+1}(\mathbf{x} - \bar{\mathbf{x}}_1)'S^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1) - \frac{N_2}{N_2+1}(\mathbf{x} - \bar{\mathbf{x}}_2)'S^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2), \quad (\text{A.17})$$

あるいは  $S$  の代わりに  $A = nS$  を用いて,

$$\frac{N_1}{N_1+1}(\mathbf{x} - \bar{\mathbf{x}}_1)'A^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1) - \frac{N_2}{N_2+1}(\mathbf{x} - \bar{\mathbf{x}}_2)'A^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2) \quad (\text{A.18})$$

という形になる.  $N_1 = N_2$  のとき, (A.17) は (A.14) の定数倍となり, この2つの判別法は同値となる. なお,  $\Sigma_1 \neq \Sigma_2$  のときの尤度比法は,

$$\begin{aligned} \lambda = & \left( \frac{N_1}{N_1+1} \right)^{pN_1/2} \left( \frac{N_2}{N_2+1} \right)^{-pN_1/2} \left( \frac{N_2+1}{N_1+1} \right)^{p/2} \left( \frac{|S_2|}{|S_1|} \right)^{1/2} \\ & \times \frac{\{1 + (N_2/(N_2+1))(\mathbf{x} - \bar{\mathbf{x}}_2)'S_2^{-1}(\mathbf{x} - \bar{\mathbf{x}}_2)\}^{(N_2+1)/2}}{\{1 + (N_1/(N_1+1))(\mathbf{x} - \bar{\mathbf{x}}_1)'S_1^{-1}(\mathbf{x} - \bar{\mathbf{x}}_1)\}^{(N_1+1)/2}} \end{aligned}$$

と置いたとき

$$\lambda \geq c \Rightarrow \mathbf{x} \in \Pi_1, \quad \lambda < c \Rightarrow \mathbf{x} \in \Pi_2 \quad (\text{A.19})$$

と表される. この条件での最尤法は (A.19) で  $c = 1$  のときである.

## A.2 漸近展開

### A.2.1 漸近展開の基礎理論

大きさ  $n$  の標本  $X_1, \dots, X_n$  が有るとき, 推定や検定などの統計的推測はこれらの標本から計算される統計量  $T_n = T(X_1, \dots, X_n)$  に基づいて行われる. この時, 推定における推定誤差の評価や検定における有意点の計算等, 推定方式を評価する為には  $T_n$  の (標本) 分布を知る必要がある. しかしながら, 統計量  $T_n$  自体が複雑な場合には,  $T_n$  の分布を性格に求めることが困難であることも少なくない. ところが, このような場合でも, 標本数  $n$  を大きくしていくと, 統計量の分布が正規分布や  $\chi^2$ -分布などといった標準的な分布に近づいていく場合がしばしばある. この場合, 標本数  $n$  が十分大きいときには, 統計量  $T_n$  の分布を  $n \rightarrow \infty$  とした極限の分布で近似することができる. これが大標本論の基礎であるが, 大標本論で与えられる第1次近似, すなわち極限の分布のみでは近似が満足いかないものであることも多い. そこで, 近似を良くする為に更に高次元の近似を求めようとするのが漸近展開の考え方である.

漸近展開において注意すべきことは, 漸近展開は理論的には標本数  $n$  が十分大きなときに収束を改善する手法であり,  $n$  が小さいところでは漸近展開によって近似が改良されるとは限らないという事である. しかしながら経験的に, 漸近展開はさまざまな状況下で有効であることが分かっている.

ある数列  $\{b_{kn}\}(k=0,1,\dots)$  を,  $n \rightarrow \infty$  のときには  $b_{kn} = o(b_{k-1,n})$  となり,  $b_{0n} = 1$  であるような列の基底集合であるとする. 例として,

$$b_{0n} = 1, b_{1n} = \frac{1}{\sqrt{n}}, b_{2n} = \frac{1}{n}, \dots, b_{kn} = \frac{1}{n^{k/2}}, \dots$$

$$b_{0n} = 1, b_{1n} = \frac{1}{n}, b_{2n} = \frac{1}{n^2}, \dots, b_{kn} = \frac{1}{n^k}, \dots$$

などが挙げられる. 任意の列  $\{g_n\}$  に対し  $n \rightarrow \infty$  としたとき, ある自然数  $h$  を固定し, あるふさわしい定数  $\gamma_1, \gamma_2, \dots, \gamma_h$  について,

$$g_n = g_n^{(0)} \{1 + \gamma_1 b_{1n} + \dots + \gamma_h b_{hn} + o(b_{hn})\} \quad (\text{A.20})$$

特に,

$$g_n = g_n^{(0)} \{1 + o(b_{0n})\} \quad (\text{A.21})$$

とできる列  $\{g_n^{(0)}\}$  が存在すると仮定する. このとき, (A.20) を  $n \rightarrow \infty$  としたときの先導項 (leading term)  $g_n^{(0)}$  を持つ  $g_n$  の漸近展開 (asymptotic expansion) という. 又, 任意の関数  $g(z)$  についても, 上で  $g_n^{(0)}$  を  $g^{(0)}(z)$ ,  $b_{kn}$  を  $b_k(z)$  と置き換えれば似たような定義が適用でき,

$$g(z) = g^{(0)}(z) \{1 + \gamma_1 b_1(z) + \dots + \gamma_h b_h(z) + o(b_h(z))\} \quad (\text{A.22})$$

と表す事ができる. これを先導項  $g^{(0)}(z)$  を持つ  $g(z)$  の漸近展開という.

最近では, 近似を考える際に, Bootstrap 法などといった計算機を用いてノンパラメトリック法の枠組みの中で近似を計算するといった方法もとられているが, 漸近展開の利点には以下のことが挙げられる. まず, 漸近展開は式の形で表されているので, 一度計算しておくと同様に応用が利く点が挙げられる. 又, 式を見て傾向等が把握できる点も魅力的であろう.

なお, 漸近展開の研究には, 次の3つの視点がある. 一つは, 漸近展開をいかにして求めるかである. 漸近展開を求める為にこれまで色々な方法が考えられている. どのような方法があるかについては後に述べる. 二つめは, 誤差項  $o(b_h(z))$  が本当にこれくらい小さくなるかという問題である. これは誤差項の評価 (validity) といわれている. 後に述べる Edgeworth 展開については, Bhattacharya and Ghosh (1978) にて validity が成り立つ条件について述べられている. 三つめは, その誤差項をもっと精密に評価する問題である. これは Ulyanov and Fujikoshi (1999) などによって研究されている.

### A.2.1 漸近展開の方法

漸近展開の展開方法には次のようなものがある.

- 部分積分を用いた方法.

標準正規分布を例にとって説明する. 標準正規分布の分布関数  $\Phi(z)$  について,

$$\Psi(z) = 1 - \Phi(z) = \int_z^\infty \phi(y) dy$$

とする. すると,  $\Psi(z)$  は  $z$  が十分大きいとき

$$\Psi(z) = \frac{\phi(z)}{z} - \int_z^\infty \frac{\phi(y)}{y^2} dy$$

と式変形できる. これを繰り返すと,

$$\Psi(z) = \frac{\phi(z)}{z} \left( 1 - \frac{1}{z^2} + \frac{1 \cdot 3}{z^4} - \frac{1 \cdot 3 \cdot 5}{z^6} + \dots \right) \quad (\text{A.23})$$

という式を得る. この式を  $z^{-2k}$  の項で止めたとき, 剰余項は  $O(z^{-2k-2})$  となる事が簡単に示される.

- Laplace 展開を用いた方法.

関数  $w(z)$  が,

$$w(z) = \int_a^b e^{-zr(y)} g(y) dy$$

という形で表されているとき,  $r(y)$  の最小値をとる  $y = \tilde{y}$  が  $(a, b)$  内にあり, かつ  $r'(\tilde{y}) = 0$ ,  $r''(\tilde{y}) > 0$ ,  $g(\tilde{y}) \neq 0$  ならば,  $w(z)$  は

$$w(z) = \frac{e^{-z\tilde{r}} \tilde{g} \sqrt{2\pi}}{\sqrt{z} \sqrt{\tilde{r}''}} \{1 + O(z^{-1})\} \quad (\text{A.24})$$

という形に展開できる. さらに次の項まで展開すると,

$$w(z) = \frac{e^{-z\tilde{r}} \sqrt{2\pi}}{\sqrt{z} \sqrt{\tilde{r}''}} \left\{ \tilde{g} + \frac{1}{z} \left( \frac{\tilde{g}''}{2\tilde{r}''} - \frac{\tilde{r}^{(3)} \tilde{g}}{2(\tilde{r}'')^2} - \frac{\tilde{r}^{(4)} \tilde{g}}{8(\tilde{r}'')^2} + \frac{5(\tilde{r}^{(3)})^2 \tilde{g}}{24(\tilde{r}'')^3} \right) + O(z^{-2}) \right\} \quad (\text{A.25})$$

という形になる. 但し,

$$\tilde{r} = r(\tilde{y}), \quad \tilde{g} = g(\tilde{y}), \quad \tilde{r}'' = r''(\tilde{y}), \quad \text{etc.}$$

であるとする. この展開は Stirling の定理の導出に使われている.

- 級数の総和を用いた方法.

$S_n$  と  $I(n)$  をそれぞれ

$$S_n = \frac{1}{2}f(a) + f(a+1) + \dots + f(n-1) + \frac{1}{2}f(n)$$

$$I(n) = \int_a^n f(x) dx$$

と定義する. 但し,  $a$  と  $n$  は  $a < n$  となる整数で,  $f(x)$  は  $x$  の  $C^\infty$  級の関数であるとする. このとき,  $S_n - I(n)$  が,

$$\frac{f'(n) - f'(a)}{12} - \frac{f^{(3)}(n) - f^{(3)}(a)}{720} + R_4$$

だけ異なる事を用いて展開を考える. 但し,  $R_4$  は  $f^{(4)}(x)$  に依存している積分であるとする. これは  $n \rightarrow \infty$  としたときの漸近展開の形ではないのだが, 利用方法を工夫すれば漸近展開の一方とすることができる.

- 級数の反転を用いた方法.

$g_n(x)$  を,  $n \rightarrow \infty$  としたときに先導項が  $g^{(0)}(x)$  であり, かつ  $1/\sqrt{n}$  の項の展開式を持つものとし, そしてこのとき, 方程式  $g^{(0)}(x) = a$  の解が知られているものとする. 方程式

$$g_n(x) = a$$

を解く為に,  $x$  を

$$x = x^{(0)} + \frac{x^{(1)}}{\sqrt{n}} + \frac{x^{(2)}}{n} + \dots$$

と置き  $g_n(x)$  の漸近展開に代入すれば,  $x^{(0)}, x^{(1)}, x^{(2)}, \dots$  が求まり,  $x$  の近似式ができる. これは Conish-Fisher 展開のように分布の分点を計算するときによく使われる.

- 直接に, あるいは  $\log$  をとってから漸近展開をする方法.
- (もしできるなら) モーメント母関数や特性関数の展開を得てから項別積分により反転する方法.
- 確率変数による漸近展開を用いた導出法.

確率変数の列  $\{Y_n\}$  を, 連続な確率変数で

$$Y_n = X_0 + X_1 b_{1n} + \dots + X_h b_{hn} + O_p(b_{h+1,n})$$

とできるものであると定義する. 但し,  $\{X_0, X_1, \dots\}$  は  $n$  に従属しない分布を持つ確率変数であるとする. このとき,  $Y_n$  の密度関数は  $X_0, X_1, \dots, X_h$  の同時分布から積分と Taylor 展開などを用いて展開する事ができる. 例えば,  $h = 2$  のときは,

$$f(Y_n; y) = f_0(y) - \frac{\partial f_0(y) \mu_{1 \cdot 0}(y)}{\partial y} \frac{1}{\sqrt{n}} + \frac{1}{n} \left( \frac{\partial^2 f_0(y) \mu_{2 \cdot 0}(y)}{\partial y^2} \cdot \frac{1}{2} - \frac{\partial f_0(y) \mu_{01 \cdot 0}(y)}{\partial y} \right) + O(n^{-3/2}) \quad (\text{A.26})$$

と表される. 但し,  $f_0(x_0)$  は  $X_0$  の密度関数,  $f_{01}(x_0, x_1)$  は  $X_0, X_1$  の同時密度関数,  $f_{012}(x_0, x_1, x_2)$  は  $X_0, X_1, X_2$  の同時密度関数で,  $\mu_{1 \cdot 0}, \mu_{01 \cdot 0}, \mu_{2 \cdot 0}$  はそれぞれ,

$$\begin{aligned} \mu_{1 \cdot 0} &= E(X_1 | X_0 = y) = \int_{-\infty}^{\infty} x_1 \frac{f_{01}(y, x_1)}{f_0(y)} dx_1, \\ \mu_{01 \cdot 0} &= E(X_2 | X_0 = y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 \frac{f_{012}(y, x_1, x_2)}{f_0(y)} dx_1 dx_2, \\ \mu_{2 \cdot 0} &= E(X_1^2 | X_0 = y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 \frac{f_{012}(y, x_1, x_2)}{f_0(y)} dx_1 dx_2 \end{aligned}$$

を表しているものとする.

- Edgeworth 展開による方法.

$Y, Y_1, \dots, Y_n$  を平均が  $\mu$ , 分散が  $\sigma^2$  である独立同一分布に従う確率変数で,  $r$  次キュミュラン

ト  $\kappa_r$  を持っているものと仮定する. 又, 標準化されたキュムラントを  $\rho_r := \kappa_r / \sigma^r$  と表す事に  
 する. この確率変数の和  $Y_1 + \dots + Y_n$  を  $S_n$  と定義し, その標準化を

$$S_n^* := \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad (\text{A.27})$$

と表すと,  $S_n^*$  のキュムラント母関数  $K(S_n^*; t)$  は,

$$\begin{aligned} K(S_n^*; t) &= -\sqrt{n}\mu t / \sigma + nK(Y; t / \sigma\sqrt{n}) \\ &= \frac{1}{2}t^2 + \frac{\rho_3 t^3}{6\sqrt{n}} + \frac{\rho_4 t^4}{24n} + O(n^{-3/2}) \end{aligned} \quad (\text{A.28})$$

と表される.  $S_n^*$  のモーメント母関数は  $M(S_n^*; t) = \exp\{K(S_n^*; t)\}$  となるので,

$$M(S_n^*; t) = \exp\left(\frac{1}{2}t^2\right)\left\{1 + \frac{\rho_3 t^3}{6\sqrt{n}} + \frac{\rho_4 t^4}{24n} + \frac{\rho_3^2 t^6}{72n} + O(n^{-3/2})\right\} \quad (\text{A.29})$$

と表される. ここで,  $Y$  の特性関数  $\varphi(t)$  が Cramér 条件

$$\limsup_{|t| \rightarrow \infty} |\varphi(t)| < 1$$

を満たすと仮定する. この仮定により, (A.29) のモーメント母関数は validity を損なわずに反転  
 を行う事ができる. よって,  $S_n^*$  の確率密度関数について以下の展開を得る.

$$f(S_n^*; x) = \phi(x)\left\{1 + \frac{\rho_3 h_3(x)}{6\sqrt{n}} + \frac{\rho_4 h_4(x)}{24n} + \frac{\rho_3^2 h_6(x)}{72n}\right\} + O(n^{-3/2}) \quad (\text{A.30})$$

但し,  $\phi(x)$  は標準正規分布の確率密度関数であり,  $h_r(x)$  は  $r$  次の Hermite 多項式であるとする.  
 このような展開を独立和の Edgeworth 展開という. なお, 同じ仮定の下での分布関数についての  
 展開は

$$F(S_n^*; x) = \Phi(x) - \phi(x)\left\{\frac{\rho_3 h_2(x)}{6\sqrt{n}} + \frac{\rho_4 h_3(x)}{24n} + \frac{\rho_3^2 h_5(x)}{72n}\right\} + O(n^{-3/2}) \quad (\text{A.31})$$

と表される.

- Saddlepoint Edgeworth 展開 (tilted Edgeworth 展開) による方法.

Edgeworth 展開では,  $\rho_3 = 0$  を仮定したとき

$$f(S_n^*; x) = \phi(x)\{1 + O(n^{-1/2})\}$$

となり, 標準正規分布との誤差は  $O(n^{-1/2})$  となる. しかし,  $x = 0$  の近傍で考えたとき,  $h_3(x) =$   
 $x^3 - 3x$  であるので  $h_3(0) = 0$  となり, 標準正規分布との誤差は  $O(n^{-1})$  となる. そこで, 誤差を  
 良くする為に,  $Y, Y_1, \dots, Y_n$  をそれぞれ独立に確率密度関数  $f_0(y)$  に従う確率変数とし, その  
 平均が  $\mu$ , 分散が  $\sigma^2$  であるとする. 又,  $K(t)$  と  $M(t)$  をそれぞれ  $Y$  のキュムラント母関数と  
 モーメント母関数であるとする.  $f_0(y)$  を例えば指数分布族

$$f(y; \lambda) = \frac{e^{\lambda y} f_0(y)}{M(\lambda)} = \exp\{\lambda y - K(\lambda)\} f_0(y)$$

と関わらせて考える. このとき,  $f(y;0) = f_0(y)$  となっている.  $Y'_1, \dots, Y'_n$  を独立に確率密度関数  $f(y;\lambda)$  に従う確率変数とし, その和を  $S_{n\lambda}$  で表す. ここで,  $S_{n0} = S_n$  は  $Y_1, \dots, Y_n$  の和を表している.  $s = y_1 + \dots + y_n$  と置く.  $g(s;\lambda)$  を  $S_{n\lambda}$  の確率密度関数とすると,  $Y_1, \dots, Y_n$  の独立性によりこれは  $S_n$  の確率密度関数  $f_{S_n}(s)$  を用いて表す事ができ,

$$g(s;\lambda) = \exp\{\lambda s - nK(\lambda)\}f_{S_n}(s)$$

となる. これを,

$$f_{S_n}(s) = \exp\{-\lambda s + nK(\lambda)\}g(s;\lambda)$$

と変形し,  $f_{S_n}(s)$  の展開を  $g(s;\lambda)$  の展開から求める.  $g(s;\lambda)$  について  $S_n$  を標準化すると,  $s$  と Edgeworth 展開 (A.30) での  $x$  との関係は,

$$\frac{s - E(S_n;\lambda)}{\sqrt{V(S_n;\lambda)}} = x$$

と表される. ここで,  $s$  が  $\lambda$  についての分布  $g(s;\lambda)$  の『中心』にあるように, すなわち,

$$E(S_n;\lambda) = s$$

を満たすように  $\lambda = \hat{\lambda}$  と取る. これは  $K'(\hat{\lambda}) = s/n$  によって定義される. こうすると,  $g(s;\lambda)$  の Edgeworth 展開は (A.30) で  $x = 0$  としたときと同じものになるので, 平均で評価された Edgeworth 級数から,  $g(s;\lambda)$  は

$$g(s;\lambda) = \{2\pi nK''(\hat{\lambda})\}^{-1/2}\{1 + O(n^{-1})\}$$

と表される. よって,  $f_{S_n}$  は

$$f_{S_n} = \frac{\exp\{-\hat{\lambda} + nK(\hat{\lambda})\}}{\{2\pi nK''(\hat{\lambda})\}^{1/2}}\{1 + O(n^{-1})\} \quad (\text{A.32})$$

と表される. このような展開を Saddlepoint Edgeworth 展開 (tilted Edgeworth 展開) という.

- Cornish-Fisher 展開

統計量の分布の漸近展開を利用して, その統計量のパーセント点を求めることを考える. 例えば, 統計量  $T_n$  の分布関数  $F_n(x)$  が  $n \rightarrow \infty$  で標準正規分布  $N(0,1)$  に従うとする. すなわち, 全ての  $x$  に対して

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$$

であるとする. このとき, 分布関数  $F$ ,  $\Phi$  の上側  $\alpha$  点をそれぞれ  $x$ ,  $u$  とすると

$$F_n(x) = \Phi(u) = 1 - \alpha$$

となる. この方程式を満たす解  $x = x(u)$  が求まれば,  $T_n$  の上側  $\alpha$  点が  $\Phi$  の上側  $\alpha$  点  $u$  を用いて求められたことになる.

統計量  $T_n$  の特性関数  $C_n(t)$  が

$$C_n(t) = e^{(it)^2/2} \left[ 1 + \frac{1}{\sqrt{n}} \{a_1(it) + a_3(it)^3\} + \frac{1}{n} \{b_2(it)^2 + b_4(it)^4 + b_6(it)^6\} + o(n^{-1}) \right]$$

と展開されるとする。このとき、適当な正則条件の下で

$$\begin{aligned} P(T_n \leq x) &= F_n(x) \\ &= \Phi(x) + \phi(x) \left\{ \frac{1}{\sqrt{n}} p_1(x) + \frac{1}{n} p_2(x) \right\} + o(n^{-1}) \end{aligned}$$

と表すことができる。但し、

$$\begin{aligned} p_1(x) &= -\{a_1 + a_3 h_2(x)\}, \\ p_2(x) &= -\{b_2 h_1(x) + b_4 h_3(x) + b_6 h_5(x)\} \end{aligned}$$

であり、 $h_r$  は  $r$  次の Hermite 多項式であるとする。この展開式に

$$x = x(u) = u + \frac{1}{\sqrt{n}} q_1(u) + \frac{1}{n} q_2(u) + o(n^{-1})$$

を代入して、 $F_n(x(u)) = \Phi(u)$  となる  $q_1(u)$ ,  $q_2(u)$  を決定することができ、

$$\begin{aligned} q_1(u) &= -p_1(u), \\ q_2(u) &= -p_2(u) - \frac{1}{2} u p_1(u) + p_1'(u) p_1(u) \end{aligned}$$

となる。

このように漸近展開には色々な方法が存在する。

## 参考文献

- [1] Anderson, T. W. (1973a). An asymptotic expansion of the distribution of the studentized classification statistic, *Ann. Statist.* **1**, 964-972.
- [2] Anderson, T. W. (1973b). Asymptotic evaluation of the probabilities of misclassification by linear discriminant function, *Discriminant Analysis and Applications* (T. Cacoullos, ed.), Academic Press, New York, 17-35.
- [3] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition. New York, Wiley.
- [4] Bai, Z. D. and Zhao, L. C. (1986). Edgeworth expansions of distribution functions of independent random variables, *Sci. China Ser. A* **29**, 1-22.
- [5] Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall.
- [6] Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansions, *J. Multivariate Anal.* **27**, 68-79.
- [7] Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansion*, John Wiley and Sons, New York.
- [8] Deev, A. D. (1970). Representation of statistics of discriminant analysis and asymptotic expansions when space dimensions are comparable with sample size, *Soviet Math. Dokl.* **11**, 1547-1550.
- [9] Fujikoshi, Y. and Kanazawa, M. (1976). The ML classification statistic in covariate discriminant analysis and its asymptotic expansions, *Essays in Prob. Statist.* (S. Ikeda and Others, eds.), Shinko-Tsusho, Tokyo, 305-320.

- [10] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant functions when the sample sizes and the dimension are large, *Random Oper. Stochastic Equations* **6**, 269-280.
- [11] Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type, *Ann. Math. Statist.* **39**, 1264-1273.
- [12] Kanazawa, M. (1979). The asymptotic cut-off point and comparison of error probabilities in covariate discriminant analysis, *J. Japan. Statist. Soc.* **9**, 7-17.
- [13] Kanazawa, M. and Fujikoshi, Y. (1977). The distribution of the studentized classification statistic  $W^*$  in covariate discriminant analysis, *J. Japan. Statist. Soc.* **7**, 81-88.
- [14] Matsumoto, C. (2004). An optimal discriminant rule in the class of linear and quadratic discriminant functions for large dimension and samples, *Hiroshima Mathematical Journal* **34**, 231-250.
- [15] Matsumoto, C. and Wakaki, H. (2003). Asymptotic expansion of the quadratic discriminant function when the dimension and sample sizes are large, *TR No. 03-14, Statistical Research Group, Hiroshima University*.
- [16] McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley and Sons, Inc., New York.
- [17] Memon, A. Z. and Okamoto, M. (1971). Asymptotic expansion of the distribution of the Z statistics in discriminant analysis, *J. Multivariate Anal.* **1**, 294-307.
- [18] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, Inc., New York.
- [19] Okamoto, M. (1963, 1968). An asymptotic expansion for the distribution of the linear discriminant function, *Ann. Math. Statist.* **34**, 1286-1301. Correction, *Ann. Math. Statist.* **39**, 1358-1359.
- [20] Page, J. T. (1985). Error-rate estimation in discriminant analysis, *Technometrics* **27**, 189-198.
- [21] Raudys, S. (1972). On the amount of priori information in designing the classification algorithm, *Tech. Cybern* **4**, 168-174 (in Russian).
- [22] Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of D- and A-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices, *J. Multivariate Anal.* **46**, 154-174.
- [23] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Sciences Press, INC.
- [24] Siotani, M. and Wang, R. H. (1977). Asymptotic expansions for error rates and comparison of the W-procedure and the Z-procedure in discriminant analysis, *In Multivariate Analysis (Vol.IV)*, P. R. Krishnaiah (Ed.), Amsterdam : North-Holland, 523-545.
- [25] Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North-Holland, New York.
- [26] Tonda, T. and Wakaki, H. (2003). EPMC estimation in discriminant analysis when the dimension and sample sizes are large, *TR No. 03-08, Statistical Research Group, Hiroshima University*.
- [27] Ulyanov, V. V. and Fujikoshi, Y. (1999). On uniform error estimates of improved  $\chi^2$ -approximation, *TR No. 99-07, Statistical Research Group, Hiroshima University*.
- [28] Wakaki, H. (1990). Comparison of linear and quadratic discriminant functions, *Biometrika* **77**, 227-229.
- [29] Wakaki, H. and Aoshima, M. A class of Linear Discriminant function, (to appear).
- [30] Wyman, F. J., Young, D. M. and Turner, D. W. (1990). A comparison of asymptotic error rate expectations for the sample linear discriminant function, *Pattern Recognition* **23**, 775-783.