HIROSHIMA MATH. J.
34 (2004), 231–250

An optimal discriminant rule in the class of linear and quadratic discriminant functions for large dimension and samples

Chieko MATSUMOTO

(Received January 7, 2004)

ABSTRACT. For the classification problem between two normal populations with a common covariance matrix, we consider a class of discriminant rules based on a general discriminant function T. The class includes the one based on Fisher's linear discriminant function and the likelihood ratio rule. Our main purpose is to derive an optimal discriminant rule by using an asymptotic expansion of misclassification probability when both the dimension and the sample sizes are large. We also derive an asymptotically unbiased estimator of the misclassification probability of T in our class.

1. Introduction

Consider the problem of classifying an observation vector $\boldsymbol{x}$ into one of two normal populations $\Pi_1 : N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\Pi_2 : N_p(\boldsymbol{\mu}_2, \Sigma)$, where the mean vector $\boldsymbol{\mu}_i$ and the common covariance matrix Σ are unknown. Suppose that we have a training sample of $\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{1N_1}$ from Π_1 and another independent training sample of $\boldsymbol{x}_{21}, \dots, \boldsymbol{x}_{2N_2}$ from Π_2. Let $\bar{\boldsymbol{x}}_i$ and S be the sample mean and the pooled sample covariance matrix given by

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

respectively, where $n = N_1 + N_2 - 2$.

We define a general discriminant function by

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

which leads to the discriminant rule: classify $\boldsymbol{x}$ to Π_1 if $T(\boldsymbol{x}) < 0$ and otherwise to Π_2, where a and b are constants. Note that the discriminant function T

2000 *Mathematics Subject Classification*. Primary 62H30, secondly 62H12.

Key words and phrases. Asymptotic expansion, Asymptotic unbiased estimator, Discriminant function, Misclassification probability, Optimal discriminant rule.

induces a class of discriminant rules by considering various a and b, where $a = O(m^{-1})$ and $b = O(1)$ when $m \to \infty$. Here, m may be the dimension p or the sample size N_i. In this paper we use the notation T as both a general discriminant function and a class of discriminant rules. The sample linear discriminant rule (called the W-rule) and the maximum likelihood ratio rule (called the Z-rule) are included in this class. In fact, when $a = 0$ and $b = -\log c$, T becomes

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

and we have the W-rule, where c is the ratio of error costs of misclassifications and the prior probabilities. When

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

and $b = 0$, T becomes

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

and we have the Z-rule (see Anderson (1984)). In this paper, we are interested in finding an optimal one in the class of discriminant rules.

Let c_i be the product of the prior probability and the cost of misclassification for $\boldsymbol{x}$ coming from Π_i. Then, the risk of a classification rule is defined as the expected cost of misclassification

$$c_1 P(2|1) + c_2 P(1|2). \tag{1.2}$$

Generally, if (1.2) is smaller, we consider the classification rule is better. Note that the constant c in the W-rule is given by $c = c_2/c_1$.

In order to make (1.2) smaller, we use some approximated misclassification probability since the exact distribution function of T is too complicated to handle. Fujikoshi (1987) considered to select the variables minimizing a risk of classification for the W-rule and the Z-rule when $c_1 = q$ and $c_2 = 1 - q$ and derived an asymptotically unbiased estimator of the risk by using an asymptotic expansion of misclassification probability when only the sample sizes are large. Wakaki and Aoshima (2004) derived an asymptotic expansion for the cut-off point which satisfies some conditions on misclassification probability and derived an optimal rule in the class (1.1) of discriminant rules, using the asymptotic approximated distributions of discriminant functions when only the sample sizes are large. They noted that the Z-rule is optimal when $c_1 = c_2$. Our purpose of this paper is to derive an optimal discriminant rule by finding out the coefficients a and b which minimize (1.2) in the class given by

(1.1), using an asymptotic approximation of misclassification probability when both the dimension and the sample sizes are large.

In section 2 we give an asymptotic approximation of misclassification probability for large dimension and sample sizes. In section 3 we derive the coefficients a and b which minimize (1.2) when both the dimension and the sample sizes are large. In section 4, some numerical experiments are carried out to examine the performance of the derived classification rule.

2. Asymptotic approximation of misclassification probability

2.1. Introduction

Since the exact distributions of discriminant functions are too complicated to handle, we often use asymptotic approximations instead of the exact distributions. In this section we give new results as well as the previous results for asymptotic approximations of misclassification probability.

When the sample sizes are large, Okamoto (1963, 1968) derived an asymptotic expansion formula for the W-rule up to $O_2(n)$, where $O_j(n)$ means the j-th order terms with respect to $(N_1^{-1}, N_2^{-1}, n^{-1})$. Its asymptotic expansion formula of misclassification probability is given as

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

where

$$a_1 = (2\Delta^2)^{-1}(d_0^{(4)} + 3pd_0^{(2)}), \qquad a_2 = (2\Delta^2)^{-1}\{d_0^{(4)} - (p-4)d_0^{(2)}\},$$

$$a_3 = \frac{1}{2}(p-1)d_0^{(2)}, \qquad d_0^{(i)} := \frac{d^i}{dc^i}\Phi(c)|_{c=-\Delta/2} \quad (i = 2, 4),$$

and Δ means the Mahalanobis distance, $\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$. For the Z-rule, Memon and Okamoto (1971) derived an asymptotic expansion formula up to $O_2(n)$. Its asymptotic expansion formula of misclassification probability is given as

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

where

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

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

Additionally, Siotani and Wang (1977) derived the asymptotic expansions of misclassification probability up to $O_3(n)$ for the W-rule and the Z-rule. It is known that the accuracy of these formulas depends on the dimension and the Mahalanobis distance between the two populations. When the dimension is large, it is known that the approximations by these expansion formulas are poor. Therefore, when both the dimension and the sample sizes are large, we need other approximations.

When both the dimension and the sample sizes are large, Deev (1970) gave an asymptotic expansion for the W-rule in the case $N_1 = N_2$. Wyman et al. (1990) compared the accuracy of several approximations for the W-rule in the case $N_1 = N_2$, and pointed that the approximation due to Raudys (1972) has overall the best accuracy for the combinations of the parameters considered in the study. Saranadasa (1993) obtained the limiting distribution function of misclassification probability for the the Z-rule using the formal Edgeworth expansion of the distribution of sum of random variables which are independent but not distributed identically. The approximated misclassification probability is given by

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

where $n = N_1 + N_2 - 2$ and $p/n \to y \in (0,1)$ when $n \to \infty$. Fujikoshi and Seo (1998) also derived the limiting distribution for a class of discriminant rules which includes both the W-rule and the Z-rule. They considered the statistic

$$\tilde{T} = \frac{1}{2}\{(\mathbf{X} - \overline{\mathbf{X}}_2)'S^{-1}(\mathbf{X} - \overline{\mathbf{X}}_2) - d(\mathbf{X} - \overline{\mathbf{X}}_1)'S^{-1}(\mathbf{X} - \overline{\mathbf{X}}_1)\}. \tag{2.4}$$

Note that $\tilde{T}$ is obtained from T in (1.1) by putting $a = -(d-1)/(d+1)$ and $b = 0$. They derived the approximated misclassification probability as

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

where

$$\zeta = \frac{1}{2}\left(\frac{N}{N-p}\right)\left\{\varDelta^2 + \frac{p}{N_1N_2}(N_1 - N_2) + p(1-d)(1+N_1^{-1})\right\},$$

$$\sigma^2 = \left(\frac{N}{N-p}\right)^3\left(\varDelta^2 + \frac{pN}{N_1N_2}\right),$$

and $N = N_1 + N_2$. By numerical experiments they show that (2.5) gives a good approximation even for small dimension. Note that the (2.3) is not the same as (2.5) for the Z-rule. Fujikoshi (2000) derived an error bound for the approximation in the W-rule.

Recently, Tonda and Wakaki (2003) derived an asymptotic expansion of misclassification probability for the W-rule which is given as

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

where

$$\begin{aligned} v &= v(\Delta) \\ &= -\frac{1}{2}\left(\frac{N-p}{N-1}\right)^{1/2}\left\{\Delta^2 + \frac{(p-1)(N_1-N_2)}{N_1N_2}\right\}\left(\Delta^2 + \frac{N(p-1)}{N_1N_2}\right)^{-1/2}, \end{aligned}$$

$f_1(\Delta)$ is the term of O_1, and O_j means the term of the j-th order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1})$. They also derived an unbiased estimator up to O_1 by using (2.6).

For the sample quadratic discriminant function (SQDF, shortly), some results have been obtained. Wakaki (1990) derived an asymptotic approximation of misclassification probability in the case of proportional covariance matrices when only the sample sizes are large, under the assumption that

$$\Sigma_i = \lambda_i^{-1} I \quad (i = 1, 2), \qquad \lambda_1\mu_1 - \lambda_2\mu_2 = 0, \qquad \lambda_1 - \lambda_2 = 1, \qquad N_1 = N_2 = N.$$

Then, the misclassification probability can be expressed as

$$P(2|1) \simeq P_{Q_0} + \frac{P_{Q_1}}{N}, \tag{2.7}$$

where P_{Q_0} and P_{Q_1} are defined in terms of the parameter p, N_i and λ_j and the non-central χ^2-distribution. When both the dimension and the sample sizes are large, Matsumoto and Wakaki (2003) derived an asymptotic expansion of misclassification probability up to $O(p^{-1})$. The method is based on the formal Edgeworth expansion of the distribution of sum of random variables which are independent but not distributed identically. In their paper, it is assumed that $n_i = N_i - 1$, $v_i = p/n_i$, $\liminf_{n\to\infty} p/n_i > 0$ and $n_i > p$. Then, the misclassification probability was given in the form

$$P(2|1) = P_1 + \frac{1}{\sqrt{p}} P_2 + \frac{1}{p} P_3 + o(p^{-1}), \tag{2.8}$$

where P_1, P_2 and P_3 are defined in terms of the standard normal distribution, the parameter p, N_i, μ_i, Σ_i and the asymptotic mean, variance, 3rd and 4th cumulants of the SQDF (for details, see Matsumoto and Wakaki (2003)).

2.2. Asymptotic expansion of misclassification probability

In this section we derive an asymptotic expansion of misclassification probability for the statistic (1.1). Our methods are as follows: First, we represent the statistic T as a function of several random variables distributed as normal and χ^2-distributions. Secondly, we expand the characteristic function of T^*, standardization of T. Finally, we derive an asymptotic expansion of distribution function by inverting an expanded characteristic function. From the result we obtain an asymptotic expansion of misclassification probability.

Note that the results when the observation vector $\boldsymbol{x}$ comes from Π_2 can be obtained from that for $\boldsymbol{x} \in \Pi_1$ by replacing (a, b, N_1, N_2) with $(-a, -b, N_2, N_1)$. Therefore, we consider the case of $\boldsymbol{x} \in \Pi_1$.

As in the paper of Fujikoshi and Seo (1998), we express T by using random variables distributed as normal and χ^2-distributions as

$$T = \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 T_3 - b, \tag{2.9}$$

where

$$\alpha_1 = \frac{1}{2}\frac{nN_1}{N_1+1}\left\{(1+a)\left(\frac{N_1+1}{N_1}\right)^2 - (1-a)\right\},$$

$$\alpha_2 = -(1-a)\frac{n}{N_1+1}\sqrt{\frac{N+1}{N_2}}, \qquad \alpha_3 = -\frac{1}{2}(1-a)\frac{n(N+1)}{(N_1+1)N_2},$$

$$T_1 = g_1 \cdot g_2^2 + g_3, \qquad T_2 = \sqrt{N_1}g_1 \cdot g_2 \cdot g_4^{1/2}, \qquad T_3 = g_1 \cdot g_4,$$

$$g_1 = \frac{1}{y_2}, \qquad g_2 = z_1 + z_2\left\{\frac{y_3 y_5}{y_4(y_6 + z_2^2)}\right\}^{1/2}, \qquad g_3 = \frac{y_3}{y_4}, \qquad g_4 = (z_3 + \xi)^2 + y_1,$$

$$z_i \sim N(0,1), \qquad y_i \sim \chi^2_{f_i},$$

$$f_1 = f_3 = f_5 = p - 1, \qquad f_2 = n - p + 1, \qquad f_4 = n - p + 2, \qquad f_6 = p - 2,$$

$$\xi = \xi(\Delta) = \left\{\frac{N_2(N_1+1)}{N+1}\right\}^{1/2}\Delta, \qquad \Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2),$$

$$N = N_1 + N_2, \qquad n = N - 2.$$

Here, z_i's and y_i's are independent. We consider an asymptotic expansion of T in the situation where N_1, N_2 and p are large. Assume that p and N_i have the same order. For a convenient notation of order, let

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

where $\rho_i, \lambda_i, \alpha$ and $\zeta = O(1)$ when $m \to \infty$. Here, m may be the dimension p or the sample size N_i. Additionally, let

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

Then, u_i is asymptotically distributed as $N(0,1)$ when f_i tends to infinity. Using them, T_1, T_2 and T_3 are shown as

$$\begin{aligned}
T_1 &= \frac{\rho_3}{\rho_4} + \frac{1}{\sqrt{m}}\left(\frac{\sqrt{2}\sqrt{\rho_3}}{\rho_4}u_3 - \frac{\sqrt{2}\rho_3}{\rho_4^{3/2}}u_4\right) \\
&\quad + \frac{1}{m}\left(\frac{2\rho_3}{\rho_4^2}u_4^2 - \frac{2\sqrt{\rho_3}}{\rho_4^{3/2}}u_3u_4 + \frac{z_1^2}{\rho_2} + \frac{2\rho}{\rho_2}z_1z_2 + \frac{\rho^2}{\rho_2}z_2^2\right) + O(m^{-3/2}), \\
T_2 &= \frac{\sqrt{\lambda_1\eta_1}}{\rho_2}(z_1 + \rho z_2) \\
&\quad + \frac{1}{\sqrt{m}}\left\{\frac{\rho\sqrt{\lambda_1\eta_1}}{\sqrt{2}\rho_2}\left(\frac{1}{\sqrt{\rho_3}}u_3z_2 - \frac{1}{\sqrt{\rho_4}}u_4z_2 + \frac{1}{\sqrt{\rho_5}}u_5z_2 - \frac{1}{\sqrt{\rho_6}}u_6z_2\right)\right. \\
&\quad \left. + \frac{\sqrt{2\lambda_1\rho_1}}{2\rho_2\sqrt{\eta_1}}u_1(z_1 + \rho z_2) + \frac{\sqrt{\lambda_1}\zeta}{\rho_2\sqrt{\eta_1}}z_3(z_1 + \rho z_2) - \frac{\sqrt{2\lambda_1\eta_1}}{\rho_2^{3/2}}u_2(z_1 + \rho z_2)\right\} \\
&\quad + \frac{1}{m}\left\{\frac{\rho\sqrt{\lambda_1\eta_1}}{4\rho_2}\left(-\frac{1}{\rho_3}u_3^2z_2 + \frac{3}{\rho_4}u_4^2z_2 - \frac{1}{\rho_5}u_5^2z_2 + \frac{3}{\rho_6}u_6^2z_2 - \frac{2}{\rho_6}z_2^3\right)\right. \\
&\quad \left. + \frac{2\sqrt{\lambda_1\eta_1}}{\rho_2^2}u_2^2(z_1 + \rho z_2) + \frac{\sqrt{\lambda_1}\rho_1}{2\rho_2\sqrt{\eta_1}^3}z_3^2(z_1 + \rho z_2) - \frac{\sqrt{\lambda_1}\rho_1}{4\rho_2\sqrt{\eta_1}^3}u_1^2(z_1 + \rho z_2)\right\} \\
&\quad + O(m^{-3/2}), \\
T_3 &= \frac{\eta_1}{\rho_2} + \frac{1}{\sqrt{m}}\left(\frac{\sqrt{2}\sqrt{\rho_1}}{\rho_2}u_1 + \frac{2\zeta}{\rho_2}z_3 - \frac{\sqrt{2}\eta_1}{\rho_2^{3/2}}u_2\right) \\
&\quad + \frac{1}{m}\left(\frac{1}{\rho_2}z_3^2 - \frac{2\sqrt{\rho_1}}{\rho_2^{3/2}}u_1u_2 - \frac{2\sqrt{2}\zeta}{\rho_2^{3/2}}z_3u_2 + \frac{2\eta_1}{\rho_2^2}u_2^2\right) + O(m^{-3/2}), \qquad (2.10)
\end{aligned}$$

where $\eta_1 = \rho_1 + \zeta^2$ and $\rho = \sqrt{\rho_3\rho_5/\rho_4\rho_6}$. Therefore, T in (2.9) can be expressed as

$$T = t_0 + t_1 + t_2, \qquad (2.11)$$

where

$$
\begin{aligned}
t_0 &= t_{00} + t_{01} z_1 + t_{02} z_2, \\
t_{00} &= -b + \frac{n}{N_1 + 1}\left\{(N_1 a + 1)\frac{f_3}{f_4} - \frac{\eta_2(N+1)}{2 f_2 N_2}\right\}, \\
t_{01} &= -\frac{n}{N_1 + 1}\sqrt{\frac{N+1}{N_2}}\sqrt{N_1}\,\frac{\sqrt{\eta_2}}{f_2}, \\
t_{02} &= -\frac{n}{N_1 + 1}\sqrt{\frac{N+1}{N_2}}\sqrt{N_1}\,\frac{\sqrt{\eta_2}}{f_2}\rho, \\
\rho &= \sqrt{\frac{p-1}{N-p}}, \qquad \eta_2 = f_1 + \xi^2,
\end{aligned}
\tag{2.12}
$$

and t_1 and t_2 are terms of $O(m^{-1/2})$ and $O(m^{-1})$, respectively (calculated easily from (2.10)). Using (2.12), T is standardized as

$$
\begin{aligned}
T^* = \frac{T - t_{00}}{\sigma} &= \frac{t_0 - t_{00}}{\sigma} + \frac{t_1}{\sigma} + \frac{t_2}{\sigma} \\
&= t_0^* + \frac{t_1}{\sigma} + \frac{t_2}{\sigma},
\end{aligned}
$$

where $t_0^* = (t_0 - t_{00})/\sigma$ and $\sigma = \sqrt{t_{01}^2 + t_{02}^2}$. Then, the characteristic function of T^* can be expressed as

$$
\begin{aligned}
C_{T^*}(s) &= E(e^{isT^*}) \\
&= E\left[\exp\left\{is\left(t_0^* + \frac{t_1}{\sigma} + \frac{t_2}{\sigma}\right)\right\}\right] + O(m^{-3/2}) \\
&= E(e^{ist_0^*}) + \frac{1}{\sigma}E(e^{ist_0^*} ist_1) + \frac{1}{\sigma}E\left[e^{ist_0^*}\left\{ist_2 + \frac{(is)^2}{2}\frac{t_1^2}{\sigma}\right\}\right] \\
&\quad + O(m^{-3/2}).
\end{aligned}
\tag{2.13}
$$

We calculate the expectations in (2.13) using the density function of z_i given by

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

and the density function of u_i which is expanded as

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

After much computation, it is shown that

$$C_{T^*}(s) = e^{-s^2/2}\left[1 + \left\{\frac{1}{\sigma}\tau_1(is) + \frac{1}{\sigma^2}\tau_2(is)^2 + \frac{1}{\sigma^3}\tau_3(is)^3 + \frac{1}{\sigma^4}\tau_4(is)^4\right\}\right] + O(m^{-3/2}), \tag{2.14}$$

where

$$\tau_1 = \alpha_{11}\left(\frac{2f_3}{f_4^2} + \frac{1+\rho^2}{f_2}\right) + \alpha_{31}\left(\frac{2\eta_2}{f_2^2} + \frac{1}{f_2}\right) + \alpha_{12}\frac{f_3}{f_4} + \alpha_{32}\frac{\eta_2}{f_2},$$

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

$$\begin{aligned}\tau_3 = {} & \alpha_{11}\frac{1}{f_2}(t_{01} + \rho t_{02})^2 + \alpha_{11}\alpha_{21}t_{02}\left(\frac{\rho\sqrt{N_1\eta_2}}{f_2 f_4} + \frac{f_3\rho\sqrt{N_1\eta_2}}{f_2 f_4^2}\right) \\ & + \alpha_{31}\alpha_{21}(t_{01} + \rho t_{02})\left(\frac{2\sqrt{N_1}\eta_2^{3/2}}{f_2^3} + \frac{2\sqrt{N_1\eta_2}}{f_2^2} - \frac{f_1\sqrt{N_1}}{f_2^2\sqrt{\eta_2}}\right),\end{aligned}$$

$$\begin{aligned}\tau_4 = {} & -\alpha_{21}\frac{\rho\sqrt{N_1\eta_2}}{2f_2 f_6}t_{02}^3 + \alpha_{21}^2\left\{\frac{(t_{01} + \rho t_{02})^2}{4}\left(\frac{2N_1\eta_2}{f_2^3} + \frac{N_1}{f_2^2} - \frac{N_1 f_1}{2f_2^2\eta_2}\right)\right. \\ & \left. + \frac{\rho^2\eta_2 N_1}{8f_2^2}t_{02}^2\left(\frac{1}{f_3} + \frac{1}{f_4} + \frac{1}{f_5} + \frac{1}{f_6}\right)\right\},\end{aligned}$$

$$\alpha_{11} = \frac{N_1 n}{N_1 + 1}a + \frac{n}{N_1 + 1}, \qquad \alpha_{12} = \frac{n}{N_1 + 1}a + \frac{n}{2N_1(N_1 + 1)},$$

$$\alpha_{21} = -\frac{n}{N_1 + 1}\sqrt{\frac{N+1}{N_2}}, \qquad \alpha_{22} = \frac{n}{N_1 + 1}\sqrt{\frac{N+1}{N_2}}a,$$

$$\alpha_{31} = -\frac{n(N+1)}{2N_2(N_1 + 1)}, \qquad \alpha_{32} = \frac{n(N+1)}{2N_2(N_1 + 1)}a.$$

Inverting (2.14), we obtain the following theorem.

THEOREM 2.1. *The distribution function of T^* can be asymptotically expanded as*

$$F_{T^*}(x) = \Phi(x) - \phi(x)\left\{\frac{1}{\sigma}\tau_1 + \frac{1}{\sigma^2}\tau_2 h_1(x) + \frac{1}{\sigma^3}\tau_3 h_2(x) + \frac{1}{\sigma^4}\tau_4 h_3(x)\right\} + O_{3/2}, \tag{2.15}$$

where $h_r(x)$ is the Hermite polynomial of degree r,

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

and O_j means the term of the j-th order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1})$.

Theorem 2.1 implies the following theorem for the misclassification probability.

THEOREM 2.2. *The misclassification probability for discriminant function* (1.1) *is asymptotically given by*

$$\begin{aligned}
P(2|1) = P(T > 0 \,|\, \boldsymbol{x} \in \Pi_1) &= P\left(T^* > -\frac{t_{00}}{\sigma} \,|\, \boldsymbol{x} \in \Pi_1\right) \\
&= 1 - F_{T^*}\left(-\frac{t_{00}}{\sigma}\right) \\
&= \Phi\left(\frac{t_{00}}{\sigma}\right) + \phi\left(-\frac{t_{00}}{\sigma}\right) \\
&\quad \times \left\{\frac{1}{\sigma}\tau_1 + \frac{1}{\sigma^2}\tau_2 h_1\left(-\frac{t_{00}}{\sigma}\right) + \frac{1}{\sigma^3}\tau_3 h_2\left(-\frac{t_{00}}{\sigma}\right) + \frac{1}{\sigma^4}\tau_4 h_3\left(-\frac{t_{00}}{\sigma}\right)\right\} + O_{3/2}.
\end{aligned} \tag{2.16}$$

The first term of (2.16) with $(a, b) = (-(d-1)/(d+1), 0)$ is the same as (2.5), the result of Fujikoshi and Seo (1998), and it makes the same form at (2.6), the result of Tonda and Wakaki (2003) for the W-rule. The validity of our expansions follows from the results by Bhattacharya and Ghosh (1978).

2.3. Estimation of misclassification probability

The expansion (2.16) for $P(2|1)$ in Theorem 2.2 includes the parameter Δ which dependes on population parameters μ_i and Σ. In many actual cases, Δ is unknown. Then, we usually use an estimator in place of Δ. Since

$$E(D^2) = E[(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)] = \frac{n}{n-p-1}\left[\Delta^2 + \frac{pN}{N_1 N_2}\right],$$

we can estimate Δ^2 by

$$D_u^2 = \frac{n-p-1}{n} D^2 - \frac{pN}{N_1 N_2}, \tag{2.17}$$

which is consistent as well as unbiased in our framework. However, when we substitute (2.17) to $\varDelta$ in (2.16), the resultant estimator of misclassification probability has the bias of order $O_{1/2}$. So, some correction is needed.

From Theorem 2.2,

$$P(2|1) = \Phi(v) + \phi(v) f(\varDelta),$$

where $v = v(\varDelta^2) = t_{00}/\sigma$. As in Tonda and Wakaki (2003), we consider an estimator

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

where $\tilde{v} = v(D_u^2)$ and Q_1 is the term of O_1. To construct an asymptotic unbiased estimator up to the term of O_1, we define Q_1 such that the bias of Q_{MW} is $O_{3/2}$. The bias of Q_{MW} can be expressed as

$$\begin{aligned} \mathrm{Bias}(Q_{MW}) &= E_{\bar{x}_1, \bar{x}_2, S}[P_{\bar{x}_1, \bar{x}_2, S}(2|1) - Q_{MW}] \\ &= P(2|1) - E_{\bar{x}_1, \bar{x}_2, S}[\Phi(\tilde{v}) - Q_1], \end{aligned} \tag{2.18}$$

where $P_{\bar{x}_1, \bar{x}_2, S}(2|1)$ means the exact misclassification probability. Using the same way for expanding the characteristic function of T in Section 2.2, we obtain the following lemma.

Lemma 2.1. *It holds that*

$$E_{\bar{x}_1, \bar{x}_2, S}[\Phi(\tilde{v})] = \Phi(v) + \phi(v) g(\varDelta) + O_{3/2}, \tag{2.19}$$

where

$$\begin{aligned} g(\varDelta) &= v'(\varDelta^2) g_2(\varDelta^2) + \frac{g_1(\varDelta^2)}{2} \{v''(\varDelta^2) - v(\varDelta)[v'(\varDelta^2)]^2\}, \\ g_1(\varDelta^2) &= \left(\frac{N+1}{N_1 N_2}\right)^2 \left\{2(p-1) + 4\xi^2 + \frac{2\eta_2}{f_2^2}\right\} + \frac{4(N+1)}{(N_1+1)^2 N_2} \eta_2 (1+\rho), \\ g_2(\varDelta^2) &= \frac{N+1}{(N_1+1) N_2} \left\{1 + \frac{2\eta_2}{f_2}\right\} + \frac{n}{N_1(N_1+1)} \frac{f_3}{f_4} - \frac{N}{N_1 N_2}, \\ v'(\varDelta^2) &= \frac{d}{dx} v(x)\bigg|_{x=\varDelta^2}, \qquad v''(\varDelta^2) = \frac{d^2}{dx^2} v(x)\bigg|_{x=\varDelta^2}. \end{aligned} \tag{2.20}$$

From Theorem 2.2 and Lemma 2.1, it follows that

$$\mathrm{Bias}(Q_{MW}) = \phi(v)\{f(\varDelta) - g(\varDelta)\} - E[Q_1] + O_{3/2}.$$

If $E[Q_1] = \phi(v)\{f(\varDelta) - g(\varDelta)\}$, the bias of Q_{MW} becomes $O_{3/2}$. Therefore, the estimator defined by

$$Q_{MW} = \Phi(\tilde{v}) + \hat{Q}_1, \qquad \hat{Q}_1 = \phi(\tilde{v})\{f(D_u) - g(D_u)\} \tag{2.21}$$

is asymptotically unbiased up to the term of O_1.

3. Coefficients to minimize the risk

In this section we derive the coefficients a and b which make an optimal discriminant rule in the class of (1.1).

The asymptotic expansion formula of misclassification probability derived from (2.16) is expressed as the function of coefficients a and b as follows:

$$\begin{aligned} P(2|1) &= \Phi(-s_1 b + s_2 a + s_{31}) + \phi(-s_1 b + s_2 a + s_{31}) \\ &\quad \times \left\{ \frac{1}{\sigma}(\tau_{10} + \tau_{11} a) + \frac{1}{\sigma^2}(\tau_{20} + \tau_{21} a + \tau_{22} a^2) h_1(s_1 b - s_2 a - s_{31}) \right. \\ &\quad + \frac{1}{\sigma^3}(\tau_{30} + \tau_{31} a) h_2(s_1 b - s_2 a - s_{31}) \\ &\quad \left. + \frac{1}{\sigma^4}\tau_4 h_3(s_1 b - s_2 a - s_{31}) \right\} + O_{3/2}, \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} s_1 &= \left(\frac{N-1}{N-p}\right)^{-3/2}\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}, \\ s_2 &= \left(\frac{N-1}{N-p}\right)^{-1/2}\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2} p, \\ s_{31} &= -\frac{1}{2}\left(\frac{N-1}{N-p}\right)^{-1/2}\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}\left\{\Delta^2 + \frac{p(N_1 - N_2)}{N_1 N_2}\right\}, \\ s_{32} &= -\frac{1}{2}\left(\frac{N-1}{N-p}\right)^{-1/2}\left(\Delta^2 + \frac{pN}{N_1 N_2}\right)^{-1/2}\left\{\Delta^2 - \frac{p(N_1 - N_2)}{N_1 N_2}\right\}, \\ \sigma &= \sqrt{t_{01}^2 + t_{02}^2}, \qquad \Delta = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \end{aligned} \tag{3.2}$$

By changing the role of N_1 for N_2 in (3.1), we obtain

$$\begin{aligned} P(1|2) &= \Phi(s_1 b - s_2 a + s_{32}) + \phi(s_1 b - s_2 a + s_{32}) \\ &\quad \times \left\{ \frac{1}{\sigma}(\tau'_{10} - \tau_{11} a) + \frac{1}{\sigma^2}(\tau_{20} - \tau'_{21} a + \tau_{22} a^2) h_1(-s_1 b + s_2 a - s_{32}) \right. \\ &\quad + \frac{1}{\sigma^3}(\tau'_{30} - \tau_{31} a) h_2(-s_1 b + s_2 a - s_{32}) \\ &\quad \left. + \frac{1}{\sigma^4}\tau_4 h_3(-s_1 b + s_2 a - s_{32}) \right\} + O_{3/2}. \end{aligned} \tag{3.3}$$

Note that τ_{ij} is composed of N_1, N_2, p and Δ and τ'_{ij} is obtained from τ_{ij} by changing the role of N_1 and N_2. (see (2.16)). Using (3.1) and (3.3), we consider to derive the coefficients a and b with which the risk (1.2) is asymptotically smaller.

However, it is very difficult to obtain the coefficients a and b actually, based on the formulas (3.1) and (3.3). So, at first, we consider to minimize only the leading term of (1.2),

$$c_1\Phi(-s_1b + s_2a + s_{31}) + c_2\Phi(s_1b - s_2a + s_{32}). \tag{3.4}$$

This leads to consider the coefficients a and b such that

$$s_2a - s_1b + \frac{s_{31} - s_{32}}{2} = (s_{31} + s_{32})^{-1} \log \frac{c_1}{c_2}. \tag{3.5}$$

Under the condition (3.5), we can obtain the coefficient a and b which makes an optimal discriminant rule in the class of (1.1) with neglecting the terms of $O_{3/2}$.

Theorem 3.1. *In the class of* (1.1) *with* (3.5), *an optimal discriminant rule is made from the coefficients* $(a,b) = (a_0(\Delta^2), b_0(\Delta^2))$, *where*

$$\begin{aligned} a_0(\Delta^2) = &\left[\frac{c_2\phi(\gamma_2)\tau_{11}}{\theta^{1/2}} - \frac{c_1\phi(\gamma_1)\tau_{11}}{\theta^{1/2}} - \frac{c_2\phi(\gamma_2)\tau'_{21}\gamma_2}{\theta} + \frac{c_1\phi(\gamma_1)\tau_{21}\gamma_1}{\theta}\right. \\ &\left. + \frac{c_2\phi(\gamma_2)\tau_{31}(-1+\gamma_2^2)}{\theta^{3/2}} - \frac{c_1\phi(\gamma_1)\tau_{31}(-1+\gamma_1^2)}{\theta^{3/2}}\right] \\ &\times \left[-\frac{2c_2\phi(\gamma_2)\tau_{22}\gamma_2}{\theta} - \frac{2c_1\phi(\gamma_1)\tau_{22}\gamma_1}{\theta}\right]^{-1}, \end{aligned} \tag{3.6}$$

$$b_0(\Delta^2) = \frac{s_2}{s_1}a_0(\Delta^2) + \frac{s_{31} - s_{32}}{2s_1} - \frac{\log\frac{c_1}{c_2}}{s_1(s_{31} + s_{32})}, \tag{3.7}$$

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

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

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

and $\phi(x)$ *is the density function of the standard normal distribution.*

In the above theorem, the quantiles $a_0(\Delta^2)$ and $b_0(\Delta^2)$ include the unknown parameter Δ which is made from population parameters $\boldsymbol{\mu}_i$ and Σ. So, we estimate Δ^2 by (2.17).

4. Simulation

In this section we carry out numerical experiments for three purposes. The first purpose is to examine the accuracy of the asymptotic expansion formula given by Theorem 2.2, the second is for the accuracy of the asymptotic unbiased estimator given by (2.21), and the third is to compare the performance of the W-rule, the Z-rule and our new rule derived in section 3.

4.1. Methods

Without loss of generality we assume that $\mu_1 = -\mu_2 = (\Delta/2, 0, \dots, 0)$ and $\Sigma = I_p = \mathrm{diag}(1, \dots, 1)$. The configurations of the values of N_1, N_2, p, Δ, c_1 and c_2 are all those possible combination of

p; 5, 10, 15, 20,

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

Δ; 1.05, 1.68, 2.56, 3.29,

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

Here, $(N_1, N_2) = (10, 10)$ is eliminated in the case of $p = 15, 20$. For each of the configurations, we calculate the approximated misclassification probability by using the asymptotic expansion formula given by Theorem 2.2. By using Monte-Carlo method, we calculate the risks of the W-rule, the Z-rule and our new rule (referred as Min-rule). We also estimate the expectations of the asymptotic unbiased estimators of the misclassification probabilities given by (2.21) for the W-rule and the Z-rule.

4.2. Results and comments

Table 1 and Table 2 show the risks of the three classification rules estimated by Monte-Carlo method in the case of $\Delta = 1.05$ and $\Delta = 1.68$, respectively. Table 4 shows that the accuracy of the approximations of $P(2|1)$ for the W-rule and the Z-rule in the case that $p = 10$ and $c_1 = c_2$, where

Table 1. Risks of the three classification rules in $\Delta = 1.05$.

$\Delta = 1.05$		$p = 10$			$p = 20$		
(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

Table 2. Risks of the three classification rules in $\Delta = 1.68$.

$\Delta = 1.68$		$p = 10$			$p = 20$		
(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

Table 3. The values of coefficients $(a_0(\Delta^2), b_0(\Delta^2))$ and Z-rule's a.

			$\Delta = 1.05$				$\Delta = 1.68$			
(c_1, c_2)	(N_1, N_2)	Z-rule	$p = 10$		$p = 20$		$p = 10$		$p = 20$	
$(1, 1)$	$(10, 10)$	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

"Sim.", "W_O", "Z_{MO}", "W_{AE}", "Z_{AE}", "W_{Est}", "Z_{Est}", "W_{Bias}" and "Z_{Bias}" mean as follows:

- "Sim.": the misclassification probability estimated by Monte-Carlo methods,
- "W_O": the asymptotic approximation given by Okamoto (1963, 1968),
- "Z_{MO}": the asymptotic approximation given by Memon and Okamoto (1971),
- "W_{AE}" and "Z_{AE}": an asymptotic expansion given by (2.16) for the W-rule and the Z-rule,
- "W_{Est}" and "Z_{Est}": the expectations of estimators given by (2.16) in which Δ is replaced with (2.17),
- "W_{Bias}" and "Z_{Bias}": the expectation of asymptotic unbiased estimator.

Figure 1 shows the difference between the values of $P(2|1)$ (by Monte-Carlo Method) and approximated or estimated values of $P(2|1)$ in Table 3. (For the results corresponding to the W-rule, see also Tonda and Wakaki (2003).) Here, all standard deviations of the estimated values given by Monte-Carlo method are not over 0.01.

In Table 1 and Table 2, we find that the risks of Min-rule are the smallest among the three classification rules for all parameters when $\Delta = 1.05$. While in the case of $\Delta = 1.68$ and $p = 10$, some risks of Min-rule are larger than

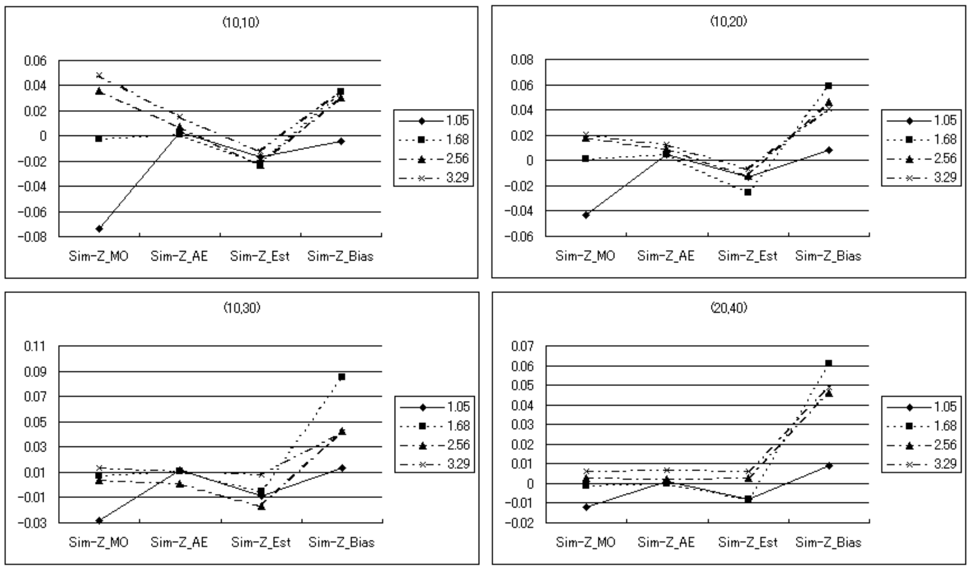

Fig. 1. Comparison of the difference of the simulation results $P(2|1)$ in the case of $p = 10$ and $c = 1$ for the Z-rule.

those of the W-rule or the Z-rule. The performance of Min-rule looks poor for large Δ. In the case that $c_1 \neq c_2$ we find similar tendency to the case that $c_1 = c_2$. We think one reason of the poor performance of Min-rule for large Δ and small p is the poor accuracy of the approximation formula given by the asymptotic expansion. We show the values of $(a_0(\Delta^2), b_0(\Delta^2))$ and the coefficient a for the Z-rule in Table 3 (for the W-rule, always $a = 0$).

In Table 4 and Figure 1 we can see that our approximation formula and asymptotically unbiased estimator work better than the classical method such as W_O and Z_{MO} for large dimension and samples. For small Δ, we can see that the expectation of asymptotic unbiased estimator has good approximation.

Acknowledgment

I wish to express my deepest gratitude to Prof. Y. Fujikoshi and Dr. H. Wakaki of Hiroshima University for their constant comments and useful advices. Also I would like to thank Prof. M. Ohtaki and Dr. K. Satoh of Hiroshima University for their useful advices of computer and making simulation. Further, I also thank the referee for some very useful suggestions.

Table 4. Results of $P(2|1)$ in the simulation with $p = 10$ and $c = 1$.

$p = 10,\ c = 1$		W-rule					Z-rule				
(N_1, N_2)	Δ	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

References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Second Edition. New York, Wiley, 1984.

[2] Z. D. Bai and L. C. Zhao, Edgeworth expansions of distribution functions of independent random variables, Sci. China Ser. A **29**, 1–22, (1986).

[3] O. E. Barndorff-Nielsen and D. R. Cox, Asymptotic Techniques for Use in Statistics, Chapman and Hall, 1989.

[4] R. N. Bhattacharya and J. K. Ghosh, On the validity of the formal Edgeworth expansions, J. Multivariate Anal. **27**, 68–79, (1978).

[5] R. N. Bhattacharya and R. Ranga Rao, Normal Approximation and Asymptotic Expansion, John Wiley and Sons, New York, 1976.

[6] A. D. Deev, Representation of statistics of discriminant analysis and asymptotic expansions when space dimensions are comparable with sample size, Soviet Math. Dokl. **11**, 1547–1550, (1970).

[7] Y. Fujikoshi, On relationship between the AIC and the overall error rates for selection of variables in a discriminant analysis, Theory Decis. Lib. Ser. B: Math. Statist. Methods, 121–138, (1987).

[8] Y. Fujikoshi, Error bounds for asymptotic approximations of the linear discriminant function when the sample sizes and dimensionality are large, J. Multivariate Anal. **73**, 1–17, (2000).

[9] Y. Fujikoshi and T. Seo, 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, (1998).

[10] P. A. Lachenbruch, On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficients, Biometrics **24**, 823–834, (1968).

[11] C. Matsumoto and H. Wakaki, Asymptotic expansion of the quadratic discriminant function when the dimension and sample sizes are large, TR No. 03-14, Statistical Research Group, Hiroshima University, 2003.

[12] G. J. McLachlan, An asymptotic unbiased technique for estimating the error rates in discriminant analysis, Biometrics **30**, 230–249, (1974).

[13] A. Z. Memon and M. Okamoto, Asymptotic expansion of the distribution of the Z statistics in discriminant analysis, J. Multivariate Anal. **1**, 294–307, (1971).

[14] R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Willey and Sons, Inc., New York, 1982.

[15] M. Okamoto, An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Statist. **34**, 1286–1301, (1963). Correction, Ann. Math. Statist. **39**, 1358–1359, (1968).

[16] J. T. Page, Error-rate estimation in discriminant analysis, Technometrics **27**, 189–198, (1985).

[17] S. Raudys, On the amount of priori information in designing the classification algorithm, Tech. Cybern **4**, 168–174 (in Russian), (1972).

[18] H. Saranadasa, 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, (1993).

[19] M. Siotani and R. H. Wang, 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, (1977).

[20] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, INC, 1985.
[21] M. S. Srivastava and C. G. Khatri, An Introduction to Multivariate Statistics, North-Holland, New York, 1979.
[22] T. Tonda and H. Wakaki, EPMC estimation in discriminant analysis when the dimension and sample sizes are large, TR No. 03-08, Statistical Research Group, Hiroshima University, 2003.
[23] H. Wakaki, Comparison of linear and quadratic discriminant functions, Biometrika **77**, 227–229, (1990).
[24] H. Wakaki and M. Aoshima, A class of Linear Discriminant function, (to appear).
[25] F. J. Wyman, D. M. Young and D. W. Turner, A comparison of asymptotic error rate expectations for the sample linear discriminant function, Pattern Recognition **23**, 775–783, (1990).

Chieko Matsumoto
Department of Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526*, Japan*
e-mail: shisato@hiroshima-u.ac.jp

Current address: Institute of Economic Research, Hitotsubashi University, Kunitachi, Tokyo, 186-8603, Japan
shisato@ier.hit-u.ac.jp