High-dimensional asymptotic results for EPMCs of W- and Z- rules

Takayuki Yamada[1], Tetsuro Sakurai[2] and Yasunori Fujikoshi[3]

[1] Institute for Comprehensive Education, Center of General Education
Kagoshima University
1-21-30 Korimoto, Kagoshima 890-0065 Japan

[2] Center of General Education
Tokyo University of Science, Suwa
5000-1, Toyohira, Chino-shi, Nagano, 391-0292, Japan

[3] Department of Mathematics, Graduate School of Science
Hiroshima University
1-3-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8626, Japan

Abstract

This paper is concerned with high-dimensional asymptotic results for W- and Z- rules when the sample size N and the dimension are large. First we give a unified location and scale mixture expression of the standard normal distribution for W and Z statistics. Then, the EPMCs (Expected Probability of Misclassifications) of W- and Z- rules are obtained in expanded forms with errors of $O(N^{-2})$. It is pointed that Z-rule has smaller EER (Expected Error Rate) than W-rule when the prior probabilities are the same, neglecting the terms of $O(N^{-2})$. Further, asymptotic unbiased estimators are proposed for the EPMCs and the EERs of W- and Z- rules. Variable selection criteria are also proposed, based on asymptotic unbiased estimators of the EERs of W- and Z- rules. It is pointed that the no additional information model based on the coefficients of the linear discriminant function is closely related to the subset of variables with the minimized EER in a high dimensional situation. Accuracies of our asymptotic results are checked numerically by conducting a Mote Carlo simulation.

AMS 2000 subject classification: primary 62H30; secondary 62H12

Key Words and Phrases: Discriminant analysis, EPMC, High-dimensional asymptotic results, Method by differential operator, Selection of variable, Unified location and scale mixtured expression, W- rule, Z-rule.

[1]E-mail address:yamada@gm.kagoshima-u.ac.jp
[2]E-mail address:sakurai@rs.tus.ac.jp
[3]E-mail address:fujikoshi_y@yahoo.co.jp

1 Introduction

This paper is concerned with the problem of classifying an observation vector $\boldsymbol{x}$ as coming from one of two populations Π_1 and Π_2. Let Π_i have p-dimensional normal populations with mean vectors $\boldsymbol{\mu}_i$ and the $p \times p$ common positive definite covariance matrix $\boldsymbol{\Sigma}$, which are denoted as $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. Consider the case that all parameters are unknown. Suppose that training data $\boldsymbol{x}_{1i}, \ldots, \boldsymbol{x}_{N_i,i}$ are independently and identically distributed (i.i.d.) as $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2$. Let W be the linear discriminant function

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

where $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$ and $\boldsymbol{S}$ are the sample mean vectors and the pooled sample covariance matrix defined by

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

Then, the linear discriminant rule with a cutoff point c, which is also called W-rule, classifies $\boldsymbol{x}$ as Π_1 if $W(\boldsymbol{x}) > c$ for a constant c, and as Π_2 if $W(\boldsymbol{x}) < c$. Furthermore, Anderson [1](see also Anderson [3]; Chapter 6) introduced the other discriminant rule which is based on the likelihood ratio criterion for testing the composite null hypothesis that $\boldsymbol{x}, \boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1N_1} \in \Pi_1$ against the composite alternative hypothesis that $\boldsymbol{x}, \boldsymbol{x}_{21}, \ldots, \boldsymbol{x}_{2N_2} \in \Pi_2$, which is called maximum likelihood rule or Z-rule. Let

$$Z(\boldsymbol{x}) = \frac{1}{2}\{(1 + N_2^{-1})^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_2) - (1 + N_1^{-1})^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1)\}.$$

Then the Z-rule with a cutoff point c classifies $\boldsymbol{x}$ as Π_1 if $Z(\boldsymbol{x}) > c$ and Π_2 if $Z(\boldsymbol{x}) < c$.

There are two types of probability of misclassification. One is the probability of allocating $\boldsymbol{x}$ into Π_2 even though it is actually belonging to Π_1. The other is the probability that $\boldsymbol{x}$ is classified as Π_1 although it is actually belonging to Π_2. These two types of expected probabilities of misclassifications (EPMCs) for W- and Z- rules are expressed as

$$e_w(2|1) = P(W(\boldsymbol{x}) < c|\boldsymbol{x} \in \Pi_1) \quad \text{and} \quad e_w(1|2) = P(W(\boldsymbol{x}) > c|\boldsymbol{x} \in \Pi_2),$$
$$e_z(2|1) = P(Z(\boldsymbol{x}) < c|\boldsymbol{x} \in \Pi_1) \quad \text{and} \quad e_z(1|2) = P(Z(\boldsymbol{x}) > c|\boldsymbol{x} \in \Pi_2).$$

We also express $e_w(2|1)$ and $e_z(2|1)$ as

$$g_w(c; N_1, N_2, \Delta^2) = e_w(2|1), \quad g_z(c; N_1, N_2, \Delta^2) = e_z(2|1).$$

As is well known, the distribution of W when $\boldsymbol{x} \in \Pi_1$ is the same as that of $-W$ when $\boldsymbol{x} \in \Pi_2$ by interchanging N_1 and N_2. Similarly, the distribution of Z when $\boldsymbol{x} \in \Pi_1$ is the same as that of $-Z$ when $\boldsymbol{x} \in \Pi_2$ by interchanging N_1 and N_2. These indicate that $e_w(1|2)$ (or $e_z(1|2)$) is obtained from $e_w(2|1)$ (or $e_z(2|1)$) by replacing (c, N_1, N_2) with $(-c, N_2, N_1)$, and hence

$$e_w(1|2) = g_w(-c; N_2, N_1, \Delta^2), \quad e_z(1|2) = g_z(-c; N_2, N_1, \Delta^2).$$

Thus, in this paper, we only deal with $e_w(2|1)$ and $e_z(2|1)$. Related to a unified expression for W-rule and Z- rule, we consider Z-rule such that classifies $\boldsymbol{x}$ as Π_1 if $Z(\boldsymbol{x}) > c^*$ and as Π_2 if $W(\boldsymbol{x}) < c^*$, where

$$c^* = \sqrt{\frac{1+N^{-1}}{(1+N_1^{-1})(1+N_2^{-1})}}c.$$

That is, we consider Z- rule with cutoff point c^*.

Note that the EPMCs of W- and Z- rules are obtained from the distribution functions of W and Z. In general, it is hard to evaluate these expected probabilities of misclassification (EPMC) explicitly, but some asymptotic results including asymptotic expansions have been obtained. It is well known that the discriminant functions $W(\boldsymbol{x})$ and $Z(\boldsymbol{x})$ converges in distribution to the normal distributions, i.e.,

$$W(\boldsymbol{x}) \text{ and } Z(\boldsymbol{x}) \overset{\mathcal{D}}{\to} N((-1)^i\Delta^2/2, \Delta^2), \tag{1}$$

if $\boldsymbol{x} \in \Pi_i$ under the asymptotic framework A0:

$$\text{A0}: \ N_1 \to \infty, N_2 \to \infty, N_1/N_2 \to \gamma \in (0,\infty), p \text{ is fixed.}$$

Here, $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Okamoto [19] derived an asymptotic expansion of the distribution of $W(\boldsymbol{x})$ up to terms of order n^{-1}, and Siotani and Wang [21], [22] extended it up to terms of order n^{-3}. Furthermore, Memon and Okamoto [15] expanded the distribution of $Z(\boldsymbol{x})$ up to terms of order n^{-2} and Siotani and Wang [21], [22] extended it up to terms of order n^{-3}. Anderson [2] derived an asymptotic expansion of Studentized $W(\boldsymbol{x})$, and an asymptotic expansion of Studentized $Z(\boldsymbol{x})$ was derived by Fujikoshi and Kanazawa [8]. These and some other asymptotic results were reviewed by Siotani [20] and by McLachlan [18]. Generally, the precision of asymptotic approximations under A0 gets worth as the dimension p becomes large. As an alternative approach to overcome this shortcoming, it has been considered to derive asymptotic distributions of discriminant functions in a high-dimensional situation where n and p tend to infinity together. Fujikoshi and Seo [9] derived the limiting distribution of a general discriminant function for a class of discriminant rules which includes both the W- rule and Z-rule under asymptotic framework A1:

$$\begin{aligned} \text{A1}: p \to \infty, \quad N_1 \to \infty, \quad N_2 \to \infty, \quad p/n \to \gamma_0 \in [0,1), \\ \text{and} \quad N_1/N_2 \to \gamma \in (0,\infty). \end{aligned}$$

Note that $m = n - p \to \infty$ under A1. Matsumoto [14] generalized Fujikoshi and Seo [9]'s limiting result to an asymptotic expansion up to terms of order $O_{3/2}$, where $O_{j/2}$ is a term of j-th order with respect to $\{p^{-1/2}, N_1^{-1/2}, N_2^{-1/2}, m^{-1/2}\}$. Fujikoshi [6] gave a general approximation of a location and scale mixture of the standard normal distribution, and gave its explicit error bound. He applied his result to Lachenbruch [13]'s approximation of $e_w(2|1)$, and gave the error bound which is O_1 under A1. These and some other asymptotic results are also reviewed in Fujikoshi et al. [10]. High-dimensional asymptotic expansions for W have been also given by Hyodo and Kubokawa [11].

In this paper, we give asymptotic expansions of the EPMCs for W- and Z- rules with the errors of order O_2 under the asymptotic framework A1. Our derivation is based on a unified location and scale mixture expression of the standard normal distribution for W and Z. It is well known (see, e.g., Fujikoshi [6]) that W can be expressed as a location and scale mixtures of the standard normal distribution. We note that Z can be also expressed as a location and scale mixtures of the standard normal distribution. Based on our asymptotic expansion formulas for the EPMCs, it is shown that Z-rule has smaller EEP (Expected Error Rate) than W-rule when the prior probabilities are the same, neglecting the terms of $O(N^{-2})$. Further, asymptotic unbiased estimators are proposed for EPMCs of W- and Z- rules. Similarly, we propose asymptotic unbiased estimator for EEPs. It is pointed that the no additional information model based on the coefficients of the linear discriminat function is closely related to the subset of variables with the minimized EER in a high dimensional situation. We propose variable selection criteria based on unbiased estimator for EEPs. Our results are checked numerically by conducting a Mote Carlo simulation.

The present paper is organized as follows. In section 2, we give a unified location and scale mixture expression for the distributions of W and Z. Further, the expressions are expressed in terms of three standard normal variables and four chi-square variables which are independent. Applying the expression to the method of differential operator, we obtain asymptotic expansions for the EPMCs of W- and Z- rules with the errors of $O(n^{-2})$. In Section 4, it is shown that the EER of Z- rule is smaller than the one of Z- rule when the prior probabilities are the same, neglecting the terms of $O(n^{-2})$. The result is proved Appendix A. In Sections 5 and 6, asymptotic unbiased estimators for the EPMCs and the EERs of W- and Z- rules are derived. In Section 7, simulation results are results to see accuracies of our asymptotic results. In Section 8, we propose variable selection criteria based on asymptotic unbiased estimators of the EERs of W- and Z- rules. Concluding remarks are given in Section 9.

Hereafter, the symbol "$\overset{\mathcal{D}}{=}$" denotes the equality in distribution. Throughout this paper, we assume that Δ^2 converges a positive constant as $p \to \infty$.

2 A unified expression of W and Z as location and scale mixtures of $N(0,1)$

Following Lachenbruch [13], for $\boldsymbol{x} \in \Pi_1$, it can be expressed that

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\left\{\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\right\} = V_w^{1/2} Z_w - U_w, \tag{2}$$

where

$$\begin{aligned}
V_w &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2),\\
Z_w &= V_w^{-1/2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1),\\
U_w &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}D^2,
\end{aligned}$$

and D^2 is the squared sample Mahalanobis distance defined by $D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Then, it is checked that V_w is a positive random variable and (U_w, V_w) are jointly independent of Z_w. Further, Z_w is distributed as $N(0,1)$. This normality follows by considering the conditional distribution of Z_w when $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$, and $\boldsymbol{S}$ are given. In this case, W is called a location and scale mixture of the standard normal distribution. Now we consider to express U_w and V_w in terms of simple variables. Let

$$\begin{aligned}
\boldsymbol{u}_w &= \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1/2} \boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\
\boldsymbol{v}_w &= \frac{1}{\sqrt{N}} \boldsymbol{\Sigma}^{-1/2}(N_1\bar{\boldsymbol{x}}_1 + N_2\bar{\boldsymbol{x}}_2 - N_1\boldsymbol{\mu}_1 - N_2\boldsymbol{\mu}_2), \\
\boldsymbol{B} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}.
\end{aligned}$$

Then $\boldsymbol{u}_w$, $\boldsymbol{v}_w$ and $\boldsymbol{B}$ are independent. In addition, $\boldsymbol{u}_w \sim N_p((1/N_1 + 1/N_2)^{-1/2}\boldsymbol{\delta}, \boldsymbol{I}_p)$ and $\boldsymbol{v}_w \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, where $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. It also holds that $n\boldsymbol{B}$ is distributed as a Wishart distribution with n degrees of freedom and covariance matrix $\boldsymbol{I}_p$, which is denoted as $W_p(n, \boldsymbol{I}_p)$. Substituting them, we have

$$\begin{aligned}
U_w &= -\frac{1}{2}\left(\frac{n}{N_2} - \frac{n}{N_1}\right)\frac{\boldsymbol{u}_w'\boldsymbol{B}^{-1}\boldsymbol{u}_w}{n} + \frac{n}{\sqrt{N_1N_2}}\frac{\boldsymbol{u}_w'\boldsymbol{B}^{-1}\boldsymbol{v}_w}{n} - \sqrt{\frac{nN_2}{NN_1}}\frac{\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}_w}{\sqrt{n}}, \\
V_w &= \frac{Nn}{N_1N_2}\frac{\boldsymbol{u}_w'\boldsymbol{B}^{-2}\boldsymbol{u}_w}{n}.
\end{aligned}$$

On the other hand, for $\boldsymbol{x} \in \Pi_1$, we can express $Z(\boldsymbol{x})$ as

$$\begin{aligned}
Z(\boldsymbol{x}) &= \frac{1}{2}(1 + N_2^{-1})^{-1}\left\{a^{1/2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1) + (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)\right\}\boldsymbol{S}^{-1}\left\{-a^{1/2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1) + (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)\right\} \\
&= \frac{1}{2}(1 + N_2^{-1})^{-1}\omega_1\omega_2\boldsymbol{u}_z'\boldsymbol{B}^{-1}\boldsymbol{t},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{u}_z &= \omega_1^{-1}\boldsymbol{\Sigma}^{-1/2}\left\{-a^{1/2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1) + (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)\right\}, \\
\boldsymbol{t} &= \omega_2^{-1}\boldsymbol{\Sigma}^{-1/2}\left\{a^{1/2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1) + (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)\right\}, \\
\omega_1^2 &= 2\left\{(1 + N_2^{-1}) - a^{1/2}\right\}, \\
\omega_2^2 &= 2\left\{(1 + N_2^{-1}) + a^{1/2}\right\}, \\
a &= \frac{1 + N_2^{-1}}{1 + N_1^{-1}}.
\end{aligned}$$

Note that $\omega_1^2 = O(p^{-1})$ and $\omega_2^2 \to 4$ under A1. The independency of $\boldsymbol{u}_z$ and $\boldsymbol{t}$ and these distributional results can be derived by using the following general result (Lemma 1) for linear combinations of i.i.d. random vectors (see, e.g., Anderson [3]; Theorem 3.3.1).

Lemma 1. *Suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are independent, and $\boldsymbol{x}_i$ is distributed as $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$. Let $\boldsymbol{H} = (h_{ij})$ be an $N \times N$ orthogonal matrix. Then $\boldsymbol{y}_i = \sum_{j=1}^N h_{ij}\boldsymbol{x}_i$ is distributed as $N_p(\nu_i, \boldsymbol{\Sigma})$, where $\nu_i = \sum_{j=1}^N h_{ij}\mu_j$, $i = 1, \ldots, N$, and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ are independent.*

From Lemma 1, we have that $\boldsymbol{u}_z$ and $\boldsymbol{t}$ are independent, and

$$\boldsymbol{u}_z \sim N_p(\omega_1^{-1}\boldsymbol{\delta}, \boldsymbol{I}_p), \ \boldsymbol{t} \sim N_p(\omega_2^{-1}\boldsymbol{\delta}, \boldsymbol{I}_p),$$

where $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Let

$$\boldsymbol{v}_z = \boldsymbol{t} - \omega_2^{-1}\boldsymbol{\delta},$$

which is distributed as $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Now we shall see that $Z(x)$ can be expressed as a location of scale mixture of the standard normal distribution. Note that

$$\begin{aligned}\boldsymbol{u}_z'\boldsymbol{B}^{-1}\boldsymbol{t} &= \boldsymbol{u}_z'\boldsymbol{B}^{-1}\boldsymbol{v}_z + \boldsymbol{u}_z'\boldsymbol{B}^{-1}\omega_2^{-1}\boldsymbol{\delta} \\ &= \sqrt{\boldsymbol{u}_z'\boldsymbol{B}^{-2}\boldsymbol{u}_z}Z_0 + \omega_2^{-1}\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}_z',\end{aligned}$$

where

$$Z_0 = \frac{1}{\sqrt{\boldsymbol{u}_z'\boldsymbol{B}^{-2}\boldsymbol{u}_z}}(\boldsymbol{u}_z'\boldsymbol{B}^{-1})\boldsymbol{v}_z.$$

The conditional distribution of Z_0 when $\boldsymbol{u}_z$ and $\boldsymbol{S}$ are given is the standard normal distribution. Since it does not depend on $\boldsymbol{u}_z$ and $\boldsymbol{S}$, Z_0 is distributed as the standard normal distribution, and is independent of $\boldsymbol{u}_z$ and $\boldsymbol{S}$. Therefore, $Z(\boldsymbol{x})$ is a location and scale mixture of the standard normal distribution. Modifying the sale and the location, we use the following location and scale mixture expression for $Z(\boldsymbol{x})$:

$$Z(\boldsymbol{x}) = \frac{1}{2}(1+N_2^{-1})^{-1}\left(\frac{Nn}{N_1N_2}\right)^{-1/2}(\sqrt{n}\omega_1)\omega_2(V_z^{1/2}Z_z - U_z),$$

where

$$\begin{aligned}U_z &= -\left(\frac{Nn}{N_1N_2}\right)^{1/2}\omega_2^{-1}\frac{\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}_z}{\sqrt{n}}, \\ V_z &= \frac{Nn}{N_1N_2}\frac{\boldsymbol{u}_z'\boldsymbol{B}^{-2}\boldsymbol{u}_z}{n}, \\ Z_z &= V_z^{-1/2}\left(\frac{Nn}{N_1N_2}\right)^{1/2}\frac{\boldsymbol{u}_z'\boldsymbol{B}^{-1}\boldsymbol{v}_z}{\sqrt{n}}.\end{aligned}$$

Note that

$$\frac{1}{2}(1+N_2^{-1})^{-1}\omega_1\omega_2 = \sqrt{1 - \frac{1}{(1+N_1^{-1})(1+N_2^{-1})}} = \sqrt{\frac{N_1^{-1}+N_2^{-1}+N_1^{-1}N_2^{-1}}{(1+N_1^{-1})(1+N_2^{-1})}}.$$

So,

$$Z(\boldsymbol{x}) = \sqrt{\frac{1+N^{-1}}{(1+N_1^{-1})(1+N_2^{-1})}}(V_z^{1/2}Z_z - U_z) = c^*(V_z^{1/2}Z_z - U_z).$$

Here, the variable $V_z^{1/2}Z_z - U_z$ is a location and scale mixture of the standard normal distribution. Further,

$$Z(\boldsymbol{x}) > c^* \Leftrightarrow V_z^{1/2}Z_z - U_z > c.$$

We have seen that the discriminant function $W(\boldsymbol{x})$ based on a cutoff point c and the discriminant function $Z(\boldsymbol{x})$ based on a cutoff point c^* can be expressed as a location (U) and scale ($V^{1/2}$) mixture of the standard normal distribution, and so these misclassification probabilities when $\boldsymbol{x} \in \Pi_1$ can be expressed as

$$E[\Phi\left\{V^{-1/2}(U+c)\right\}], \tag{3}$$

where Φ is the distribution function of the standard normal distribution. In order to treat for W and Z in a unified way, we define two random variables U and V as follows: for $\boldsymbol{x} \in \Pi_1$,

$$U = \rho_1 \frac{\boldsymbol{u}'\boldsymbol{B}^{-1}\boldsymbol{u}}{n} + \rho_2 \frac{\boldsymbol{v}'\boldsymbol{B}^{-1}\boldsymbol{u}}{n} - \rho_3 \frac{\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}}{\sqrt{n}}, \quad V = \tau^2 \frac{\boldsymbol{u}'\boldsymbol{B}^{-2}\boldsymbol{u}}{n}, \tag{4}$$

where $\boldsymbol{u} \sim N_p(\sqrt{n}\omega\boldsymbol{\delta}, \boldsymbol{I}_p)$, $\boldsymbol{v} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $n\boldsymbol{B} \sim W_p(n, \boldsymbol{I}_p)$, $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, and $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{B}$ are independent. Here, $\rho_i = \rho_i(N_1, N_2)$, $i = 1, 2, 3$, $\tau = \tau(N_1, N_2) \neq 0$, and $\omega = \omega(N_1, N_2) \neq 0$ are constants which are $O(1)$ under A1. In addition, $\omega^{-2} = Nn/(N_1N_2) + O(n^{-1})$. The above results are summarized as in the following Lemma.

Lemma 2. *Assume that $\boldsymbol{x} \in \Pi_1$. Let (U, V) be the random variables as in (4), and let Z be the standard normal random variable which is independent of (U, V). Then*

$$W(\boldsymbol{x}) \overset{\mathcal{D}}{=} V^{1/2}Z - U,$$

where (U, V) in (4) is defined with the following ρ_1, ρ_2, ρ_3, τ and ω:

$$\rho_1 = \frac{1}{2}\left(\frac{n}{N_1} - \frac{n}{N_2}\right), \quad \rho_2 = \frac{n}{\sqrt{N_1N_2}}, \quad \rho_3 = \sqrt{\frac{nN_2}{NN_1}},$$
$$\tau = \sqrt{\frac{Nn}{N_1N_2}}, \quad \omega = \left(\frac{n}{N_1} + \frac{n}{N_2}\right)^{-1/2} = \sqrt{\frac{N_1N_2}{Nn}}. \tag{5}$$

Similarly,

$$(1/c^*)Z(\boldsymbol{x}) \overset{\mathcal{D}}{=} V^{1/2}Z - U,$$

where $c^* = \left[(1+N^{-1})/\{(1+N_1^{-1})(1+N_2^{-1})\}\right]^{1/2}$, (U, V) in (4) is defined with the following ρ_1, ρ_2, ρ_3, τ and ω:

$$\rho_1 = 0, \quad \rho_2 = 0, \quad \rho_3 = \sqrt{\frac{Nn}{N_1N_2}}\omega_2^{-1}, \quad \tau = \sqrt{\frac{Nn}{N_1N_2}}, \quad \omega = (n\omega_1^2)^{-1/2},$$
$$\omega_1^2 = 2\left\{(1+N_2^{-1}) - a^{1/2}\right\}, \quad \omega_2^2 = 2\left\{(1+N_2^{-1}) + a^{1/2}\right\}, \quad a = \frac{1+N_2^{-1}}{1+N_1^{-1}}. \tag{6}$$

In order to evaluate the expectation as in (3), it is important to express $\boldsymbol{u}'\boldsymbol{B}^{-1}\boldsymbol{u}$, $\boldsymbol{v}'\boldsymbol{B}^{-1}\boldsymbol{u}$, $\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}$, and $\boldsymbol{u}'\boldsymbol{B}^{-2}\boldsymbol{u}$ in terms of simple variables whose moments are computable. We use the following lemma given by Yamada et al. [23] which expresses them as functions of the independent standard normal and chi-squared variables.

Lemma 3. *Let $\boldsymbol{v}_1 \sim N_p(\boldsymbol{\delta}, \boldsymbol{I}_p)$, $\boldsymbol{v}_2 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $\boldsymbol{A} \sim W_p(n, \boldsymbol{I}_p)$, and $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{A}$ are independent. Then*

$$\begin{pmatrix} \boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_2'\boldsymbol{A}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_1'\boldsymbol{A}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_1'\boldsymbol{A}^{-2}\boldsymbol{v}_1 \end{pmatrix} \overset{\mathcal{D}}{=} \begin{pmatrix} \dfrac{\Delta}{Y_1}\left(Z_1 + \Delta - \sqrt{\dfrac{Y_2}{Y_3}} Z_2\right) \\ \sqrt{\dfrac{1}{Y_1^2}\left(1 + \dfrac{Y_2}{Y_3}\right)\{(Z_1+\Delta)^2 + Z_2^2 + Y_4\}} Z_3 \\ \dfrac{1}{Y_1}\{(Z_1+\Delta)^2 + Z_2^2 + Y_4\} \\ \dfrac{1}{Y_1^2}\left(1 + \dfrac{Y_2}{Y_3}\right)\{(Z_1+\Delta)^2 + Z_2^2 + Y_4\} \end{pmatrix},$$

where $\Delta = \sqrt{\boldsymbol{\delta}'\boldsymbol{\delta}}$; $Z_i \sim N(0,1)$, $i = 1,2,3$; $Y_i \sim \chi^2_{f_i}$, $i = 1,2,3,4$; all the seven variables Z_1, Z_2, Z_3, Y_1, Y_2, Y_3, Y_4 are independent;

$$f_1 = n - p + 1,\ \ f_2 = p - 1,\ \ f_3 = n - p + 2,\ \ f_4 = p - 2.$$

Results which are similar to Lemma 3 have benn given in the following papers. Fujikoshi and Seo [9]; Lemma 2.2 gave stochastic expression for triplet of $\boldsymbol{v}_1'\boldsymbol{A}^{-1}\boldsymbol{v}_1$, $\boldsymbol{v}_1'\boldsymbol{A}^{-1}\boldsymbol{v}_2$ and $\boldsymbol{v}_2'\boldsymbol{A}^{-1}\boldsymbol{v}_2$, and Fujikoshi [7]; Lemma 4.1 gave for triplet of $\boldsymbol{v}_1'\boldsymbol{A}^{-1}\boldsymbol{v}_1$, $\boldsymbol{v}_1'\boldsymbol{A}^{-2}\boldsymbol{v}_1$ and $\boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{v}_1$. Hyodo and Kubokawa [11] has also given a different expression for the four statistics in Lemma 3. However, their expression does not hold simultaneously. In fact, they have used the same expression as Lemma 4.1 in Fujikoshi [7] for the triplet of $\boldsymbol{v}_1'\boldsymbol{A}^{-1}\boldsymbol{v}_1$, $\boldsymbol{v}_1'\boldsymbol{A}^{-2}\boldsymbol{v}_1$ and $\boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{v}_1$, and added an expression for $\boldsymbol{v}_2'\boldsymbol{A}^{-1}\boldsymbol{v}_1$, which was derived separately from the triplet.

From Lemma 3, we can write the U and V as in (4) as follows:

$$\begin{pmatrix} U \\ V \end{pmatrix} \overset{\mathcal{D}}{=} \begin{pmatrix} u(Y_1/f_1, Y_2/f_2, Y_3/f_3, Y_4/f_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}) \\ v(Y_1/f_1, Y_2/f_2, Y_3/f_3, Y_4/f_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}) \end{pmatrix},$$

where

$$\begin{aligned} u(y_1, y_2, y_3, y_4, z_1, z_2, z_3) &= u_1(y_1, y_4, z_1, z_2) + u_2(y_1, y_2, y_3, y_4, z_1, z_2, z_3) \\ &\quad - u_3(y_1, y_2, y_3, y_4, z_1, z_2), \qquad (7) \\ u_1(y_1, y_4, z_1, z_2) &= \frac{a_1}{y_1}\{(z_1 + \omega\Delta)^2 + z_2^2 + a_4 y_4\}, \\ u_2(y_1, y_2, y_3, y_4, z_1, z_2, z_3) &= \frac{a_2}{y_1}\sqrt{\left(1 + a_5^2\frac{y_2}{y_3}\right)\{(z_1 + \omega\Delta)^2 + z_2^2 + a_4 y_4\}} z_3, \\ u_3(y_1, y_2, y_3, y_4, z_1, z_2) &= a_3\frac{\Delta}{y_1}\left(z_1 + \omega\Delta - a_5\sqrt{\frac{y_2}{y_3}} z_2\right), \\ v(y_1, y_2, y_3, y_4, z_1, z_2, z_3) &= \frac{a_6^2}{y_1^2}\left(1 + a_5^2\frac{y_2}{y_3}\right)\{(z_1 + \omega\Delta)^2 + z_2^2 + a_4 y_4\}. \qquad (8) \end{aligned}$$

Here,

$$a_i = \frac{n}{f_1}\rho_i \quad (i = 1,2,3), \quad a_4 = \frac{f_4}{n}, \quad a_5 = \sqrt{\frac{f_2}{f_3}}, \quad a_6 = \frac{n}{f_1}\tau.$$

Note that $a_i = O(1)$ under A1.

3 Asymptotic expansions for the EPMCs of $W(\boldsymbol{x})$ and $Z(\boldsymbol{x})$ under A1

In order to obtain asymptotic expansions for the EPMCs of $W(\boldsymbol{x})$ and $Z(\boldsymbol{x})$, we may derive an asymptotic expansion of $P(\sqrt{V}Z - U < c)$ under A1. Further, instead of the distribution of $\sqrt{V}Z - U$, we consider its standardized version defined by

$$T = \sqrt{\frac{V}{v_0}}Z - \left(\frac{U - u_0}{\sqrt{v_0}}\right),$$

where

$$\begin{aligned} u_0 &= u(E(Y_1/f_1), E(Y_2/f_2), E(Y_3/f_3), E(Y_4/f_4), E(Z_1/\sqrt{n}), E(z_2/\sqrt{n}), E(Z_3/\sqrt{n})) \\ &= u(1,1,1,1,0,0,0), \\ v_0 &= v(E(Y_1/f_1), E(Y_2/f_2), E(Y_3/f_3), E(Y_4/f_4), E(Z_1/\sqrt{n}), E(z_2/\sqrt{n}), E(Z_3/\sqrt{n})) \\ &= v(1,1,1,1,0,0,0). \end{aligned}$$

Then, it holds that

$$P(\sqrt{V}Z - U \leq x) = P(T \leq v_0^{-1/2}(x + u_0)).$$

Now we consider an asymptotic expansion of the distribution of T by expanding its characteristic function

$$C(t) = E\left\{\exp(itT)\right\}.$$

Based on the fact that T is conditionally normal when (U, V) is given, the conditional characteristic function can be expressed as

$$\Psi(y_1, y_2, y_3, y_4, z_1, z_2, z_3) = \exp(it\mu - t^2\sigma^2/2),$$

where

$$\begin{aligned} \mu = \mu(y_1, y_2, y_3, y_4, z_1, z_2, z_3) &= -\frac{u(y_1, y_2, y_3, y_4, z_1, z_2, z_3) - u_0}{\sqrt{v_0}}, \\ \sigma^2 = \sigma^2(y_1, y_2, y_3, y_4, z_1, z_2, z_3) &= \frac{v(y_1, y_2, y_3, y_4, z_1, z_2, z_3)}{v_0}. \end{aligned}$$

Therefor we have

$$C(t) = E\left[\Psi\left(\frac{Y_1}{f_1}, \frac{Y_2}{f_2}, \frac{Y_3}{f_3}, \frac{Y_4}{f_4}, \frac{Z_1}{\sqrt{n}}, \frac{Z_2}{\sqrt{n}}, \frac{Z_3}{\sqrt{n}}\right)\right].$$

To get an asymptotic expansion of $C(t)$, we use a powerful method known as the method by the differential operator which was used by James [12], Okamoto [19], etc. Since the function $\Psi(\cdot)$ is analytic about the point $(Y_1/f_1, Y_2/f_2, Y_3/f_3, Y_4/f_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}) = (1,1,1,1,0,0,0)$, we can expand it a Taylor series as follows.

$$\begin{aligned} &\Psi\left(\frac{Y_1}{f_1}, \frac{Y_2}{f_2}, \frac{Y_3}{f_3}, \frac{Y_4}{f_4}, \frac{Z_1}{\sqrt{n}}, \frac{Z_2}{\sqrt{n}}, \frac{Z_3}{\sqrt{n}}\right) \\ &= \exp(\mathcal{A})\Psi(y_1, y_2, y_3, y_4, z_1, z_2, z_3) \mid_0, \end{aligned} \tag{9}$$

where

$$\mathcal{A} = \sum_{i=1}^{4} \left(\frac{Y_i}{f_i} - 1 \right) \frac{\partial}{\partial y_i} + \sum_{i=1}^{3} \frac{Z_i}{\sqrt{n}} \frac{\partial}{\partial z_i},$$

and the notation $|_0$ stands for the value at the point that $(y_1, y_2, y_3, y_4, z_1, z_2, z_3) = (1, 1, 1, 1, 0, 0, 0)$. Then,

$$C(t) = \Theta\Psi(y_1, y_2, y_3, y_4, z_1, z_2, z_3) \ |_0, \tag{10}$$

where

$$\Theta = E[\exp(\mathcal{A})].$$

Note that Y_1, Y_2, Y_3, Y_4, Z_1, Z_2, and Z_3 are independent, and $Y_i \sim \chi^2_{f_i}$, $i = 1, 2, 3, 4$, $Z_i \sim N(0, 1)$, $i = 1, 2, 3$. Considering the expectation with respect to Y_i's and Z_i's, we have

$$\begin{aligned}
\Theta &= \exp\left\{ -\sum_{i=1}^{4} \frac{\partial}{\partial y_i} - \frac{1}{2} \sum_{i=1}^{4} f_i \log\left(1 - \frac{2}{f_i} \frac{\partial}{\partial y_i}\right) + \frac{1}{2n} \sum_{i=1}^{3} \frac{\partial^2}{\partial z_i^2} \right\} \\
&= \exp\left(\sum_{i=1}^{4} \frac{1}{f_i} \frac{\partial^2}{\partial y_i^2} + \frac{1}{2n} \sum_{i=1}^{3} \frac{\partial^2}{\partial z_i^2} + R_1 \right) \\
&= 1 + \sum_{i=1}^{4} \frac{1}{f_i} \frac{\partial^2}{\partial y_i^2} + \frac{1}{2n} \sum_{i=1}^{3} \frac{\partial^2}{\partial z_i^2} + R_2,
\end{aligned} \tag{11}$$

where R_1 and R_2 are remainder terms which are $O(n^{-2})$ under A1. Substituting (11) into (10), we have

$$C(t) = e^{-t^2/2} \left(1 + \frac{1}{n} \sum_{k=1}^{4} b_k (it)^k + R_3 \right),$$

where R_3 is a remainder term which has the same property as R_1. Inverting the above expansion of the characteristic function, we can obtain an asymptotic expansion of the distribution of T up to the order $O(n^{-1})$ under A1 which is given as the following theorem.

Theorem 1. *Assume that $\boldsymbol{x} \in \Pi_1$. Let (U, V) be the random variables defined as* (4)*, and let Z be a standard normal random variable which is independent of (U, V). Let $y = v_0^{-1/2}(x + u_0)$, where*

$$\begin{aligned}
u_0 &= \frac{n}{m+1} \left\{ (\rho_1 \omega^2 - \rho_3 \omega) \Delta^2 + \frac{p-2}{n} \rho_1 \right\}, \\
v_0 &= \frac{n^2(n+1)}{(m+1)^2(m+2)} \frac{Nn}{N_1 N_2} \left(\omega^2 \Delta^2 + \frac{p-2}{n} \right).
\end{aligned}$$

Then it holds that

$$P(V^{-1/2} Z - U \leq x) = \Phi(y) - \frac{1}{n} \sum_{k=1}^{4} b_k H_{k-1}(y) \phi(y) + O(n^{-2})$$

under A1*, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, $\phi(\cdot)$ is the derivative of $\Phi(\cdot)$, and $H_k(x)$ denotes the Hermite polynomial of degree k, especially, $H_0(x) =$*

$1,\ H_1(x) = x,\ H_2(x) = x^2 - 1,\ H_3(x) = x^3 - 3x,\ H_4(x) = x^4 - 6x^2 + 3.$ *Here,*

$$
\begin{aligned}
b_1 &= -\frac{2n}{m+1}\frac{1}{\sqrt{v_0}}(u_0 + \rho_1),\\
b_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{1}{\frac{p-2}{n} + \omega^2\Delta^2} + \frac{\rho_2^2}{2\tau^2} + \frac{1}{2v_0}\left\{\frac{2n}{m+1}u_0^2\right.\\
&\quad \left. + \frac{2n(p-2)}{(m+1)^2}\rho_1^2 + \frac{n^2}{(m+1)^2}(2\rho_1\omega - \rho_3)^2\Delta^2 + \frac{n^2(p-1)}{(m+1)^2(m+2)}\rho_3^2\Delta^2\right\},\\
b_3 &= -\frac{1}{\sqrt{v_0}}\left\{\frac{2n}{m+1}u_0 + \frac{u_0 + \frac{n}{m+1}\rho_1\omega^2\Delta^2}{\frac{p-2}{n} + \omega^2\Delta^2}\right\},\\
b_4 &= \frac{n}{m+1} + \frac{n(p-1)}{4(m+2)(n+1)} + \frac{1}{4}\frac{1}{\frac{p-2}{n} + \omega^2\Delta^2}\left(1 + \frac{\omega^2\Delta^2}{\frac{p-2}{n} + \omega^2\Delta^2}\right).
\end{aligned}
$$

Corollary 2. *Let $g_w(c; N_1, N_2, \Delta^2)$ be the expected probability of misclassification of W- rule with cutoff point c when $\boldsymbol{x} \in \Pi_1$. Let $y_w = v_w^{-1/2}(c + u_w)$, where*

$$
\begin{aligned}
u_w = u_w(N_1, N_2, \Delta^2) &= -\frac{1}{2}\frac{n}{m+1}\left\{\Delta^2 - \left(\frac{p-2}{N_1} - \frac{p-2}{N_2}\right)\right\},\\
v_w = v_w(N_1, N_2, \Delta^2) &= \frac{n^2(n+1)}{(m+1)^2(m+2)}\left(\Delta^2 + \frac{N(p-2)}{N_1N_2}\right).
\end{aligned}
$$

Then, it holds that under A1,

$$
g_w(c; N_1, N_2, \Delta^2) = \Phi(y_w) - \frac{1}{n}\sum_{k=1}^{4}\ell_k H_{k-1}(y_w)\phi(y_w) + O(n^{-2}),
$$

where $\ell_k = \ell_k(N_1, N_2, \Delta^2)$ for $k = 1, 2, 3, 4$ are given as follows.

$$
\begin{aligned}
\ell_1 &= -\frac{2}{\sqrt{v_w}}\frac{n}{m+1}\left\{u_w + \frac{1}{2}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)\right\},\\
\ell_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{\frac{Nn}{N_1N_2}}{\Delta^2 + \frac{N(p-2)}{N_1N_2}} + \frac{n}{2N} + \frac{1}{2v_w}\left[\frac{2n}{m+1}u_w^2\right.\\
&\quad + \frac{n(p-2)}{2(m+1)^2}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 + \frac{n^2}{(m+1)^2}\frac{N_1n}{NN_2}\Delta^2\\
&\quad \left. + \frac{n^2(p-1)}{(m+1)^2(m+2)}\frac{N_2n}{NN_1}\Delta^2\right],\\
\ell_3 &= -\frac{1}{\sqrt{v_w}}\left\{\frac{2n}{m+1}u_w + \frac{Nn}{N_1N_2}\frac{u_w + \frac{1}{2}\frac{n}{m+1}\frac{N_2 - N_1}{N}\Delta^2}{\Delta^2 + \frac{N(p-2)}{N_1N_2}}\right\},\\
\ell_4 &= \frac{n}{m+1} + \frac{n(p-1)}{4(m+2)(n+1)} + \frac{1}{4}\frac{\frac{Nn}{N_1N_2}}{\Delta^2 + \frac{N(p-2)}{N_1N_2}}\left(1 + \frac{\Delta^2}{\Delta^2 + \frac{N(p-2)}{N_1N_2}}\right).
\end{aligned}
$$

Corollary 3. *Let $g_z(c^*; N_1, N_2, \Delta^2)$ be the expected probability of misclassification of Z- rule with a cutoff point c^* when $\boldsymbol{x} \in \Pi_1$, where $c^* = \sqrt{(1 + N^{-1})/\{(1 + N_1^{-1})(1 + N_2^{-1})\}}c$. Let $y_z = v_z^{-1/2}(c + u_z)$, where*

$$
\begin{aligned}
u_z = u_z(N_1, N_2, \Delta^2) &= -\frac{n}{m+1}\sqrt{\frac{N}{N_1N_2}}\omega_1^{-1}\omega_2^{-1}\Delta^2,\\
v_z = v_z(N_1, N_2, \Delta^2) &= \frac{n^2(n+1)}{(m+1)^2(m+2)}\frac{N}{N_1N_2}\omega_1^{-2}\left\{\Delta^2 + (p-2)\omega_1^2\right\},
\end{aligned}
$$

and

$$\omega_1 = \omega_1(N_1, N_2) = \sqrt{2\{(1+N_2^{-1}) - a^{1/2}\}},$$
$$\omega_2 = \omega_2(N_1, N_2) = \sqrt{2\{(1+N_2^{-1}) + a^{1/2}\}},$$
$$a = a(N_1, N_2) = \frac{1+N_2^{-1}}{1+N_1^{-1}}.$$

Then, it holds that under A1*,*

$$g_z(c^*; N_1, N_2, \Delta^2) = \Phi(y_z) - \frac{1}{n}\sum_{k=1}^{4} \zeta_k H_{k-1}(y_z)\phi(y_z) + O(n^{-2}),$$

where $\zeta_k = \zeta_k(N_1, N_2, \Delta^2)$ *for* $k = 1, 2, 3, 4$ *are given as follows.*

$$\begin{aligned}
\zeta_1 &= -\frac{2n}{m+1}\frac{u_z}{\sqrt{v_z}}, \\
\zeta_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{n\omega_1^2}{\Delta^2 + (p-2)\omega_1^2} \\
&\quad + \frac{1}{2v_z}\left\{\frac{2n}{m+1}u_z^2 + \frac{n^2}{(m+1)^2}\frac{Nn}{N_1N_2}\omega_2^{-2}\Delta^2 + \frac{n^2(p-1)}{(m+1)^2(m+2)}\frac{Nn}{N_1N_2}\omega_2^{-2}\Delta^2\right\}, \\
\zeta_3 &= -\frac{1}{\sqrt{v_z}}\left\{\frac{2n}{m+1}u_z + \frac{n\omega_1^2 u_z}{\Delta^2 + (p-2)\omega_1^2}\right\}, \\
\zeta_4 &= \frac{n}{m+1} + \frac{n(p-1)}{4(m+2)(n+1)} + \frac{1}{4}\frac{n\omega_1^2}{\Delta^2 + (p-2)\omega_1^2}\left\{1 + \frac{\Delta^2}{\Delta^2 + (p-2)\omega_1^2}\right\}.
\end{aligned}$$

There are some results on asymptotic results on the EPMC of W- and Z- rules under A1, see, e.g., Seo and Fujikoshi [9], Fujikoshi [6], Matsumoto [14], Hyodo and Kubokawa [11], etc. It may be noted that our results have been given a unified way for W- and Z- rules, and so that their comparison becomes more easy. In fact, in the next section, using Corollaries 2 and 3 we show that Z-rule has an optimality in the comparison with W-rule.

4 Comparison of EERs

Let π_i be the prior probabilities of $\boldsymbol{x}$ drown from Π_i for $i = 1, 2$. Then, the expected error rate (EER) for W-rule with a cutoff point c_w is expressed as

$$\mathrm{EER}_w(c_w) = \pi_1 P(W(\boldsymbol{x}) < c_w | \boldsymbol{x} \in \Pi_1) + \pi_2 P(W(\boldsymbol{x}) > c_w | \boldsymbol{x} \in \Pi_2).$$

From Corollary 2, the limit under A1 is given as

$$\begin{aligned}
&\lim_{\mathrm{A1}} \mathrm{EER}_w(c_w) \\
&= \pi_1 \Phi\left(-\frac{1}{2}\frac{-\tilde{c}_w + \Delta_0^2 + (\gamma - \gamma^{-1})\gamma_0}{\sqrt{\Delta_0^2 + (\gamma + \gamma^{-1} + 2)\gamma_0}}\sqrt{1-\gamma_0}\right) + \pi_2 \Phi\left(-\frac{1}{2}\frac{\tilde{c}_w + \Delta_0^2 + (\gamma^{-1} - \gamma)\gamma_0}{\sqrt{\Delta_0^2 + (\gamma^{-1} + \gamma + 2)\gamma_0}}\sqrt{1-\gamma_0}\right),
\end{aligned}$$

where $\tilde{c}_w = 2(1-\gamma_0)c_w$, $\Delta_0^2 = \lim_{p\to\infty} \Delta^2$, $\lim N_1/N_2 = \gamma$ and $\lim p/n = \gamma_0$. The minimum value with respect to $\tilde{c}_w$ or equivalently c_w is attained at

$$c_w = c_{w,0} = \frac{1}{2(1-\gamma_0)}\left[(\gamma - \gamma^{-1})\gamma_0 + \frac{2}{1-\gamma_0}\left\{1 + \frac{(\gamma + \gamma^{-1} + 2)\gamma_0}{\Delta_0^2}\right\}\log\frac{\pi_2}{\pi_1}\right].$$

This result was pointed by Hyodo and Kubokawa [11], and they studied asymptotic unbiased estimator for $\mathrm{EER}_w(c_{w,0})$.

From the above result we can see that the limiting $\mathrm{EER}_w(c_w)$ takes the minimum value at

$$c_{w,m} = c_{w,m} = \frac{1}{2}\frac{N}{N-p}\left[\frac{p}{N_2} - \frac{p}{N_1} + 2\frac{N}{N-p}\left(1 + \frac{Np}{N_1 N_2}\frac{1}{\Delta^2}\right)\log\frac{\pi_2}{\pi_1}\right]. \tag{12}$$

For the case when the prior probabilities are equal,

$$c_{w,m} = \frac{1}{2}\frac{N}{N-p}\left(\frac{p}{N_2} - \frac{p}{N_1}\right),$$

and then,

$$\begin{aligned}
&\lim_{\mathrm{A}_1}\left[\frac{1}{2}P(W(\boldsymbol{x}) < c_{w,m}|\boldsymbol{x} \in \Pi_1) + \frac{1}{2}P(W(\boldsymbol{x}) > c_{w,m}|\boldsymbol{x} \in \Pi_2)\right] \\
&= \Phi\left(-\frac{1}{2}\frac{\Delta_0^2}{\sqrt{\Delta_0^2 + (\gamma + \gamma^{-1} + 2)\gamma_0}}\sqrt{1-\gamma_0}\right).
\end{aligned} \tag{13}$$

On the other hand, the EER for Z-rule with a cutoff point c_z is expressed as

$$\mathrm{EER}_z(c_z) = \pi_1 P(Z(\boldsymbol{x}) < c_z|\boldsymbol{x} \in \Pi_1) + \pi_2 P(Z(\boldsymbol{x}) > c_z|\boldsymbol{x} \in \Pi_2).$$

Let

$$c_z^* = \sqrt{\frac{1 + N^{-1}}{(1 + N_1^{-1})(1 + N_2^{-1})}}c_z.$$

Using Corollary 3,

$$\begin{aligned}
&\lim_{\mathrm{A1}}\mathrm{EER}_z(c_z^*) \\
&= \pi_1\Phi\left(-\frac{1}{2}\frac{-c_z^* + \Delta_0^2}{\sqrt{\Delta_0^2 + (\gamma + \gamma^{-1} + 2)\gamma_0}}\sqrt{1-\gamma_0}\right) + \pi_2\Phi\left(-\frac{1}{2}\frac{c_z^* + \Delta_0^2}{\sqrt{\Delta_0^2 + (\gamma^{-1} + \gamma + 2)\gamma_0}}\sqrt{1-\gamma_0}\right),
\end{aligned}$$

where $c_z^* = 2(1-\gamma_0)\lim_{\mathrm{A1}} c_z^* = 2(1-\gamma_0)c_z$. The minimum value with respect to c_z^* is attained at

$$c_z^* = c_{z,0}^* = \frac{2}{1-\gamma_0}\left\{1 + \frac{(\gamma + \gamma^{-1} + 2)\gamma_0}{\Delta_0^2}\right\}\log\frac{\pi_2}{\pi_1}.$$

This implies that the limiting EER for Z-rule, i.e., $\lim_{\mathrm{A1}}\mathrm{EER}_z(c_z^*)$ takes the minimum value at

$$c_z = c_{z,m} = \left\{\frac{1 + N^{-1}}{(1 + N_1^{-1})(1 + N_2^{-1})}\right\}^{-1/2}\left(\frac{N}{N-p}\right)^2\left(1 + \frac{Np}{N_1 N_2}\frac{1}{\Delta^2}\right)\log\frac{\pi_2}{\pi_1}. \tag{14}$$

When the prior probabilities are equal, $c_{z,m} = 0$, and then, the limiting error rate under A1 is the same as (13). These imply that when $\pi_1 = \pi_2$,

$$\lim_{\mathrm{A1}}\mathrm{EER}_w(c_{w,m}) = \lim_{\mathrm{A1}}\mathrm{EER}_z(0),$$

which is equal to the right-hand side of the equality in (13). In order to see the difference when $\pi_1 = \pi_2$, we need to compare the next terms of these asymptotic expansions. The final result is given in the next theorem.

Theorem 4. *Let* $\mathrm{EER}_w(c_w)$ *and* $\mathrm{EER}_z(c_z)$ *be the expected error rates of* W*- rule with a cutoff point* c_w *and* Z*- rule with a cutoff point* c_z*, respectively.*

(1) *The minimums of* $\mathrm{EER}_w(c_w)$ *and* $\mathrm{EER}_z(c_z)$ *are attained at* $c_w = c_{w,m}$ *and* $c_w = c_{z,m}$ *given in* (12) *and* (14)*, respectively.*

(2) *When* $\pi_1 = \pi_2$*, it holds that*

$$\mathrm{EER}_w(c_{w,m}) - \mathrm{EER}_z(c_{z,m}) = -\frac{1}{4v_w}\frac{(n-1)(p-2)}{(m+1)^3}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 H_1(y_c)\phi(y_c) + O(n^{-2}),$$

where

$$y_c = \frac{-(1/2)\Delta^2}{\sqrt{\left\{\Delta^2 + \frac{N(p-2)}{N_1N_2}\right\}\frac{n+1}{m+2}}}.$$

Further, since $H_1(y_c) = y_c < 0$*, we have that* $\mathrm{EER}_z(0)$ *is less than or equal to* $\mathrm{EER}_w(c_{w,m})$*, neglecting the term of* $O(n^{-2})$*. When* $N_1 = N_2$*, the difference becomes* $O(n^{-2})$*.*

The proof of Theorem 4 (2) is given in Appendix A.

We will show an asymptotic expansion for each of $\mathrm{EER}_w(0)$ and $\mathrm{EER}_z(0)$ up to the terms of $O(n^{-1})$ under A0. Asymptotic expansion of $\mathrm{EER}_w(0)$ is obtained by using Corollary 2 in Okamoto [19] (which is cited in Fujikoshi et al. [10], Corollary 9.3.1), which is as follows.

$$\mathrm{EER}_w(0) = \Phi\left(-\frac{1}{2}\Delta\right) + \phi\left(-\frac{1}{2}\Delta\right)\left[\frac{1}{16}\left\{\frac{4(p-1)}{\Delta} + \Delta\right\}\left(\frac{1}{N_1} + \frac{1}{N_2}\right) + \frac{(p-1)\Delta}{4n}\right] + O(n^{-2}).$$

Since $c_{wm} = O(n^{-1})$ under A0, we can show that asymptotic expansion of $\mathrm{EER}_w(c_{wm})$ up to the term of $O(n^{-1})$ is the same as the one of $\mathrm{EER}_w(0)$.

By virtue of COROLLARY in Memon and Okamoto [15] (which is cited in Fujikoshi et al. [10], Corollary 9.3.2), we have

$$\mathrm{EER}_z(0) = \Phi\left(-\frac{1}{2}\Delta\right) + \phi\left(-\frac{1}{2}\Delta\right)\left[\frac{1}{16}\left\{\frac{4(p-1)}{\Delta} + \Delta\right\}\left(\frac{1}{N_1} + \frac{1}{N_2}\right) + \frac{(p-1)\Delta}{4n}\right] + O(n^{-2}).$$

These results imply that $\mathrm{EER}_w(0)$ and $\mathrm{EER}_z(0)$ are the same up to the terms of $O(n^{-1})$ under A0.

Hyodo and Kubokawa [11] proposed a variable selection procedure for W-rule in two-group discriminant analysis for high-dimensional data. Their criteria is based on an estimator of $\mathrm{EER}_w(c_{w,m})$ which is unbiased up to the term of order $O_p(n^{-1})$ under A1. From the above result, we think that Hyodo and Kubokawa [11]'s criteria with being reconsidered for Z-rule outperforms their criteria in terms of selecting the true set of variables. In Section 6 we give an asymptotic estimator of $\mathrm{EER}_w(c_{w,m})$ with errors of $O(N^{-2})$.

5 Estimation for EPMCs

In this section, we consider to estimate the expected probabilities of misclassification; $e_w(2|1)$, $e_w(1|2)$, $e_z(2|1)$ and $e_z(1|2)$. These depend on unknown parameter Δ^2. It is important to estimate Δ^2. A well known unbiased estimator is given by

$$\widehat{\Delta}^2 = \frac{n-p-1}{n} D^2 - \frac{pN}{N_1 N_2}.$$

Here, we denote $\widehat{\Delta^2}$ by $\widehat{\Delta}^2$. Such conventional notation is used, hereafter. Firstly, we give a stochastic expression of D^2, which is essentially stated in Lemma 3.

Lemma 4. *The following equality holds in distribution:*

$$D^2 \overset{\mathcal{D}}{=} \frac{n}{y_1}\left(\Delta^2 + \frac{N}{N_1 N_2} z_1^2 - 2\sqrt{\frac{N}{N_1 N_2}}\Delta z_1 + \frac{N}{N_1 N_2} y_2\right),$$

where $y_1 \sim \chi^2(n-p+1)$, $y_2 \sim \chi^2(p-1)$, $z_1 \sim N(0,1)$, and y_1, y_2, z_1 are independent.

From Lemma 4 it can be seen that $\widehat{\Delta}^2$ has consistency under A1. Therefore, from Corollary 2, a consistent estimator of $g_w(c; N_1, N_2, \Delta^2)$ under A1 is obtained as

$$\Phi\left(\frac{c + u_w(N_1, N_2, \widehat{\Delta}^2)}{\sqrt{v_w(N_1, N_2, \widehat{\Delta}^2)}}\right).$$

Similarly, a consistent estimator of $g_z(c^*; N_1, N_2, \Delta^2)$ is obtained as

$$\Phi\left(\frac{c + u_z(N_1, N_2, \widehat{\Delta}^2)}{\sqrt{v_z(N_1, N_2, \widehat{\Delta}^2)}}\right). \tag{15}$$

However, $v_z(N_1, N_2, \widehat{\Delta}^2)$ does not always take non-negative values, and so it is important to modify the estimator such that it takes always non-negative values. For the purpose, instead of using $\widehat{\Delta}^2$, we use

$$\widehat{\Delta}_A^2 = \frac{n-p+1}{n} D^2 - \frac{p-2}{n}\omega^{-2}.$$

Then, it can be seen that

$$v_w(N_1, N_2, \Delta_A^2) = \frac{n(n+1)}{(m+1)(m+2)} D^2,$$
$$v_z(N_1, N_2, \Delta_A^2) = \frac{n(n+1)}{(m+1)(m+2)} D^2,$$

which take non-negative values. So, we propose a consistent estimator of misclassification probability given as in the following theorem.

Theorem 5. *Assume that* $\lim_{p\to\infty} \Delta^2 > 0$. *Then, under* A1,

$$P(W(\boldsymbol{x}) < c | \boldsymbol{x} \in \Pi_1) - \Phi\left(\frac{c + u_w(N_1, N_2, \widehat{\Delta}_A^2)}{\sqrt{v_w(N_1, N_2, \widehat{\Delta}_A^2)}}\right) \xrightarrow{p} 0,$$
$$P(Z(\boldsymbol{x}) < c^* | \boldsymbol{x} \in \Pi_1) - \Phi\left(\frac{c + u_z(N_1, N_2, \widehat{\Delta}_A^2)}{\sqrt{v_z(N_1, N_2, \widehat{\Delta}_A^2)}}\right) \xrightarrow{p} 0.$$

We extend the result given in Theorem 5 to the one based on asymptotic expansions. From Theorem 1, we can see that y, u_0, v_0, $b_1, \ldots, b_4$ are functions of Δ^2. Define $\widehat{y}$, $\widehat{u}_0$, $\widehat{v}_0$, $\widehat{b}_1, \ldots, \widehat{b}_4$ be the ones obtained from y, u_0, v_0, $b_1, \ldots, b_4$ by replacing Δ^2 with $\widehat{\Delta}^2_A$.

Theorem 6. *Assume that $\boldsymbol{x} \in \Pi_1$, and let U, V and Z be the same ones as in Theorem 1. Then, under* A1,

$$P(V^{-1/2}Z - U \leq x) = E\left[\Phi(\widehat{y}) + \frac{1}{n}\sum_{k=1}^{4}(\widehat{\varepsilon}_k - \widehat{b}_k)H_{k-1}(\widehat{y})\phi(\widehat{y})\right] + O(n^{-2}),$$

where $\widehat{\varepsilon}_k = \varepsilon_k(\widehat{\Delta}^2_A)$ with

$$\begin{aligned}
\varepsilon_1(\Delta^2) &= -\frac{1}{\sqrt{v_0}}\frac{n}{m+1}(\rho_1\omega^2 - \rho_3\omega)(d_2(\Delta^2) + nr_1),\\
\varepsilon_2(\Delta^2) &= \frac{\lambda}{2}(d_2(\Delta^2) + nr_1) + \frac{1}{2}\frac{1}{v_0}\left(\frac{n}{m+1}\right)^2(\rho_1\omega^2 - \rho_3\omega)^2 d_1(\Delta^2),\\
\varepsilon_3(\Delta^2) &= -\frac{\lambda}{2}\frac{1}{\sqrt{v_0}}\frac{n}{m+1}(\rho_1\omega^2 - \rho_3\omega)d_1(\Delta^2),\\
\varepsilon_4(\Delta^2) &= \frac{\lambda^2}{8}d_1(\Delta^2).
\end{aligned}$$

Here,

$$\begin{aligned}
d_1(\Delta^2) &= \frac{2n}{f_1}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}^2 + \frac{4Nn}{N_1N_2}\Delta^2 + \frac{2n}{f_2}\left\{\frac{N(p-1)}{N_1N_2}\right\}^2,\\
d_2(\Delta^2) &= \frac{2n}{f_1}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\} + \frac{Nn}{N_1N_2},\\
\lambda &= \frac{1}{v_0}\frac{n^2(n+1)}{(m+1)^2(m+2)}\frac{Nn}{N_1N_2}\omega^2,\\
r_1 &= \frac{N(p-1)}{N_1N_2} - \frac{p-2}{n}\omega^{-2}.
\end{aligned}$$

The proof of Theorem 6 is given in Appendix B. From Theorem 6, we can obtain estimators of the misclassification probabilities for W-rule and Z-rule, which are unbiased up to the term with the order $O(n^{-1})$ under A1. These results are summarized as the following corollaries.

Corollary 7. *Let $\widehat{y}_w = \widehat{v}_w^{-1/2}(c + \widehat{u}_w)$, where $\widehat{u}_w = u_w(N_1, N_2, \widehat{\Delta}^2_w)$ and $\widehat{v}_w = v_w(N_1, N_2, \widehat{\Delta}^2_w)$ with*

$$\widehat{\Delta}^2_w = \widehat{\Delta}^2_A = \frac{n-p+1}{n}D^2 - \frac{N(p-2)}{N_1N_2}.$$

Then,

$$g_w(c; N_1, N_2, \Delta^2) = E\left[\Phi(\widehat{y}_w) + \frac{1}{n}\sum_{k=1}^{4}(\widehat{\varepsilon}_{w,k} - \widehat{\ell}_k)H_{k-1}(\widehat{y}_w)\phi(\widehat{y}_w)\right] + O(n^{-2}),$$

where $\widehat{\ell}_k = \ell_k(N_1, N_2, \widehat{\Delta}^2_w)$ and $\widehat{\varepsilon}_{w,k} = \varepsilon_{w,k}(\widehat{\Delta}^2_w)$. Here, $\varepsilon_{w,k}(\Delta^2)$ is the same as $\varepsilon_k(\Delta^2)$ given in

Theorem 6, i.e.,

$$\begin{aligned}
\varepsilon_{w,1}(\Delta^2) &= \frac{1}{2\sqrt{v_w}}\frac{n}{m+1}(d_2(\Delta^2)+nr_{w,1}),\\
\varepsilon_{w,2}(\Delta^2) &= \frac{\lambda_w}{2}(d_2(\Delta^2)+nr_{w,1})+\frac{1}{8}\frac{1}{v_w}\left(\frac{n}{m+1}\right)^2 d_1(\Delta^2),\\
\varepsilon_{w,3}(\Delta^2) &= \frac{\lambda_w}{4}\frac{1}{\sqrt{v_w}}\frac{n}{m+1}d_1(\Delta^2),\\
\varepsilon_{w,4}(\Delta^2) &= \frac{\lambda_w^2}{8}d_1(\Delta^2),
\end{aligned}$$

with

$$\lambda_w = \frac{1}{v_w}\frac{n^2(n+1)}{(m+1)^2(m+2)}, \text{ and } r_{w,1} = \frac{N}{N_1N_2}.$$

Corollary 8. *Let $\widehat{y}_z = \widehat{v}_z^{-1/2}(c+\widehat{u_z})$, where $\widehat{u}_z = u_z(N_1, N_2, \widehat{\Delta}_z^2)$ and $\widehat{v}_z = v_z(N_1, N_2, \widehat{\Delta}_z^2)$ with*

$$\widehat{\Delta}_z^2 = \widehat{\Delta}_A^2 = \frac{n-p+1}{n}D^2 - (p-2)\omega_1^2.$$

Then,

$$g_z(c^*; N_1, N_2, \Delta^2) = E\left[\Phi(\widehat{y}_z) + \frac{1}{n}\sum_{k=1}^{4}(\widehat{g}_{z,k} - \widehat{\zeta}_k)H_{k-1}(\widehat{y}_z)\phi(\widehat{y}_z)\right] + O(n^{-2}),$$

where $\widehat{\zeta}_k = \zeta_k(N_1, N_2, \widehat{\Delta}_z^2)$ and $\widehat{\varepsilon}_{z,k} = \varepsilon_{z,k}(\widehat{\Delta}_z^2)$. Here, $\varepsilon_{z,k}(\Delta^2)$ is the same as $\varepsilon_k(\Delta^2)$ given in Theorem 6, i.e.,

$$\begin{aligned}
\varepsilon_{z,1}(\Delta^2) &= \frac{1}{\sqrt{v_z}}\frac{n}{m+1}\sqrt{\frac{N}{N_1N_2}}\omega_1^{-1}\omega_2^{-1}(d_2(\Delta^2)+nr_{z,1}),\\
\varepsilon_{z,2}(\Delta^2) &= \frac{\lambda_z}{2}(d_2(\Delta^2)+nr_1)+\frac{1}{2}\frac{1}{v_z}\left(\frac{n}{m+1}\right)^2\frac{N}{N_1N_2}\omega_1^{-2}\omega_2^{-2}d_1(\Delta^2),\\
\varepsilon_{z,3}(\Delta^2) &= \frac{\lambda_z}{2}\frac{1}{\sqrt{v_z}}\frac{n}{m+1}\sqrt{\frac{N}{N_1N_2}}\omega_1^{-1}\omega_2^{-1}d_1(\Delta^2),\\
\varepsilon_{z,4}(\Delta^2) &= \frac{\lambda_z^2}{8}d_1(\Delta^2),
\end{aligned}$$

with

$$\lambda_z = \frac{1}{v_z}\frac{n^2(n+1)}{(m+1)^2(m+2)}\frac{N}{N_1N_2}\omega_1^{-2}, \text{ and } r_{z,1} = \frac{N(p-1)}{N_1N_2} - (p-2)\omega_1^2.$$

6 Asymptotic unbiased estimator of EER

In this section, we propose asymptotically unbiased estimators of EERs for W-rule and Z-rule under A1. Our estimators are constructed using the following result.

Lemma 5. *Let*

$$\widehat{y}_c = \frac{-(1/2)\widehat{\Delta}_w^2}{\sqrt{\left\{\widehat{\Delta}_w^2 + \frac{N(p-2)}{N_1N_2}\right\}\frac{n+1}{m+2}}}.$$

Then, the following equalities hold under A1:

$$E\left[\Phi\left(\widehat{y}_c\right)+\frac{1}{n}\sum_{k=1}^{4}\varepsilon_{w,k}\left(\widehat{\Delta}_w^2\right)H_{k-1}\left(\widehat{y}_c\right)\phi\left(\widehat{y}_c\right)\right]=\Phi\left(y_c\right)+O(n^2),$$
$$E\left[\bar{\ell}_k\left(\widehat{\Delta}_w^2\right)\right]=\bar{\ell}_k\left(\Delta^2\right)+O(n^{-1})\quad(k=1,\ldots,4).$$

Since Lemma 5 can be similarly proved with Theorem 6, we omit the proof. From (25) and Lemma 5, we can give an asymptotically unbiased estimator of $\mathrm{EER}_w(c_{wm})$ under A1 as in the following theorem.

Theorem 9. *An estimator of* $\mathrm{EER}_w(c_{wm})$ *whose bias is of* $O(n^{-2})$ *under* A1 *is given by*

$$\widehat{\mathrm{EER}}_w(c_{wm})=\Phi\left(\widehat{y}_c\right)-\frac{1}{n}\sum_{k=1}^{4}\widehat{\eta}_{w,k}H_{k-1}\left(\widehat{y}_c\right)\phi\left(\widehat{y}_c\right),\tag{16}$$

where $\widehat{\eta}_{w,k}=\eta_{w,k}\left(\widehat{\Delta}_w^2\right)$ *and*

$$\eta_{w,k}(\Delta^2)=\varepsilon_{w,k}(\Delta^2)-\bar{\ell}_k(\Delta^2).$$

To construct asymptotically unbiased estimator for $\mathrm{EER}_z(0)$, we use the following result.

Lemma 6. *For* $\widehat{y}_c$ *defined in* Lemma 6*,*

$$E\left[\frac{1}{\widehat{v}_w}H_1\left(\widehat{y}_c\right)\phi\left(\widehat{y}_c\right)\right]=\frac{1}{v_w}H_1(y_c)\phi(y_c)+O(n^{-1})$$

under A1*.*

Lemma 6 can be similarly proved with Theorem 6, and so we omit the proof. From 37, Theorem 9 and Lemma 6, we can get an asymptotically unbiased estimator of $\mathrm{EER}_z(0)$ under A1 which is given in the following theorem.

Theorem 10. *An estimator of* $\mathrm{EER}_z(0)$ *whose bias is of* $O(n^{-2})$ *under* A1 *is given by*

$$\widehat{\mathrm{EER}}_z(0)=\widehat{\mathrm{EER}}_w(c_{wm})+\frac{1}{4\widehat{v}_w}\frac{(n-1)(p-2)}{(m+1)^3}\left(\frac{n}{N_1}-\frac{n}{N_2}\right)^2H_1\left(\widehat{y}_c\right)\phi\left(\widehat{y}_c\right).\tag{17}$$

We note that $\widehat{\mathrm{EER}}_z(0)$ is the same as $\widehat{\mathrm{EER}}_w(c_{w,m})$ for the case in which $N_1=N_2$.

In Section 4, we mentioned that $\mathrm{EER}_w(0)$ and $\mathrm{EER}_z(0)$ are the same up to the term of order $O(n^{-1})$ under A0. McLachlan [16] gave an asymptotically unbiased estimator of $\mathrm{EER}_w(0)$ under A0, which is also an asymptotically unbiased estimator of $\mathrm{EER}_z(0)$, which are stated as follows.

Theorem 11. *An estimator of* $\mathrm{EER}_w(0)$ *whose bias is of* $O(n^{-2})$ *under* A0 *is given by*

$$\widehat{\mathrm{EER}}_w(0)=\Phi\left(-\frac{1}{2}D\right)+\phi\left(-\frac{1}{2}D\right)\left[\frac{1}{2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\frac{p-1}{D}+\frac{D}{32n}\{4(4p-1)-D^2\}\right].\tag{18}$$

This is also an estimator of $\mathrm{EER}_z(0)$ *whose bias is of* $O(n^{-2})$*.*

Since an asymptotic expansion of $\mathrm{EER}_w(c_{w,m})$ up to the term of $O(n^{-1})$ is the same as the one of $\mathrm{EER}_w(0)$ under A0, the right-hand side of the equality in (18) is also an estimator of $\mathrm{EER}_w(c_{w,m})$ whose bias is of $O(n^{-2})$ under A0.

7 Numerical comparisons for asymptotic approximations

To compare the accuracies of the derived asymptotic expansion approximations for the misclassification probabilities with the ones of Fujikoshi and Seo [9]'s limiting approximation, we calculated these values when $p = 8, 32$, $(N_1, N_2) = (30, 10)$, $(25, 15)$, $(20, 20)$, $(15, 25)$, $(10, 30)$, $\Delta = 1.05, 1.68, 2.56, 3.29$, where the setting of Δ is followed to Wyman et al. [24]. The setting for N and p were treated as the case in which $p : N = 1 : 5$ when $p = 8$, and the case in which $p : N = 1 : 5$ when $p = 32$. We considered for the case in which the cut-off point c is zero. Table 1 gives the values of $e_W(2|1)$ and $e_Z(2|1)$ when $p = 8$, and Table 2 gives the values of $e_W(2|1)$ and $e_Z(2|1)$ when $p = 32$. In these tables, we described the value of Fujikoshi and Seo [9]'s limiting approximation at column "FS's Aprox", and the value of asymptotic expansion based on Corollary 2 for W-rule and Corollary 3 for Z-rule at column "YSF's AE". To compare the accuracy, it is needed the values of misclassification probabilities calculated by simulation. When we treat the distributions of W- rule and Z- rule, without loss of generality from invariant property of the distribution for the orthogonal transformation of observation vector, we may assume that two given normal populations with the same covariance matrix are

$$\Pi_1 : N_p(-(\delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p), \quad \Pi_2 : N_p((\delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p),$$

where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$. To compute misclassification probability, generate 10^4 training samples. For each training samples, we generate 10^4 test samples in which observation vectors are i.i.d. as $N_p(-(\delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p)$. The value of the conditional misclassification probability is calculated by

$$\frac{\text{number of misclassification}}{10^4} \tag{19}$$

in each training samples. We took the average of these 10^4 values of conditional misclassification probability, and wrote it as the value of misclassification probability in column "Sim" in Tables 1 and 2.

From Tables 1 and 2, we can see that our proposed asymptotic expansion approximations have good accuracy compared to Fujikoshi and Seo [9]'s limiting approximation when $N_1 \neq N_2$. In addition, for the case in which $N_1 = N_2$, our proposed asymptotic expansion has almost the same precision of approximation with Fujikoshi and Seo [9]'s limiting approximation.

In Table 3, we give the values of $\mathrm{EER}_w(0)$, $\mathrm{EER}_w(c_{w,m})$ and $\mathrm{EER}_z(0)$ obtained by simulation for the case $\pi_1 = \pi_2$ obtained by simulation. Simulating setting and computation method are the same as ones in Table 1. We can see a tendency from Table 3 that the magnitude of these error rate has the order "$\mathrm{EER}_z(0) < \mathrm{EER}_w(c_{w,m}) < \mathrm{EER}_w(0)$" for almost all simulation settings. There are little difference between the magnitudes of $\mathrm{EER}_z(0)$ and $\mathrm{EER}_w(c_{w,m})$ when $N_1 = N_2$ (The difference appears in less than 4th place of decimal point.).

We also checked the precision of the proposed asymptotically unbiased estimators of $e_W(2|1)$ and $e_Z(2|1)$ by simulation. Simulating setting and procedure of calculation are the same as ones in Table 1.

Table 1: Comparison of approximations of $e_w(2|1)$ and $e_z(2|1)$ for cut-off point 0 when $p = 8$

(N_1, N_2)	Δ	$e_w(2\|1)$			$e_z(2\|1)$		
		FS's Aprox	YSF's AE	Sim	FS's Aprox	YSF's AE	Sim
$(30, 10)$	1.05	0.310	0.313	0.312	0.369	0.372	0.371
	1.69	0.223	0.226	0.225	0.261	0.265	0.264
	2.56	0.125	0.127	0.128	0.144	0.149	0.149
	3.29	0.070	0.072	0.072	0.080	0.085	0.084
$(25, 15)$	1.05	0.337	0.338	0.338	0.362	0.363	0.363
	1.69	0.239	0.240	0.240	0.255	0.256	0.256
	2.56	0.133	0.134	0.133	0.141	0.142	0.142
	3.29	0.074	0.075	0.075	0.078	0.080	0.080
$(20, 20)$	1.05	0.360	0.360	0.359	0.360	0.360	0.359
	1.69	0.254	0.253	0.251	0.254	0.252	0.251
	2.56	0.140	0.140	0.140	0.140	0.140	0.140
	3.29	0.078	0.078	0.078	0.078	0.078	0.078
$(15, 25)$	1.05	0.388	0.386	0.384	0.362	0.360	0.358
	1.69	0.271	0.269	0.268	0.255	0.252	0.252
	2.56	0.149	0.148	0.148	0.141	0.139	0.139
	3.29	0.082	0.083	0.083	0.078	0.078	0.077
$(10, 30)$	1.05	0.431	0.428	0.427	0.369	0.366	0.365
	1.69	0.302	0.298	0.297	0.261	0.256	0.255
	2.56	0.165	0.164	0.164	0.144	0.140	0.140
	3.29	0.091	0.091	0.092	0.080	0.078	0.078

Table 2: Comparison of approximations of $e_w(2|1)$ and $e_z(2|1)$ for cut-off point 0 when $p = 32$

(N_1, N_2)	Δ	$e_w(2\|1)$			$e_z(2\|1)$		
		FS's Aprox	YSF's AE	Sim	FS's Aprox	YSF's AE	Sim
$(30, 10)$	1.05	0.377	0.382	0.383	0.458	0.460	0.461
	1.69	0.339	0.344	0.344	0.406	0.412	0.411
	2.56	0.277	0.285	0.285	0.328	0.337	0.338
	3.29	0.228	0.238	0.238	0.267	0.280	0.279
$(25, 15)$	1.05	0.418	0.421	0.421	0.454	0.456	0.455
	1.69	0.371	0.375	0.375	0.400	0.404	0.404
	2.56	0.300	0.307	0.304	0.321	0.329	0.326
	3.29	0.244	0.254	0.253	0.261	0.271	0.270
$(20, 20)$	1.05	0.453	0.454	0.453	0.453	0.454	0.453
	1.69	0.399	0.402	0.401	0.399	0.402	0.401
	2.56	0.319	0.325	0.326	0.319	0.325	0.326
	3.29	0.259	0.268	0.268	0.259	0.268	0.268
$(15, 25)$	1.05	0.490	0.490	0.490	0.454	0.455	0.455
	1.69	0.430	0.432	0.432	0.400	0.402	0.403
	2.56	0.343	0.349	0.348	0.321	0.326	0.326
	3.29	0.277	0.286	0.284	0.261	0.168	0.266
$(10, 30)$	1.05	0.540	0.538	0.538	0.458	0.458	0.458
	1.69	0.477	0.477	0.478	0.406	0.407	0.408
	2.56	0.382	0.386	0.387	0.328	0.331	0.331
	3.29	0.308	0.317	0.317	0.267	0.272	0.272

Table 3: Comparison of EER

(N_1, N_2)	Δ	$p=8$ $\mathrm{EER}_w(0)$	$\mathrm{EER}_w(c_{wm})$	$\mathrm{EER}_z(0)$	$p=32$ $\mathrm{EER}_w(0)$	$\mathrm{EER}_w(c_{wm})$	$\mathrm{EER}_z(0)$
(30, 10)	1.05	0.371	0.370	0.369	0.460	0.460	0.459
	1.69	0.262	0.262	0.261	0.410	0.410	0.409
	2.56	0.145	0.145	0.144	0.335	0.334	0.334
	3.29	0.082	0.081	0.081	0.276	0.276	0.275
(25, 15)	1.05	0.362	0.362	0.362	0.455	0.455	0.455
	1.69	0.255	0.255	0.254	0.403	0.403	0.403
	2.56	0.141	0.141	0.141	0.326	0.326	0.326
	3.29	0.079	0.079	0.079	0.269	0.269	0.269
(20, 20)	1.05	0.359	0.359	0.359	0.454	0.454	0.454
	1.69	0.252	0.252	0.252	0.401	0.401	0.401
	2.56	0.140	0.140	0.140	0.326	0.326	0.326
	3.29	0.078	0.078	0.078	0.268	0.268	0.268
(15, 25)	1.05	0.361	0.361	0.361	0.455	0.455	0.455
	1.69	0.254	0.254	0.254	0.404	0.404	0.404
	2.56	0.141	0.141	0.141	0.327	0.327	0.327
	3.29	0.079	0.079	0.079	0.269	0.268	0.268
(10, 30)	1.05	0.370	0.369	0.369	0.459	0.459	0.458
	1.69	0.261	0.261	0.260	0.411	0.410	0.410
	2.56	0.145	0.145	0.145	0.336	0.335	0.334
	3.29	0.082	0.081	0.081	0.277	0.276	0.276

As a competitor, we used the estimator obtained as Fujikoshi and Seo [9]'s limiting approximation by replacing Δ^2 with $\widehat{\Delta}_{\mathrm{p}}^2 = \max\{\widehat{\Delta}^2, 0\}$, and wrote the value in column "FS's Est" in Table 4 for $p = 8$, and in Table 5 for $p = 43$. The value in parenthesis in column "FS's Est" is obtained by computing the mean squared error

$$\mathrm{MSE} = \frac{1}{10^4}\sum_{i=1}^{10^4}(\text{FS's Est}^{(i)} - \text{Sim})^2, \tag{20}$$

where "FS's Est$^{(i)}$" stands for the value calculated by i-th training sample, and "Sim" stands for misclassification probability computed by simulation. From Tables 4 and 5 we can see that the proposed estimators have good accuracy compared to method being used Fujikoshi and Seo [9]'s approximation. The MSE for proposed estimator is small than for the method being used Fujikoshi and Seo [9]'s approximation for the case in which $\Delta = 2.56, 3.29$.

We tried to compare the precision of asymptotically unbiased estimator of expected error rate by simulation for $p = 1, 2, 5, 8, 16, 32$, $N_1 = N_2 = 20$. The setting of Δ is the same as Table 5. We computed values of 2-types of conditional misclassification probabilities by (19), and take average of these 2 values. This calculation is repeated 10^3 times. We wrote the average of these 10^3 values in column "Sim" in Table 6. The column "A0" in Table 6 gives the averages of 10^3 replicated values of (18), and the column "A1" gives the averages of 10^3 replicated values of (16). We computed the mean squared error by using similar calculation to (20), and wrote it in parenthesis in each case. Table 6 shows that the precision of approximation (18) becomes worsen as the dimensionality gets large. We can check that the precision of approximation (16) is good in each case of p.

Table 4: Estimated values of $e_w(2|1)$ and $e_z(2|1)$ for cut-off point 0 when $p = 8$

(N_1, N_2)	Δ	$e_w(2\|1)$ FS's Est (MSE)	$e_w(2\|1)$ YSF's Est (MSE)	$e_w(2\|1)$ sim	$e_z(2\|1)$ FS's Est (MSE)	$e_z(2\|1)$ YSF's Est (MSE)	$e_z(2\|1)$ sim
(30, 10)	1.05	0.330(0.004)	0.312(0.005)	0.312	0.396(0.007)	0.370(0.008)	0.371
	1.68	0.248(0.005)	0.225(0.004)	0.225	0.292(0.007)	0.264(0.006)	0.264
	2.56	0.147(0.003)	0.127(0.002)	0.128	0.170(0.004)	0.149(0.003)	0.149
	3.29	0.089(0.002)	0.072(0.001)	0.072	0.102(0.002)	0.085(0.002)	0.084
(25, 15)	1.05	0.361(0.005)	0.337(0.005)	0.338	0.389(0.006)	0.362(0.006)	0.363
	1.68	0.263(0.005)	0.239(0.004)	0.240	0.281(0.006)	0.255(0.005)	0.256
	2.56	0.154(0.003)	0.133(0.002)	0.133	0.164(0.003)	0.142(0.003)	0.142
	3.29	0.093(0.002)	0.075(0.001)	0.075	0.098(0.002)	0.080(0.001)	0.080
(20, 20)	1.05	0.387(0.006)	0.359(0.006)	0.359	0.387(0.006)	0.359(0.006)	0.359
	1.68	0.279(0.005)	0.252(0.004)	0.251	0.279(0.005)	0.252(0.004)	0.251
	2.56	0.161(0.003)	0.138(0.002)	0.140	0.161(0.003)	0.138(0.002)	0.140
	3.29	0.097(0.002)	0.078(0.001)	0.078	0.097(0.002)	0.079(0.001)	0.078
(15, 25)	1.05	0.419(0.008)	0.386(0.008)	0.384	0.390(0.007)	0.360(0.007)	0.358
	1.68	0.302(0.007)	0.269(0.005)	0.268	0.283(0.006)	0.252(0.005)	0.252
	2.56	0.174(0.004)	0.148(0.003)	0.148	0.164(0.004)	0.139(0.003)	0.139
	3.29	0.103(0.002)	0.083(0.002)	0.083	0.098(0.002)	0.078(0.002)	0.077
(10, 30)	1.05	0.465(0.011)	0.426(0.012)	0.427	0.397(0.007)	0.364(0.008)	0.365
	1.68	0.340(0.011)	0.298(0.008)	0.297	0.293(0.008)	0.256(0.006)	0.255
	2.56	0.195(0.006)	0.163(0.004)	0.164	0.170(0.005)	0.140(0.003)	0.140
	3.29	0.116(0.003)	0.092(0.002)	0.092	0.102(0.003)	0.078(0.002)	0.078

Table 5: Estimated values of $e_w(2|1)$ and $e_z(2|1)$ for cut-off point 0 when $p = 32$

(N_1, N_2)	Δ	$e_w(2\|1)$ FS's Est (MSE)	$e_w(2\|1)$ YSF's Est (MSE)	$e_w(2\|1)$ sim	$e_z(2\|1)$ FS's Est (MSE)	$e_z(2\|1)$ YSF's Est (MSE)	$e_z(2\|1)$ sim
(30, 10)	1.05	0.397(0.001)	0.381(0.007)	0.383	0.485(0.002)	0.460(0.012)	0.461
	1.68	0.386(0.004)	0.343(0.007)	0.344	0.470(0.007)	0.410(0.012)	0.411
	2.56	0.360(0.010)	0.286(0.007)	0.285	0.436(0.016)	0.339(0.011)	0.338
	3.29	0.324(0.013)	0.235(0.007)	0.238	0.389(0.021)	0.277(0.010)	0.279
(25, 15)	1.05	0.445(0.002)	0.421(0.008)	0.421	0.484(0.003)	0.455(0.010)	0.455
	1.68	0.431(0.006)	0.375(0.008)	0.375	0.468(0.008)	0.404(0.010)	0.404
	2.56	0.391(0.013)	0.305(0.008)	0.304	0.424(0.017)	0.327(0.010)	0.326
	3.29	0.350(0.016)	0.253(0.008)	0.253	0.377(0.020)	0.270(0.009)	0.270
(20, 20)	1.05	0.485(0.003)	0.455(0.009)	0.453	0.485(0.003)	0.455(0.009)	0.453
	1.68	0.468(0.008)	0.402(0.009)	0.401	0.468(0.008)	0.402(0.009)	0.401
	2.56	0.423(0.016)	0.325(0.009)	0.326	0.423(0.016)	0.325(0.009)	0.326
	3.29	0.373(0.020)	0.267(0.009)	0.268	0.373(0.020)	0.267(0.009)	0.268
(15, 25)	1.05	0.524(0.003)	0.490(0.012)	0.490	0.484(0.003)	0.455(0.010)	0.455
	1.68	0.505(0.009)	0.430(0.012)	0.432	0.468(0.008)	0.400(0.010)	0.403
	2.56	0.459(0.020)	0.349(0.011)	0.348	0.426(0.017)	0.326(0.009)	0.326
	3.29	0.406(0.025)	0.286(0.010)	0.284	0.378(0.021)	0.268(0.009)	0.266
(10, 30)	1.05	0.573(0.004)	0.538(0.018)	0.538	0.485(0.003)	0.457(0.013)	0.458
	1.68	0.557(0.011)	0.479(0.017)	0.478	0.472(0.007)	0.408(0.012)	0.408
	2.56	0.512(0.026)	0.387(0.016)	0.387	0.435(0.018)	0.331(0.011)	0.331
	3.29	0.457(0.033)	0.314(0.014)	0.317	0.389(0.023)	0.270(0.010)	0.272

Table 6: Comparison of asymptotic unbiased estimate of EER for W-rule when $N_1 = N_2 = 20$

Δ	$p=1$			$p=2$		
	A0	A1	Sim	A0	A1	Sim
1.05	0.301(0.004)	0.305(0.007)	0.302	0.313(0.004)	0.312(0.004)	0.314
1.68	0.203(0.003)	0.203(0.003)	0.203	0.214(0.003)	0.214(0.003)	0.210
2.56	0.103(0.001)	0.103(0.001)	0.103	0.105(0.002)	0.105(0.002)	0.108
3.29	0.053(0.001)	0.053(0.001)	0.052	0.055(0.001)	0.055(0.001)	0.055
Δ	$p=5$			$p=8$		
	A0	A1	Sim	A0	A1	Sim
1.05	0.340(0.006)	0.339(0.006)	0.340	0.357(0.006)	0.358(0.006)	0.359
1.68	0.232(0.004)	0.233(0.004)	0.233	0.253(0.004)	0.254(0.004)	0.253
2.56	0.119(0.002)	0.120(0.002)	0.123	0.135(0.002)	0.138(0.002)	0.140
3.29	0.065(0.001)	0.066(0.001)	0.066	0.077(0.001)	0.079(0.001)	0.078
Δ	$p=16$			$p=32$		
	A0	A1	Sim	A0	A1	Sim
1.05	0.397(0.008)	0.400(0.007)	0.400	0.401(0.020)	0.457(0.009)	0.454
1.68	0.301(0.006)	0.306(0.006)	0.303	0.325(0.025)	0.403(0.009)	0.403
2.56	0.178(0.004)	0.187(0.004)	0.188	0.205(0.032)	0.321(0.009)	0.327
3.29	0.111(0.003)	0.122(0.002)	0.119	0.133(0.029)	0.269(0.008)	0.264

8 Criteria for selection of variables

We consider the problem for selection of variables in two-group discriminant analysis. McLachlan [16] and [17] proposed a criterion which is based on asymptotic unbiased estimator of the expected error rate under A0. In this section, we will derive such a criterion under A1.

The problem of selection of variables is to identify a sub-vector $\boldsymbol{x}(j) = (x_{j_1}, \ldots, x_{j_k})'$ of $\boldsymbol{x}$ which is corresponded to a subset of subscripts $\{1, 2, \ldots, p\}$, where $k = k(j)$ is the cardinal number of j, i.e. $k(j) = \#j$. Let $\mathcal{J}$ be the family of all possible subsets of $\{1, \ldots, p\}$. Then the problem may be regarded as how to select the best subset of j from $\mathcal{J}$. If we only use $\boldsymbol{x}(j)$ in discriminant analysis, the corresponding W and Z discriminant functions become

$$
\begin{aligned}
W(\boldsymbol{x}(j)) &= (\bar{\boldsymbol{x}}_1(j) - \bar{\boldsymbol{x}}_2(j))'\boldsymbol{S}(j)^{-1}\left\{\boldsymbol{x}(j) - \frac{1}{2}(\bar{\boldsymbol{x}}_1(j) + \bar{\boldsymbol{x}}_2(j))\right\}, \\
Z(\boldsymbol{x}(j)) &= \frac{1}{2}\left\{(1 + N_2^{-1})^{-1}(\boldsymbol{x}(j) - \bar{\boldsymbol{x}}_2(j))'\boldsymbol{S}(j)^{-1}(\boldsymbol{x}(j) - \bar{\boldsymbol{x}}_2(j))\right. \\
&\quad \left. -(1 + N_1^{-1})^{-1}(\boldsymbol{x}(j) - \bar{\boldsymbol{x}}_1(j))'\boldsymbol{S}(j)^{-1}(\boldsymbol{x}(j) - \bar{\boldsymbol{x}}_1(j))\right\},
\end{aligned}
$$

where $\bar{\boldsymbol{x}}_g(j)$ and $\boldsymbol{S}(j)$ denote $\bar{\boldsymbol{x}}_g$ and $\boldsymbol{S}$ corresponding to $\boldsymbol{x}(j)$, respectively. The expected error rate for W-rule on the model j is expressed as

$$
\mathrm{EER}_w(c_w; j) = \pi_1 P(W(\boldsymbol{x}(j)) < c_w | \boldsymbol{x} \in \Pi_1) + \pi_2 P(W(\boldsymbol{x}(j)) > c_w | \boldsymbol{x} \in \Pi_2),
$$

and the one for Z-rule is expressed as

$$
\mathrm{EER}_z(c_z; j) = \pi_1 P(Z(\boldsymbol{x}(j)) < c_z | \boldsymbol{x} \in \Pi_1) + \pi_2 P(Z(\boldsymbol{x}(j)) > c_z | \boldsymbol{x} \in \Pi_2).
$$

8.1 Criterion for selection of variables

For ease of explanation, we consider the case that $\pi_1 = \pi_2$. Set the cut-off point for W-rule corresponding to $\boldsymbol{x}(j)$ as

$$c_{w,m}(j) = \frac{1}{2}\frac{N}{N-k(j)}\left(\frac{k(j)}{N_2} - \frac{k(j)}{N_1}\right).$$

From the statement in Section 4, the limiting value for $\mathrm{EER}_w(c_{w,m}(j); j)$ takes the minimum value under the high-dimensional asymptotic framework A1(j);

$$\begin{aligned} \mathrm{A1}(j) : k(j) \to \infty, \quad N_1 \to \infty, \quad N_2 \to \infty, \quad k(j)/n \to \gamma(j) \in [0,1), \\ \text{and} \quad N_1/N_2 \to \gamma \in (0,\infty). \end{aligned}$$

We see that $\lim_{\mathrm{A1}(j)} \mathrm{EER}_z(0; j)$ takes the minimum value, and see that

$$\lim_{\mathrm{A1}(j)} \mathrm{EER}_z(0; j) = \lim_{\mathrm{A1}(j)} \mathrm{EER}_w(c_{w,m}(j); j). \tag{21}$$

To obtain the criterion for variable selection, firstly, we derive asymptotically unbiased estimators of $\mathrm{EER}_w(c_{w,m}(j); j)$ and $\mathrm{EER}_z(0; j)$.

Since the framework A1(j) is an imitation which is obtained from A1 by replacing p in A1 with $k(j)$, it follows from the expansion (25) that

$$\mathrm{EER}_w(c_{w,m}(j); j) = \Phi(y_c(j)) - \frac{1}{n}\sum_{i=1}^{4} \bar{\ell}_i(j) H_{i-1}(y_c(j))\phi(y_c(j)) + O(n^{-2}), \tag{22}$$

where

$$y_c(j) = \frac{-(1/2)\Delta(j)^2}{\sqrt{\left\{\Delta(j)^2 + \frac{N(k(j)-2)}{N_1N_2}\right\}\frac{n+1}{m(j)+2}}},$$

$m(j) = n - k(j)$, $\Delta(j)$ and $\bar{\ell}_i(j)$ are denoted Δ and $\bar{\ell}_i$ which is given in (26), respectively, corresponding to $\boldsymbol{x}(j)$. By using the same reason, we find from Theorem 4 (2) that

$$\begin{aligned} \mathrm{EER}_z(0; j) = &\ \mathrm{EER}_w(c_{w,m}) \\ &+ \frac{1}{4v_w(j)}\frac{(n-1)(k(j)-2)}{(m(j)+1)^3}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 H_1\left(y_c(j)\right)\phi\left(y_c(j)\right) + O(n^{-2}), \end{aligned}$$

where $v_w(j)$ is the v_w corresponding to $\boldsymbol{x}(j)$. The following theorem gives asymptotic unbiased estimates of $\mathrm{EER}_w(c_{w,m})$ and $\mathrm{EER}_z(0; j)$.

Theorem 12. *Let*

$$\widehat{y}_c(j) = \frac{-(1/2)\widehat{\Delta}_w(j)^2}{\sqrt{\left\{\widehat{\Delta}_w(j)^2 + \frac{N(k(\boldsymbol{j})-2)}{N_1N_2}\right\}\frac{n+1}{m(j)+2}}},$$

where $\widehat{\Delta}_w(j)^2$ is the $\widehat{\Delta}_w^2$ given in Corollary 7 corresponding to the model j. Let

$$G_{w,h}(j) = \widehat{y}_c(j) + \frac{1}{n}\sum_{i=1}^{4} \widehat{\eta}_{w,i}(j) H_{i-1}\left(\widehat{y}_c(j)\right),$$

where $\widehat{\eta}_{w,i}(j) = \eta_{w,i}(\widehat{\Delta}_w(j)^2; j)$, *and*

$$\eta_{w,i}(\Delta(j)^2; j) = \varepsilon_{w,i}\left(\Delta(j)^2\right) - \bar{\ell}_i(j).$$

Then under the high-dimensional asymptotic framework A1(j),

$$\mathrm{EER}_w(c_{w,m}(j); j) = E[\Phi(G_{w,h}(j))] + O(n^{-2}).$$

Let

$$G_{z,h}(j) = G_{w,h}(j) + \frac{1}{4\widehat{v}_w(j)} \frac{(n-1)(k(j)-2)}{(m(j)+1)^3} \left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 H_1\left(\widehat{y}_c(j)\right) \phi\left(\widehat{y}_c(j)\right),$$

where

$$\widehat{v}_w(j) = \frac{n^2(n+1)}{(m(j)+1)^2(m(j)+2)} \left\{\widehat{\Delta}_w(j)^2 + \frac{N(k(j)-2)}{N_1 N_2}\right\}.$$

Then under the high-dimensional asymptotic framework A1(j),

$$\mathrm{EER}_z(0; j) = E[\Phi(G_{z,h}(j))] + O(n^{-2}).$$

We propose the selection method for W-rule with the cut-off point $c_{wm}(j)$ base on

$$M_{w,h}(j) = \Phi\left(G_{w,h}(j)\right),$$

and propose for Z-rule with the cut-off point 0 based on

$$M_{z,h}(j) = \Phi\left(G_{z,h}(j)\right).$$

The selected model $\widehat{j}_{M,w,h}$ is obtained by satisfying that $M_{w,h}(\widehat{j}_{M,w,h}) = \min_{j \in \mathcal{J}} M_{w,h}(j)$, and $\widehat{j}_{M,z,h}$ is obtained by satisfying that $M_{z,h}(\widehat{j}_{M,z,h}) = \min_{j \in \mathcal{J}} M_{z,h}(j)$. Since $\Phi(\cdot)$ is a monotone increasing function, $\widehat{j}_{M,w,h}$ and $\widehat{j}_{M,z,h}$ minimize $G_{w,h}(j)$ and $G_{z,h}(j)$, respectively.

8.2 Relationship between expected error rate and no additional information model

Firstly, we consider no additional information model $\Omega(j)$ which leads $\boldsymbol{x}(j)$ to be the best subsets of variables. Define $\Omega(j)$ as

$$\Omega(j): \quad a_k \neq 0, \text{ for any } k \in j, \quad \text{and} \quad a_k = 0, \text{ for any } k \in j^c \cap \{1, 2, \ldots, p\}, \tag{23}$$

where $\boldsymbol{a} = (a_1, \ldots, a_p)' = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

Let j_0 be a fixed subset in $\mathcal{J}$. We may assume without loss of generality that $j_0 = \{1, \ldots, k_0\}$. We call $\Omega(j_0)$ true model if $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}\}$ satisfies the condition in (23). Let

$$\mathcal{J}_1 = \{j \in \mathcal{J} : j \supseteq j_0\} \quad \text{and} \quad \mathcal{J}_2 = \mathcal{J}_1^c \cap \mathcal{J}.$$

It is known (see, e.g., Fujikoshi [4]) that $\Omega(j_0)$ is true if and only if

$$\Delta(j) = \Delta \text{ for any } j \in \mathcal{J}_1 \quad \text{and} \quad \Delta(j) < \Delta \text{ for any } j \in \mathcal{J}_2. \tag{24}$$

Theorem 13. *Assume that $\Omega(j_0)$ is true and k_0 is fixed. Then, it holds that*

(i) $\lim_{\mathrm{A1}} \{\mathrm{EER}_w(c_{wm}(j); j) - \mathrm{EER}_w(c_{wm}(j_0); j_0)\} \geq 0$ *for* $j \in \mathcal{J}_1 \setminus \{j_0\}$,

(ii) $\lim_{\mathrm{A1}} \{\mathrm{EER}_w(c_{wm}(j); j) - \mathrm{EER}_w(c_{wm}(j_0); j_0)\} > 0$ *for* $j \in \mathcal{J}_2$.

In addition, it holds that

(iii) $\lim_{\mathrm{A1}} \{\mathrm{EER}_z(0; j) - \mathrm{EER}_z(0; j_0)\} \geq 0$ *for* $j \in \mathcal{J}_1 \setminus \{j_0\}$,

(iv) $\lim_{\mathrm{A1}} \{\mathrm{EER}_z(0; j) - \mathrm{EER}_z(0; j_0)\} > 0$ *for* $j \in \mathcal{J}_2$.

Proof. From (21), it is sufficient to prove (i) and (ii) only. It follows from (13) that

$$\lim_{\mathrm{A1}(j)} \mathrm{EER}_w(c_{w,m}(j); j) = \Phi\left(\frac{-(1/2)\Delta_0(j)^2}{\sqrt{\{\Delta_0(j)^2 + (2+\gamma+\gamma^{-1})\gamma(j)\}\, 1-\gamma(j)}}\right),$$

where $\Delta_0(j)$ is a positive value which is defined as $\Delta_0(j) = \lim_{k(j)\to\infty} \Delta(j)$. From the assumption, we have

$$\lim_{\mathrm{A1}} \mathrm{EER}_w(c_{w,m}(j_0); j_0) = \lim_{\mathrm{A0}} \mathrm{EER}_w(c_{w,m}(j_0); j_0) = \Phi\left(-\frac{1}{2}\Delta(j_0)\right).$$

Since $0 \leq \gamma(j) < 1$, it holds that

$$\frac{-(1/2)\Delta_0(j)^2}{\sqrt{\{\Delta_0(j)^2 + (2+\gamma+\gamma^{-1})\gamma(j)\}\, \frac{1}{1-\gamma(j)}}} \geq \frac{-(1/2)\Delta_0(j)^2}{\sqrt{\Delta_0(j)^2}},$$

where the equality hold for the case in which $\gamma(j) = 0$. The condition (24) implies that

$$\frac{-(1/2)\Delta_0(j)^2}{\sqrt{\Delta_0(j)^2}} = -\frac{1}{2}\Delta(j_0) \quad (j \in \mathcal{J}_1),$$
$$\frac{-(1/2)\Delta_0(j)^2}{\sqrt{\Delta_0(j)^2}} > -\frac{1}{2}\Delta(j_0) \quad (j \in \mathcal{J}_2),$$

which prove (i) and (ii) for the case in which $k(j) \to \infty$ as $p \to \infty$. Assume that $k(j)$ is fixed. Then,

$$\lim_{\mathrm{A1}} \mathrm{EER}_w(c_{w,m}(j); j) = \lim_{\mathrm{A0}} \mathrm{EER}_w(c_{w,m}(j); j) = \Phi\left(-\frac{1}{2}\Delta(j)\right),$$

which leads to (i) and (ii). □

From Theorem 13 , we can regard $\Omega(j)$ as a minimal realization of the parametric model such that $\mathrm{EER}_w(c_{w,m}(j); j)$ and $\mathrm{EER}_z(0; j)$ are minimum in the sense of (i)-(iv). Note that the minimization for $\Omega(j)$ leads the model j in which selected variables are overspecified for the minimization of $\mathrm{EER}_w(c_{w,m}(j); j)$ and $\mathrm{EER}_z(0; j)$ in the sense of (i) and (iii).

8.3 Simulation

From Theorem 13, $\mathrm{EER}_w(c_{wm}(j_0); j_0)$ and $\mathrm{EER}_z(0; j_0)$ become minimum for any $j \in \mathcal{J}$ in the limiting sense when the model $\Omega(j_0)$ is true. So in this simulation, we set the parameters $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}\}$ which satisfies (23). The covariance matrix $\boldsymbol{\Sigma}$ is assumed to be $\boldsymbol{I}_p$. We set $\boldsymbol{\mu}_2 = -\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_1 = -(\alpha/2)(1, \ldots, 1, 0, \ldots, 0)'$, where each of the first k elements of $\boldsymbol{\mu}_1$ is $-\alpha/2$ and the remaining is 0. Simulation experiments were carried out for the case in which $N_1 = N_2 = 80$, $k = p/10, p/5, p/2$, $p = 30, 50$, $\alpha = 1, 2, 4$. When $\#j$ is small, the precision of the approximation of $M_{w,h}(j)$ is not good. So we use a switching procedure that the model is selected by $M(j)$ for $\#j < 5$ and is selected by $M_{w,h}(j)$ for $\# \geq 5$, where $M(j)$ is the base of selection criterion given in Fujikoshi [5]. Since all candidate models are too much for large p (e.g., the number of candidate models is 1023 when $p = 10$), we use the forward step wise selection method. The details of the selection method is given in Algorithm 1.

Algorithm 1 The algorithm for forward step wise selection method

$j_{\mathrm{can}} \Leftarrow \{\}$
$j_{\mathrm{temp}} \Leftarrow \arg\min_{\#j=1} M(j)$
$t \Leftarrow \min_{\#j=1} M(j)$
$\ell \Leftarrow 2$
while $\ell < 5$ **do**
    if $t > \arg\min_{\#j=\ell} M(j)$ **then**
        $j_{\mathrm{temp}} \Leftarrow j_{\mathrm{temp}} \cup \arg\min_{\#j=1} M(j)$
        $t \Leftarrow \min_{\#j=1} M(j)$
        $\ell \Leftarrow \ell + 1$
    else
        $j_{\mathrm{can}} \Leftarrow j_{\mathrm{temp}}$
    end if
end while
while $j_{\mathrm{can}} = \{\} \wedge 5 \leq \ell \leq p$ **do**
    if $t > \arg\min_{\#j=\ell} M_{w,h}(j)$ **then**
        $j_{\mathrm{temp}} \Leftarrow j_{\mathrm{temp}} \cup \arg\min_{\#j=1} M_{w,h}(j)$
        $t \Leftarrow \min_{\#j=1} M_{w,h}(j)$
        $\ell \Leftarrow \ell + 1$
    else
        $j_{\mathrm{can}} \Leftarrow j_{\mathrm{temp}}$
    end if
end while
return $j_{\mathrm{can}} \cup j_{\mathrm{temp}}$

Table 7 gives the frequencies of selected variables for 100 trials, where $j_{0,k+\ell} = \{j \in \mathcal{J}_1 : \#j = k+\ell\}$. We can see from the table that there are few models to be selected in $\mathcal{J}_2$ when $(p, k) = (30, 3)$, $(p, k) = (30, 6)$ and $(p, k) = (50, 5)$. On the other hand, the proposed method rarely selects the model in $\mathcal{J}_2$ when $(p, k) = (50, 25)$. It is considered that these results are guaranteed by assertions (i) and (ii) in Theorem 13. However, it seems that the proposed method rarely selects true model. It is expected to construct a variable selection criterion which is consistent.

Table 7: Frequencies of proposed variable selection for 100 trials

p	k	α	j_0	$j_{0,k+1}$	$j_{0,k+2}$	$\bigcup_{\ell=3}^{p-k} j_{0,k+\ell}$	$\mathcal{J}_2$
		1	1	5	12	82	0
	3	2	0	9	13	78	0
		4	0	6	10	84	0
		1	0	4	9	83	4
30	6	2	3	7	15	75	0
		4	3	8	11	78	0
		1	4	6	11	9	70
	15	2	6	9	13	12	60
		4	6	7	12	15	60
		1	0	0	0	96	2
	5	2	0	0	0	100	0
		4	0	1	0	99	0
		1	0	0	1	65	34
50	10	2	0	1	2	89	8
		4	0	0	4	93	2
		1	0	0	0	0	100
	25	2	0	0	1	1	98
		4	0	0	0	0	100

9 Concluding remarks

In this paper, asymptotic expansions of expected probabilities of misclassification for W- and Z- rules were derived up to the term of $O(n^{-1})$ under the high-dimensional asymptotic framework A1. It may be noted that the orders of their errors are $O(n^{-2})$. We compared expected error rates for these 2-rules asymptotically for the case in which prior probabilities are equal. It is known that under the large-sample asymptotic framework A0 expected error rates for these 2-rules are equal asymptotically. However, from our high-dimensional asymptotic results we have shown that the expected error rate for Z-rule is lower than or equal to the one for W-rule.

We proposed an asymptotic unbiased estimators of the expected probability of misclassification for W- and Z- rules under the high-dimensional asymptotic framework A1. In addition, asymptotic unbiased estimators of the expected error rate for these 2-rules were derived. Based on these unbiased estimators, we proposed variable selection criteria. Our asymptotic approximations were numerically examined.

A Proof of Theorem 4 (2)

In this section, we give a proof of Theorem 4 (2).

Proof of Theorem 4 (2). From Corollary 2,

$$g_w(c_{wm}; N_1, N_2, \Delta^2) = \Phi(y_{wm}) - \frac{1}{n}\sum_{k=1}^{4} \ell_k H_{k-1}(y_{wm})\phi(y_{wm}) + O(n^{-2})$$

under A1, where $y_{wm} = y_c + (1/n)R_4$ with

$$y_c = \frac{-(1/2)\Delta^2}{\sqrt{\left\{\Delta^2 + \frac{N(p-2)}{N_1N_2}\right\}\frac{n+1}{m+2}}},$$

$$R_4 = R_4(N_1, N_2) = \frac{\left(\frac{n}{N_2} - \frac{n}{N_1}\right)\left\{1 + \frac{p}{n} - \frac{p}{2(m+2)} - \frac{p}{n(m+2)}\right\}}{\sqrt{\left\{\Delta^2 + \frac{N(p-2)}{N_1N_2}\right\}\frac{n+1}{m+2}}}.$$

Then, we have

$$g_w(c_{wm}; N_1, N_2, \Delta^2) = \Phi(y_c) + \frac{1}{n}\left\{R_4(N_1, N_2) - \sum_{k=1}^{4} \ell_k(N_1, N_2)H_{k-1}(y_c)\right\}\phi(y_c) + O(n^{-2}).$$

It follows from the duality that

$$g_w(-c_{wm}; N_2, N_1, \Delta^2) = \Phi(y_c) + \frac{1}{n}\left\{R_4(N_2, N_1) - \sum_{k=1}^{4} \ell_k(N_2, N_1)H_{k-1}(y_c)\right\}\phi(y_c) + O(n^{-2}).$$

From these expansions, we obtain

$$\mathrm{EER}_w(c_{wm}) = \Phi(y_c) - \frac{1}{n}\sum_{k=1}^{4} \bar{\ell}_k H_{k-1}(y_c)\phi(y_c) + O(n^{-2}), \tag{25}$$

where

$$\bar{\ell}_k = \frac{1}{2}\{\ell_k(N_1, N_2) + \ell_k(N_2, N_1)\}$$

for $k = 1, 2, 3, 4$. Here,

$$\begin{aligned}
\bar{\ell}_1 &= -\frac{2n}{m+1}y_c, \\
\bar{\ell}_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{\frac{Nn}{N_1N_2}}{\Delta^2 + \frac{N(p-2)}{N_1N_2}} + \frac{n}{2N} + \frac{1}{2v_w}\left[\frac{2n}{m+1}\left\{y_c^2 v_w\right.\right. \\
&\quad \left. + \frac{1}{4}\left(\frac{n}{m+1}\right)^2\left(\frac{p-2}{N_1} - \frac{p-2}{N_2}\right)^2\right\} + \frac{n(p-2)}{2(m+1)^2}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 \\
&\quad \left. + \frac{n^2(n+1)}{2(m+1)^2(m+2)}\left(\frac{N_1 n}{NN_2} + \frac{N_2 n}{NN_1}\right)\Delta^2\right], \\
\bar{\ell}_3 &= -y_c\left(\frac{2n}{m+1} + \frac{Nn}{N_1N_2}\frac{1}{\Delta^2 + \frac{N(p-2)}{N_1N_2}}\right), \\
\bar{\ell}_4 &= \ell_4(N_1, N_2).
\end{aligned} \tag{26}$$

On the other hand, it follows from Corollary 3 that

$$g_z(0; N_1, N_2, \Delta^2) = \Phi(y_{zm}) - \frac{1}{n}\sum_{k=1}^{4} \zeta_k H_{k-1}(y_c)\phi(y_c) + O(n^{-2}) \tag{27}$$

where

$$y_{zm} = y_{zm}(N_1, N_2) = \frac{-\{\omega_2(N_1, N_2)\}^{-1}\Delta^2}{\sqrt{[\Delta^2 + (p-2)\{\omega_1(N_1, N_2)\}^2]\,\frac{n+1}{m+2}}}.$$

Write

$$a(N_1,N_2)=\frac{1+N_2^{-1}}{1+N_1^{-1}}=1+(N_2^{-1}-N_1^{-1})(1+N_1^{-1})^{-1}=1+x_1.$$

From Maclaurin expansion for $(1+x_1)^{1/2}$, we have

$$a^{1/2}=1+\frac{1}{2}(N_2^{-1}-N_1^{-1})(1+N_1^{-1})^{-1}-\frac{1}{8}(N_2^{-1}-N_1^{-1})^2(1+N_1^{-1})^{-2}+O(n^{-3}). \tag{28}$$

It also holds that

$$\frac{1}{1+N_1^{-1}}=1-N_1^{-1}+\frac{N_1^{-2}}{1+N_1^{-1}}. \tag{29}$$

Substituting (29) into (28), we obtain

$$a^{1/2}=1+\frac{1}{2}(N_2^{-1}-N_1^{-1})(1-N_1^{-1})-\frac{1}{8}(N_2^{-1}-N_1^{-1})^2+O(n^{-3}),$$

and so,

$$n\{\omega_1(N_1,N_2)\}^2=\frac{Nn}{N_1N_2}+\frac{1}{4n}\left(\frac{Nn}{N_1N_2}\right)^2-\frac{1}{n}\left(\frac{n}{N_1}\right)^2+O(n^{-2}), \tag{30}$$

$$\{\omega_2(N_1,N_2)\}^2=4\left(1+\frac{3}{4}\frac{N}{N_1N_2}-\frac{1}{N_1}\right)+O(n^{-2})=4(1+x_2)+O(n^{-2}). \tag{31}$$

From Maclaurin expansion for $(1-x_2)^{-1/2}$, we have

$$\{\omega_2(N_1,N_2)\}^{-1}=\frac{1}{2}\left(1-\frac{3}{8}\frac{N}{N_1N_2}+\frac{1}{2N_1}\right)+O(n^{-2}). \tag{32}$$

Substituting (30) and (32) into $y_{zm}(N_1,N_2)$, and expanding it asymptotically under A1, we have

$$y_{zm}(N_1,N_2)=y_c+\frac{1}{n}\left(A(N_1,N_2)+\frac{1}{2v_w}B(N_1,N_2)\right)H_1(y_c)+O(n^{-2}), \tag{33}$$

where

$$A(N_1,N_2)=-\frac{1}{2}\left(\frac{3}{4}\frac{Nn}{N_1N_2}-\frac{n}{N_1}\right),$$

$$B(N_1,N_2)=-\frac{n(n+1)(p-2)}{(m+1)^2(m+2)}\left\{\frac{1}{4}\left(\frac{Nn}{N_1N_2}\right)^2-\left(\frac{n}{N_1}\right)^2\right\}.$$

Substituting (33) into (27), and expanding it asymptotically under A1, we obtain

$$g_z(0;N_1,N_2,\Delta^2)=\Phi(y_c)+\frac{1}{n}\left[\left\{A(N_1,N_2)+\frac{1}{2v_w}B(N_1,N_2)\right\}H_1(y_c)\right.$$
$$\left.-\sum_{k=1}^{4}\zeta_k(N_1,N_2)H_{k-1}(y_c)\right\}\phi(y_c)+O(n^{-2}). \tag{34}$$

It follows from the duality that

$$g_z(0;N_2,N_1,\Delta^2)=\Phi(y_c)+\frac{1}{n}\left[\left\{A(N_2,N_1)+\frac{1}{2v_w}B(N_2,N_1)\right\}H_1(y_c)\right.$$
$$\left.-\sum_{k=1}^{4}\zeta_k(N_2,N_1)H_{k-1}(y_c)\right\}\phi(y_c)+O(n^{-2}). \tag{35}$$

From (34) and (35), we can obtain an asymptotic expansion of $\mathrm{EER}_z(0)$, which is as follows.

$$\mathrm{EER}_z(0) = \Phi(y_c) + \frac{1}{n}\left\{\left(\bar{A} + \frac{1}{2v_w}\bar{B}\right)H_1(y_c) - \sum_{k=1}^{4}\bar{\zeta}_k H_{k-1}(y_c)\right\}\phi(y_c) + O(n^{-2}), \tag{36}$$

where

$$\begin{aligned}
\bar{A} &= \frac{1}{2}\{A(N_1, N_2) + A(N_2, N_1)\} = -\frac{1}{8}\frac{Nn}{N_1N_2},\\
\bar{B} &= \frac{1}{2}\{B(N_1, N_2) + B(N_2, N_1)\} = \frac{n(n+1)(p-2)}{4(m+1)^2(m+2)}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2,\\
\bar{\zeta}_k &= \frac{1}{2}\{\zeta_k(N_1, N_2) + \zeta_k(N_2, N_1)\}
\end{aligned}$$

for $k = 1, 2, 3, 4$. Substituting (30) and (32) into $\bar{\zeta}_k$, and expanding it asymptotically, we have

$$\begin{aligned}
\bar{\zeta}_1 &= \bar{\ell}_1 + O(n^{-1}),\\
\bar{\zeta}_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{\frac{Nn}{N_1N_2}}{\Delta^2 + \frac{N(p-2)}{N_1N_2}}\\
&\quad + \frac{1}{2v_w}\left\{\frac{2n}{m+1}y_c^2 v_w + \frac{n^2(n+1)}{4(m+1)^2(m+2)}\frac{Nn}{N_1N_2}\Delta^2\right\} + O(n^{-1}),\\
\bar{\zeta}_3 &= \bar{\ell}_3 + O(n^{-1}),\\
\bar{\zeta}_4 &= \bar{\ell}_4 + O(n^{-1}).
\end{aligned}$$

From (25) and (36),

$$\mathrm{EER}_w(c_{wm}) - \mathrm{EER}_z(0) = -\frac{1}{n}\left(\bar{\ell}_2 - \bar{\zeta}_2 + \bar{A} + \frac{1}{2v_w}\bar{B}\right)H_1(y_c)\phi(y_c) + O(n^{-2}).$$

It can be computed that

$$\begin{aligned}
\bar{\ell}_2 - \bar{\zeta}_2 = \frac{n}{2N} + \frac{1}{2v_w}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 &\left\{\frac{n(n-1)(p-2)}{2(m+1)^3}\right.\\
&\left. + \frac{n^2(n+1)}{4(m+1)^2(m+2)}\frac{N_1N_2}{Nn}\Delta^2\right\} + O(n^{-1}).
\end{aligned}$$

Since it can be expressed that

$$-\frac{1}{8}\frac{Nn}{N_1N_2} + \frac{n}{2N} = -\frac{N_1N_2}{8Nn}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2,$$

we find that

$$\bar{\ell}_2 - \bar{\zeta}_2 + \bar{A} + \frac{1}{2v_w}\bar{B} = \frac{1}{4v_w}\frac{n(n-1)(p-2)}{(m+1)^3}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 + O(n^{-1}).$$

This gives

$$\mathrm{EER}_w(c_{wm}) - \mathrm{EER}_z(0) = -\frac{1}{4v_w}\frac{(n-1)(p-2)}{(m+1)^3}\left(\frac{n}{N_1} - \frac{n}{N_2}\right)^2 H_1(y_c)\phi(y_c) + O(n^{-2}). \tag{37}$$

□

B Proof of Theorem 6

In this section, we give a proof of Theorem 6.

Proof of Theorem 6. For dealing with the expectation of $\widehat{\Delta}_A^2$, it is no loss of generality from Lemma 4 that $\widehat{\Delta}_A^2$ may be

$$\widehat{\Delta}_A^2 = \frac{f_1}{y_1}\left(\Delta^2 + \frac{N}{N_1N_2}z_1^2 - 2\sqrt{\frac{N}{N_1N_2}}\Delta z_1 + \frac{N}{N_1N_2}y_2\right) - \frac{p-2}{n}\omega^{-2}, \tag{38}$$

where y_1, y_2 and z_1 are defined in Lemma 4, $f_1 = m+1$ and $f_2 = p-1$. Let $w_i = \sqrt{f_i/2}(y_i/f_i - 1)$ for $i = 1, 2$. It can be expressed that

$$\frac{f_1}{y_1} = 1 + \sum_{k=1}^{4}\left(-\sqrt{\frac{2}{f_1}}w_1\right)^k + \left(-\sqrt{\frac{2}{f_1}}w_1\right)^5 \frac{1}{1+\sqrt{2/f_1}w_1}.$$

Replacing f_1/y_1 in (38) with this expression, and expanding the resultant expression, we have

$$\widehat{\Delta}_A^2 = \Delta^2 + r_1 + \frac{1}{n^{1/2}}D_1 + \frac{1}{n}D_2 + R_2, \tag{39}$$

where

$$\begin{aligned}
r_1 &= \frac{N(p-1)}{N_1N_2} - \frac{p-2}{n}\omega^{-2},\\
D_1 &= D_1(w_1, w_2, z_1)\\
&= -\sqrt{\frac{2n}{f_1}}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}w_1 - 2\sqrt{\frac{Nn}{N_1N_2}}\Delta z_1 + \frac{N(p-1)}{N_1N_2}\sqrt{\frac{2n}{f_2}}w_2,\\
D_2 &= D_2(w_1, w_2, z_1)\\
&= \frac{2n}{f_1}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}w_1^2 + \sqrt{\frac{2n}{f_1}}w_1\left\{2\sqrt{\frac{Nn}{N_1N_2}}\Delta z_1 - \sqrt{\frac{2n}{f_2}}\frac{N(p-1)}{N_1N_2}w_2\right\} + \frac{Nn}{N_1N_2}z_1^2,
\end{aligned}$$

R_2 is a remainder term consisting of $n^{-3/2}$ times a homogeneous polynomial of degree 3 in w_1, w_2 and z_1 of which the coefficients are $O(1)$ under A1, plus n^{-2} times a homogeneous polynomial of degree 4, plus a remainder term that is $O(n^{-5/2})$ under A1 for fixed w_1, w_2 and z_1. From the assumption that $\omega^{-2} = Nn/(N_1N_2) + O(n^{-1})$, we find that

$$r_1 = \frac{N}{N_1N_2} + O(pn^{-2}),$$

which yields that $r_1 = O(n^{-1})$ under A1. It follows from (39) that

$$\begin{aligned}
\widehat{u}_0 &= u_0 + \frac{n}{m+1}(\rho_1\omega^2 - \rho_3\omega)\left(r_1 + n^{-1/2}D_1 + n^{-1}D_2 + R_2\right), \\
\widehat{v}_0 &= v_0 + \frac{n^2(n+1)}{(m+1)^2(m+2)}\frac{Nn}{N_1N_2}\omega^2\left(r_1 + n^{-1/2}D_1 + n^{-1}D_2 + R_2\right)
\end{aligned} \tag{40}$$

Taylor series expansion of $\widehat{v}_0^{-1/2}$ at $\widehat{v}_0 = v_0$ up to the term with the order $O((\widehat{v}_0 - v_0)^5)$ gives that

$$\widehat{v}_0^{-1/2} = v_0^{-1/2}\left\{1 - \frac{\lambda}{2}r_1 - \frac{\lambda}{2n^{1/2}}D_1 + \frac{1}{n}\left(-\frac{\lambda}{2}D_2 + \frac{3}{8}\lambda^2D_1^2\right) + R_3\right\}, \tag{41}$$

where

$$\lambda = \frac{1}{v_0} \frac{n^2(n+1)}{(m+1)^2(m+2)} \frac{Nn}{N_1 N_2} \omega^2,$$

R_3 is a remainder term consisting of $n^{-3/2}$ times a homogeneous polynomial of degree 3 in w_1, w_2 and z_1 of which the coefficients are $O(1)$ under A1, plus $n^{-3/2}$ times a homogeneous polynomial of degree 1, plus n^{-2} times a homogeneous polynomial of degree 4, plus n^{-2} times a homogeneous polynomial of degree 2, plus terms which is $O(n^{-2})$ under A1, plus a remainder term that is $O(n^{-5/2})$ under A1 for fixed w_1, w_2 and z_1. From the expansions (40) and (41), we obtain

$$\begin{aligned}\widehat{y} &= \widehat{v}_0^{-1/2}(x + \widehat{u}_0) \\ &= y + q_1 r_1 + \frac{1}{n^{1/2}} q_1 D_1 + \frac{1}{n} Q_2 + R_4,\end{aligned}$$

where

$$\begin{aligned}q_1 &= \frac{1}{\sqrt{v_0}} \left\{ \frac{n}{m+1}(\rho_1\omega^2 - \rho_3\omega) - \frac{\lambda}{2}(x + u_0) \right\}, \\ Q_2 &= \frac{1}{\sqrt{v_0}} \left\{ \frac{n}{m+1}(\rho_1\omega^2 - \rho_3\omega) \left(D_2 - \frac{\lambda D_1^2}{2} \right) - \left(\frac{\lambda D_2}{2} - \frac{3}{8}\lambda^2 D_1^2 \right)(x + u_0) \right\},\end{aligned}$$

and R_4 is a remainder term which have the same property as R_3. Taylor series expansion of $\Phi(\widehat{y})$ at $\widehat{y} = y$ up to the term with order $O((\hat{y} - y)^5)$ gives

$$\Phi(\widehat{y}) = \Phi(y) + \frac{1}{n^{1/2}} q_1 D_1 \phi(y) + \frac{1}{n} \left(Q_2 - \frac{q_1^2}{2} D_1^2 y + n q_1 r_1 \right) \phi(y) + R_5,$$

where R_5 is a remainder term which have the same property as R_3. Then we have

$$E\left[\Phi(\hat{y})\right] = \Phi(y) - \frac{1}{n} \left(\frac{y}{2} q_1^2 E[D_1^2] - q_2 - n q_1 r_1 \right) \phi(y) + E[R_5], \tag{42}$$

in which we use the fact that $E[D_1] = 0$, and we denote $q_2 = E[Q_2]$. From the same derivation in Anderson [2], it can be shown that $E[R_5] = O(n^{-2})$ under A1. Substituting $y^3 = H_3(y) + 3H_1(y)$, $y^2 = H_2(y) + H_0(y)$, $y = H_1(y)$ and $1 = H_0(y)$ for $(yq_1^2/2)E[D_1^2] - q_2 - nq_1 r_1$, we have

$$\frac{y}{2} q_1^2 E[D_1^2] - q_2 - n q_1 r_1 = \sum_{k=1}^{4} \varepsilon_k(\Delta^2) H_{k-1}(y),$$

where

$$\begin{aligned}\varepsilon_1(\Delta^2) &= -\frac{1}{\sqrt{v_0}} \frac{n}{m+1} (\rho_1\omega^2 - \rho_3\omega)(d_2(\Delta^2) + n r_1), \\ \varepsilon_2(\Delta^2) &= \frac{\lambda}{2}(d_2(\Delta^2) + n r_1) + \frac{1}{2} \frac{1}{v_0} \left(\frac{n}{m+1} \right)^2 (\rho_1\omega^2 - \rho_3\omega)^2 d_1(\Delta^2), \\ \varepsilon_3(\Delta^2) &= -\frac{\lambda}{2} \frac{1}{\sqrt{v_0}} \frac{n}{m+1} (\rho_1\omega^2 - \rho_3\omega) d_1(\Delta^2), \\ \varepsilon_4(\Delta^2) &= \frac{\lambda^2}{8} d_1(\Delta^2).\end{aligned}$$

Here, $d_1(\Delta^2) = E[D_1^2]$ and $d_2(\Delta^2) = E[D_2]$. The expectations of D_1^2 and D_2 are given in the following closed forms.

$$E[D_1^2] = \frac{2n}{f_1}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}^2 + \frac{4Nn}{N_1N_2}\Delta^2 + \frac{2n}{f_2}\left\{\frac{N(p-1)}{N_1N_2}\right\}^2,$$
$$E[D_2] = \frac{2n}{f_1}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\} + \frac{Nn}{N_1N_2}.$$

Let

$$\tilde{\varepsilon}_k\left(\widehat{\Delta}_A^2\right) = \varepsilon_k\left(\widehat{\Delta}_A^2\right)H_{k-1}(\widehat{y})\phi(\widehat{y}).$$

Then the function $\tilde{g}_k\left(\widehat{\Delta}_A^2\right)$ is smooth (for fixed x) on $(-(p-2)\omega^{-2}/n, \infty)$. Taylor series expansion of $\tilde{\varepsilon}_k\left(\widehat{\Delta}_A^2\right)$ at $\widehat{\Delta}_A^2 = \Delta^2$ up to the term with the order $O((\widehat{\Delta}_A^2 - \Delta^2)^3)$ gives

$$\tilde{\varepsilon}_k\left(\widehat{\Delta}_A^2\right) = \tilde{\varepsilon}_k(\Delta^2) + R_6,$$

where R_6 is a remainder term consisting of $n^{-1/2}$ times a homogeneous polynomial of degree 1 in w_1, w_2 and z_1 of which the coefficients are $O(1)$ (for fixed x) under A1, plus n^{-1} times a homogeneous polynomial of degree 2, plus terms which is $O(n^{-1})$, plus a remainder term that is $O(n^{-3/2})$ under A1 for fixed w_1, w_2 and z_1 and x. From the derivation in Anderson [2] again, we find that $E[R_6] = O(n^{-1})$ under A1, and so

$$E\left[\varepsilon_k\left(\widehat{\Delta}_A^2\right)H_{k-1}(\hat{y})\phi(\hat{y})\right] = \varepsilon_k\left(\widehat{\Delta}^2\right)H_{k-1}(y)\phi(y) + O(n^{-1}). \tag{43}$$

Using the same derivation on the above, it can be shown that

$$E\left[\widehat{b}_kH_{k-1}(\widehat{y})\phi(\hat{y})\right] = b_kH_{k-1}(y)\phi(y) + O(n^{-1}). \tag{44}$$

Hence, under A1,

$$E\left[\Phi(\widehat{y}) + \frac{1}{n}\sum_{k=1}^{4}(\widehat{\varepsilon}_k - \widehat{b}_k)H_{k-1}(\widehat{y})\phi(\widehat{y})\right] = \Phi(y) - \frac{1}{n}\sum_{k=1}^{4}b_kH_{k-1}(y)\phi(y) + O(n^{-2}),$$

where $\widehat{\varepsilon}_k = \varepsilon_k\left(\widehat{\Delta}_A^2\right)$. □

Acknowledgements

The first author is partially supported by Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (Wakate B), 26800088, 2014-2016. The third author's research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), 16K00047, 2016-2018.

References

[1] Anderson, T.W. (1958). *An Introduction of Multivariate Statistical Analysis 1st ed.* Wiley, NY.

[2] Anderson, T.W. (1973). An asymptotic expansion of the distribution of the Studentized classification statistic W. *Ann. Statist.*, **1**, 964–972.

[3] Anderson, T.W. (2003). *An Introduction of Multivariate Statistical Analysis 3rd ed.* Wiley, Hoboken.

[4] Fujikoshi, Y. (1983). A criterion for variable selection in multiple discriminant analysis. *Hiroshima Math. J.*, **13**, 203–214.

[5] Fujikoshi, Y. (1985). Selection of variables in two-group discriminant analysis by error rate and Akaike's information criteria. *J. Multivariate Anal.*, **17**, 27–37.

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

[7] Fujikoshi, Y. (2002). Selection of variables for discriminant analysis in a high-dimensional case. *Sankhyā*, **64**, 256–267.

[8] Fujikoshi, Y. and Kanazawa, M. (1976). The ML classification statistic in covariate discriminant analysis and its asymptotic expansions. *Essays in Probability and Statistics*, 305–320.

[9] 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.

[10] Fujikoshi, Y., Ulyanov, V.V. and Shimizu, R. (2010). *Multivariate Statistics High-Dimensional and Large-Sample Approximations*, Wiley, Hoboken, NJ.

[11] Hyodo, M. and Kubokawa, T. (2014). A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data. *J. Multivariate Anal.*, **123**, 364–379.

[12] James, G.S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **41**, 19-43.

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

[14] Matsumoto, C. (2004). An optimal discriminant rule in the class of linear and quadratic discriminant functions for large dimension and samples. *Hiroshima Math. J.*, **34** 231–250.

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

[16] McLachlan, G.J. (1976). A criterion for selecting variables for the linear discriminant function. *Biometrics*, **32**, 529–534.

[17] McLachlan, G.J. (1980). On the relationship between the F test and the overall error rate for variable selection in two-group discriminant analysis. *Biometrics*, **36**, 501–510.

[18] McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* Wiley, NY.

[19] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, **34**, 1286–1301.

[20] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistics. *Handbook of Statist.* (P. R. Krishnaiah and L. N. Kanal, eds.), North-Holland, 47–60.

[21] Siotani, M. and Wang, R.H. (1975). Further expansion formulae for error rates and comparison of the W- and Z-procedures in discriminant analysis. Technical Report No. 33, Department of Statistics, Kansas State University, Manhattan, Kansas.

[22] 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. *Multivariate Analysis-IV* (P.R.Krishnaiah, ed.), North-Holland, Amsterdam, 523–545.

[23] Yamada, T., Himeno, T. and Sakurai, T. (2015). Cut-off point of linear discriminant rule for large dimension. TR 15-04, Hiroshima Statistical Research Group, Hiroshima University.

[24] Wyman, F.J., Young, D.M. and Turner, D.W. (1990). A comparison of asymptotic error rate expansions for the sample linear discriminant function. *Pattern Recognition*, **23**, 775–783.