Estimation of misclassification probability for a distance-based classifier in high-dimensional data

Yuki Yamada[‡], Masashi Hyodo[†] and Takashi Seo[‡]

[†] *Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjyuku-ku, Tokyo 162-8601, Japan. E-Mail: hyodoh_h@yahoo.co.jp*
[‡] *Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162-8601, Japan.*

Abstract

The Euclidean distance-based classifier is often used to classify an observation into one of several populations in high-dimensional data. One of the most important problems in discriminant analysis is estimating the probability of misclassification. In this paper, we propose a consistent estimator of misclassification probabilities when the dimension of the vector, p, may exceed the sample size, N, and the underlying distribution need not necessarily be normal. A new estimator of quadratic form is also obtained as a by-product. Finally, we numerically verify the high accuracy of our proposed estimator in finite sample applications, inclusive of high-dimensional scenarios.

AMS 2000 subject classification: 62H30, 41A60.

Key words:
asymptotic normality, Euclidean distance-based classifier, high-dimensional data, probability of misclassification.

1. Introduction

In this paper, we focus on a discrimination problem that is concerned with the allocation of a given object, $\boldsymbol{x}$, a random vector represented by a set of features $(x_1, \ldots, x_p)$, to one of two populations, Π_1 and Π_2. Let $\boldsymbol{x}$ be an observation vector into one of the two population groups Π_1 and Π_2. Then, we assume that

$$\boldsymbol{x} = \Gamma^{(g)}\boldsymbol{z} + \boldsymbol{\mu}^{(g)} \ (g = 1, 2).$$

Further, let $\boldsymbol{x}_1^{(g)}, \boldsymbol{x}_2^{(g)}, \ldots, \boldsymbol{x}_{N_g}^{(g)}$ be p-dimensional observation vectors from the g-th population Π_g such that

$$\boldsymbol{x}_j^{(g)} = \Gamma^{(g)} \boldsymbol{z}_j^{(g)} + \boldsymbol{\mu}^{(g)} \ (j = 1, \ldots, N_g,\ g = 1, 2).$$

Here, $\Gamma^{(g)}\Gamma^{(g)\prime} = \Sigma^{(g)} (\geq \mathrm{O})$ and $\boldsymbol{z} = (z_1, \ldots, z_p)'$ and $\boldsymbol{z}_j^{(g)} = (z_{1j}^{(g)}, \ldots, z_{pj}^{(g)})'$ are independent and identically distributed (i.i.d.) random vectors such that $\mathrm{E}[\boldsymbol{z}] = \mathrm{E}[\boldsymbol{z}_j^{(g)}] = \boldsymbol{0}$ and $\mathrm{Var}[\boldsymbol{z}] = \mathrm{Var}[\boldsymbol{z}_j^{(g)}] = I_p$.
In our study, we consider two cases, (C1) and (C2), as follows.
(C1) ${}^{\exists}\kappa_{4i}^{(g)},\ \kappa_{4i},\ \kappa_{4\max}^{(g)}, \kappa_{4\max} \in (0, \infty)$ such that

$$\begin{aligned}
&\mathrm{E}[z_i^4] = \kappa_{4i} + 3 \leq \kappa_{4\max} + 3, \\
&\mathrm{E}[z_{ij}^{(g)^4}] = \kappa_{4i}^{(g)} + 3 \leq \kappa_{4\max}^{(g)} + 3 \ (i = 1, \ldots, p), \\
&\mathrm{E}[z_i^2 z_k^2] = \mathrm{E}[z_{ij}^{(g)^2} z_{kj}^{(g)^2}] = 1, \\
&\mathrm{E}[z_i z_k z_l z_m] = \mathrm{E}[z_{ij}^{(g)} z_{kj}^{(g)} z_{lj}^{(g)} z_{mj}^{(g)}] = 0 \ (i \neq k, l, m).
\end{aligned}$$

(C2) z_{ij} and $z_{ij}^{(g)}$ are independent for i, j, g, and ${}^{\exists}\kappa_{4i}^{(g)},\ \kappa_{4i},\ \kappa_{4\max}^{(g)},\ \kappa_{4\max} \in (0, \infty)$ such that

$$\mathrm{E}[z_i^4] = \kappa_{4i} + 3 \leq \kappa_{4\max} + 3 \text{ and } \mathrm{E}[z_{ij}^{(g)^4}] = \kappa_{4i}^{(g)} + 3 \leq \kappa_{4\max}^{(g)} + 3.$$

Here, (C1) is a weaker condition than (C2). However, under (C2), assumptions about the mean vector and covariance become weak.

We are interested in investigating the discrimination procedure that can accommodate $p > \max\{N_1, N_2\}$ cases, with the main focus on the performance accuracy in the asymptotic framework that allows p to grow together with N_1 and N_2. Recently, Aoshima and Yata (2014) considered the Euclidean distance-based classifier for the high-dimensional multi-class problem with different class covariance matrices. Aoshima and Yata (2014) proposed the Euclidean distance discriminant function as

$$W = \left\{2\boldsymbol{x} - \left(\overline{\boldsymbol{x}}^{(1)} + \overline{\boldsymbol{x}}^{(2)}\right)\right\}' \left(\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}\right) - \frac{1}{N_2}\mathrm{tr}S^{(2)} + \frac{1}{N_1}\mathrm{tr}S^{(1)}, \quad (1.1)$$

where

$$\overline{\boldsymbol{x}}^{(g)} = \frac{1}{N_g}\sum_{i=1}^{N_g} \boldsymbol{x}_i^{(g)},\ S^{(g)} = \frac{1}{N_g - 1}\sum_{i=1}^{N_g} (\boldsymbol{x}_i^{(g)} - \overline{\boldsymbol{x}}^{(g)})(\boldsymbol{x}_i^{(g)} - \overline{\boldsymbol{x}}^{(g)})' \ (g = 1, 2).$$

Then, the Euclidean distance discriminant rule given by W assigns a new observation $\boldsymbol{x}$ to Π_1 if $W > c$, and to Π_2 otherwise, where c is an appropriate cut-off point. In particular, Aoshima and Yata (2014) derived asymptotic conditions to ensure that the expected misclassification error converges to zero and obtained an asymptotic approximation of the misclassification probability.

In this study, we focus on the misclassification probability of the Euclidean distance discriminant rule. For a specific c, the performance accuracy of the Euclidean distance discriminant rule will be represented by the resulting pair of misclassification error probabilities. More specifically, we define the misclassification probability of the Euclidean distance discriminant rule by

$$e(2|1) = \Pr(W \leq c|\boldsymbol{x} \in \Pi_1),\ e(1|2) = \Pr(W > c|\boldsymbol{x} \in \Pi_2).$$

Our main objective is to derive the limiting value of the misclassification probability and propose a consistent and asymptotically unbiased estimator in high-dimensional settings. In general, it is difficult to obtain the exact value of the misclassification probability. Many studies have attempted to obtain asymptotic approximations for the misclassification probability of the Fisher linear discriminant rule when $p < N_1 + N_2 - 2$ under a framework where N_1 and N_2 are large and p is fixed. For a review of these results, see, e.g., Okamoto (1963, 1968) and Siotani (1982). An asymptotic approximations under a framework where N_1, N_2, and p are all large have also been studied (see, e.g., Lachenbruch (1968) and Fujikoshi and Seo (1998)). Fujikoshi (2000) obtained an explicit formula of error bounds for an approximation of the misclassification probability. Further, Konishi and Honda (1990) and Kubokawa et al. (2013) deal with estimation of misclassification error probabilities of the Fisher linear discriminant rule. Recently, Aoshima and Yata (2014) showed the asymptotic normality of the Euclidean distance discriminant rule under the high-dimensional asymptotic framework

$$\text{(A0)} \qquad N_1,\ N_2,\ p \to \infty$$

and some assumptions, which represents the relationship between the quadratic forms $\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}$ and the sum of traces (see (A2') in Section 2 for details). Here, $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. In this paper, we derive the limiting value of the misclassification probability under (C1) and the above assumptions or under (C2) and assumptions that are weaker than the above assumptions. By deriving an

estimator of unknown values among the limiting values of misclassification probabilities, we propose a consistent and asymptotically unbiased estimator of misclassification probabilities. Further, we derive the unbiased estimator of $\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}$ as a by-product.

The remainder of this paper is organized as follows. In Section 2, we show the asymptotic normality of the Euclidean distance discriminant rule. In Section 3, we derive a consistent estimator of misclassification probabilities. Further, the limiting approximations of the defined cut-off point are established by using this estimator. Section 4 summarizes the results of numerical experiments conducted to verify the validity of the proposed estimators along with a number of high-dimensional scenarios where p far exceeds the sample size. Finally, we conclude the paper in Section 5 and present some auxiliary lemmas in the Appendix.

2. Asymptotic normality of Euclidean distance-based classifier

In this section, we show the asymptotic normality of the Euclidean distance-based classifier W. It is difficult to obtain the exact distribution of the Euclidean distance-based classifier. We assume the high-dimensional asymptotic framework (A0) and also make the following assumptions:

$$\text{(A1)} \quad \lim_{p\to\infty} \frac{\mathrm{tr}\Sigma^{(g)^4}}{\left(\mathrm{tr}\Sigma^{(g)^2}\right)^2} \to 0, \quad 0 < \lim_{p\to\infty} \frac{\mathrm{tr}\Sigma^{(h)}\Sigma^{(i)}}{\mathrm{tr}\Sigma^{(g)^2}} < \infty \quad (g,h,i=1,2),$$

$$\text{(A2)} \quad \lim_{N_1,N_2,p\to\infty} \frac{\boldsymbol{\delta}'\Sigma^{(g')}\boldsymbol{\delta}}{N_{g'}\sigma_g^2} \to 0 \quad (g,g'=1,2,\ g\neq g'),$$

$$\text{(A3)} \quad \lim_{p,N_1,N_2\to\infty} \frac{\max\{\gamma_1^{(g)^2},\ldots,\gamma_p^{(g)^2}\}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}{\sigma_g^4} \to 0 \quad (g=1,2),$$

where $\gamma_i^{(g)}$ $(i=1,\ldots,p)$ denotes i-th element of $\Gamma^{(g)'}\boldsymbol{\delta}$, and

$$\sigma_g^2 \;=\; 4\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta} + 4\frac{1}{N_g}\mathrm{tr}\Sigma^{(g)^2} + 4\frac{1}{N_{g'}}\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)} \ (\ g,g'=1,\ 2,\ g\neq g').$$

Aoshima and Yata (2014) proved asymptotic normality by assuming (C1), (A0), (A1) and

$$\text{(A2')} \lim_{p,N_1,N_2\to\infty} \frac{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}{\delta_g^2} \to 0 \ \ (g=1,2)$$

instead of (A2) and (A3). Here,

$$\delta_g^2 = \frac{4\mathrm{tr}\Sigma^{(g)^2}}{N_g} + \frac{4\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}}{N_{g'}} + \frac{2\mathrm{tr}\Sigma^{(1)^2}}{N_1(N_1-1)} + \frac{2\mathrm{tr}\Sigma^{(2)^2}}{N_2(N_2-1)}$$

for $g, g' = 1, 2$, $g \neq g'$. Note that assumptions (A2) and (A3) are weaker than assumption (A2'). Assuming stronger condition (C2) than the condition (C1), we can relax the assumption (A2'). The following theorem establishes the asymptotic normality of W not only under (C1), (A0), (A1) and (A2'), but also under (C2), (A0)-(A3).

Theorem 2.1. *Under assumptions* (C1),(A0),(A1) *and* (A2') *or assumptions* (C2),(A0)-(A3)*, it holds that*

$$\frac{W + (-1)^g \boldsymbol{\delta}'\boldsymbol{\delta}}{\sigma_g} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. Let $T = W + (-1)^g \boldsymbol{\delta}'\boldsymbol{\delta}$ and decompose T as $T = T_1 + T_2$, where

$$\begin{aligned} T_1 &= 2(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\boldsymbol{\delta} + 2(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\{(\overline{\boldsymbol{x}}^{(1)} - \boldsymbol{\mu}^{(1)}) - (\overline{\boldsymbol{x}}^{(2)} - \boldsymbol{\mu}^{(2)})\}, \\ T_2 &= -\frac{1}{N_1(N_1-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_1} (\boldsymbol{x}_j^{(1)} - \boldsymbol{\mu}^{(1)})'(\boldsymbol{x}_k^{(1)} - \boldsymbol{\mu}^{(1)}) \\ &\quad + \frac{1}{N_2(N_2-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_2} (\boldsymbol{x}_j^{(1)} - \boldsymbol{\mu}^{(1)})'(\boldsymbol{x}_k^{(1)} - \boldsymbol{\mu}^{(1)}) - 2\boldsymbol{\delta}'(\overline{\boldsymbol{x}}^{(g')} - \boldsymbol{\mu}^{(g')}). \end{aligned}$$

First, we show the asymptotic normality of T under (C1),(A0),(A1) and (A2'). From Lemma A.2, it holds that

$$T = \sum_{i=1}^{N_1+N_2} \psi_i + o_p(\sigma_g),$$

where

$$\begin{aligned} \psi_i &= \frac{2}{N_1}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_i^{(1)} - \boldsymbol{\mu}^{(1)}) \ (i = 1, \ldots, N_1), \\ \psi_i &= -\frac{2}{N_2}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_{i-N_1}^{(2)} - \boldsymbol{\mu}^{(2)}) \ (i = N_1+1, \ldots, N_1+N_2). \end{aligned}$$

Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \sigma\{\psi_1\}$, $\mathcal{F}_{i-1} = \sigma\{\psi_1, \psi_2, \ldots, \psi_{i-1}\}$. Then it is straightforward to show that $\mathrm{E}[\psi_i] = 0$ and $\mathrm{E}[\psi_i|\mathcal{F}_{i-1}] = 0$. Thus, ψ_i is a martingale difference sequence. Since $\delta_g^2/\sigma_g^2 \to 1$ and Theorem 3 in Aoshima and Yata (2014), we show the asymptotic normality of T under (C1),(A0),(A1) and (A2').

Next, we show the asymptotic normality of T under (C1),(A0)-(A3). From Lemma A.2, $T = T_1 + o(\sigma_g)$ under (C1),(A0)-(A3). Let $\overline{\boldsymbol{y}}^{(g)} = \overline{\boldsymbol{x}}^{(g)} - \boldsymbol{\mu}^{(g)}$, $\overline{\boldsymbol{y}}^{(g)} = \Gamma^{(g)}\overline{\boldsymbol{z}}^{(g)}$ and $\Gamma^{(g)} = H^{(g)}\Lambda^{(g)^{1/2}}$. Here, $H^{(g)}$ is an orthogonal matrix such that $H^{(g)}\Lambda^{(g)}H^{(g)'} = \Sigma^{(g)}$, where $\Lambda^{(g)} = \mathrm{diag}(\lambda_1^{(g)}, \ldots, \lambda_p^{(g)})$ and $\lambda_i^{(g)}$ is i-th eigenvalues of $\Sigma^{(g)}$. Then T_1 can be factorized as $T_1 = \sum_{i=1}^p \epsilon_i$, where

$$\epsilon_i = 2\gamma_i^{(g)} z_i + 2\lambda_i^{(g)} z_i \overline{z}_i^{(g)} - 2\lambda_i^{(g)^{1/2}} z_i \boldsymbol{h}_i^{(g)\prime} \overline{\boldsymbol{y}}^{(g')}.$$

Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_1 = \sigma\{\epsilon_1\}$, $\mathcal{F}_{i-1} = \sigma\{\epsilon_1, \ldots, \epsilon_{i-1}\}$. Then it is straightforward to show that $\mathrm{E}[\epsilon_i] = 0$ and $\mathrm{E}[\epsilon_i|\mathcal{F}_{i-1}] = 0$. Thus, ϵ_i is a martingale difference sequence. To apply the martingale central limit theorem, we need to show that

$$\frac{\sum_{i=1}^p \sigma_{g,i}^2}{\sigma_g^2} \xrightarrow{p} 1 \qquad \text{and} \qquad \frac{\sum_{i=1}^p \mathrm{E}[\epsilon_i^4]}{\sigma_g^4} \to 0, \tag{2.1}$$

where $\sigma_{g,i}^2 = \mathrm{E}[\epsilon_i^2|\mathcal{F}_{i-1}]$.

We show the first part of (2.1). Note that

$$\sigma_{g,i}^2 = 4\gamma_i^{(g)^2} + \frac{4\lambda_i^{(g)^2}}{N_g} + 4\lambda_i^{(g)} \overline{\boldsymbol{y}}^{(g')\prime} \boldsymbol{h}_i^{(g)} \boldsymbol{h}_i^{(g)\prime} \overline{\boldsymbol{y}}^{(g')} - 8\gamma_i \lambda_i^{(g)^{1/2}} \boldsymbol{h}_i^{(g)\prime} \overline{\boldsymbol{y}}^{(g')},$$

and

$$\mathrm{E}\left[\sum_{i=1}^p \sigma_{g,i}^2\right] = \sigma_g^2.$$

We need to show that $\mathrm{Var}[R_1] = o(\sigma_g^4)$ and $\mathrm{Var}[R_2] = o(\sigma_g^4)$, where

$$R_1 = \overline{\boldsymbol{y}}^{(g')\prime} \Sigma^{(g)} \overline{\boldsymbol{y}}^{(g')}, \; R_2 = \boldsymbol{\delta}' \Sigma^{(g)} \overline{\boldsymbol{y}}^{(g')}.$$

$\mathrm{Var}[R_1]$ is given by

$$\begin{aligned} \mathrm{Var}[R_1] &= \mathrm{E}[R_1^2] - \frac{1}{N_{g'}^2}\{\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})\}^2 \\ &\leq \frac{1}{N_{g'}^2}\left(\frac{\kappa_{4\max}^{(g')}}{N_{g'}} + 2\right)\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})^2 = o(\sigma_g^4). \end{aligned}$$

By applying the Cauchy-Schwarz inequality,

$$\boldsymbol{\delta}'\Sigma^{(g)}\Sigma^{(g')}\Sigma^{(g)}\boldsymbol{\delta} \leq \boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\sqrt{\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})^2}.$$

Thus, $\mathrm{Var}[R_2]$ is given by

$$\mathrm{Var}[R_2] = \mathrm{E}[R_2^2] = \frac{1}{N_{g'}}\boldsymbol{\delta}'\Sigma^{(g)}\Sigma^{(g')}\Sigma^{(g)}\boldsymbol{\delta} = o(\sigma_g^4).$$

Hence, the proof for the first part of (2.1) is complete.

To show the second part of (2.1) we decompose ϵ_i into the sum of three parts,

$$\epsilon_i = \epsilon_{i1} + \epsilon_{i2} + \epsilon_{i3},$$

where

$$\epsilon_{i1} = \gamma_i^{(g)} z_i,\ \epsilon_{i2} = 2\lambda_i^{(g)} z_i \overline{z}_i^{(g)},\ \epsilon_{i3} = 2\lambda_i^{(g)^{1/2}} z_i \boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(g')}.$$

By applying Hölder's inequality, $\mathrm{E}[\epsilon_i^4] \leq 27\mathrm{E}[\epsilon_{i1}^4 + \epsilon_{i2}^4 + \epsilon_{i3}^4]$. Thus, we need to show that $\sum_{i=1}^p \mathrm{E}[\epsilon_{i\ell}^4] = o(\sigma_g^4)$ for $\ell = 1, 2, 3$. Note that

$$\begin{aligned}
\sum_{i=1}^p \mathrm{E}[\epsilon_{i1}^4] &\leq 16(\kappa_{4\max}^{(g)} + 3)\max\{\gamma_1^{(g)^2}, \ldots, \gamma_p^{(g)^2}\}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta} = o(\sigma_g^4), \\
\sum_{i=1}^p \mathrm{E}[\epsilon_{i2}^4] &\leq \frac{16}{N_g^2}(\kappa_{4\max}^{(g)} + 3)\left(\frac{\kappa_{4\max}^{(g)}}{N_g} + 3\right)\mathrm{tr}\Sigma^{(g)^4} = o(\sigma_g^4), \\
\sum_{i=1}^p \mathrm{E}[\epsilon_{i3}^4] &\leq \frac{16}{N_{g'}^2}(\kappa_{4\max}^{(g)} + 3)\left(\frac{\kappa_{4\max}^{(g')}}{N_{g'}} + 3\right)\sum_{i=1}^p \lambda_i^{(g)^2}(\boldsymbol{h}_i^{(g)\prime}\Sigma^{(g')}\boldsymbol{h}_i^{(g)})^2 \\
&\leq \frac{16}{N_{g'}^2}(\kappa_{4\max}^{(g)} + 3)\left(\frac{\kappa_{4\max}^{(g')}}{N_{g'}} + 3\right)\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})^2 = o(\sigma_g^4).
\end{aligned}$$

This proves the second part of (2.1) and completes the proof of the asymptotic normality of W under (C2),(A0)-(A3). □

3. Estimation of misclassification probability

By using Theorem 2.1, we obtain the limiting values of misclassification probabilities as

$$e(g'|g) = \Phi(w_g) + o(1), \tag{3.1}$$

where

$$w_g = -\frac{\boldsymbol{\delta}'\boldsymbol{\delta} + (-1)^g c}{\sigma_g} \ (g, g' = 1, 2, g \neq g').$$

The limiting values (3.1) include the unknown values $\boldsymbol{\delta}'\boldsymbol{\delta}$ and σ_g. We use an unbiased estimator of $\boldsymbol{\delta}'\boldsymbol{\delta}$:

$$\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}} \ = \ (\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)})'(\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}) - \frac{\mathrm{tr}S^{(1)}}{N_1} - \frac{\mathrm{tr}S^{(2)}}{N_2}.$$

The unbiased estimator $\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}}$ has been used in two sample tests (Chen and Qin (2010), Aoshima and Yata (2011)). Now consider the estimator of σ_g. We define the unbiased estimators of $\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}$, $\mathrm{tr}\Sigma^{(g)^2}$ and $\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}$ as follows:

$$\begin{aligned}
\widehat{\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}} &= \mathrm{tr}S^{(1)}S^{(2)}, \\
\widehat{\mathrm{tr}\Sigma^{(g)^2}} &= \frac{N_g - 1}{N_g(N_g-2)(N_g-3)}\Big\{(N_g-1)(N_g-2)\mathrm{tr}S^{(g)^2} + (\mathrm{tr}S^{(g)})^2 \\
&\quad -N_g Q^{(g)}\Big\} \ (g = 1, 2), \\
\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}} &= \frac{1}{(N_g-1)(N_g-2)}\left\{(N_g-2)V^{(g)} - 2U^{(g)}\right\} - \frac{1}{N_{g'}}\mathrm{tr}S^{(g)}S^{(g')} \\
&\quad + \frac{1}{N_g(N_g-2)(N_g-3)}\Big\{2N_g Q^{(g)} - (N_g-1)\left(\mathrm{tr}S^{(g)}\right)^2 \\
&\quad -(N_g-1)^2\mathrm{tr}S^{(g)^2}\Big\} \ (g, g' = 1, 2, \ g \neq g'),
\end{aligned}$$

where

$$\begin{aligned}
Q^{(g)} &= \frac{1}{N_g - 1}\sum_{j=1}^{N_g}\left\{(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})'(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})\right\}^2, \\
V^{(g)} &= \sum_{j=1}^{N_g}\left\{(\overline{\boldsymbol{x}}^{(g)} - \overline{\boldsymbol{x}}^{(g')})'(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})\right\}^2, \\
U^{(g)} &= \sum_{j=1}^{N_g}(\overline{\boldsymbol{x}}^{(g)} - \overline{\boldsymbol{x}}^{(g')})'(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})'(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)}).
\end{aligned}$$

The unbiased estimator $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ was proposed by Himeno and Yamada (2014) and Srivastava et al. (2014), and they showed the consistency of this estimator. Further, note that $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ is the same as that proposed by Chen et

al. (2010). In this paper, we derive the unbiased estimator $\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}$, and we investigate the leading term of variance of these estimators (see Appendix). By using these estimators, we propose the estimator of σ_g. We provide the truncated estimator

$$\max\left\{\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g},0\right\} \tag{3.2}$$

so that the estimator of σ_g may be negative. Then it holds that

$$\begin{aligned}&\left|\max\left\{\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g},0\right\}-\left(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}+\frac{\mathrm{tr}\Sigma^{(g)^2}}{N_g}\right)\right| \\ &\leq\left|\left(\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g}\right)-\left(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}+\frac{\mathrm{tr}\Sigma^{(g)^2}}{N_g}\right)\right| \quad a.s.\end{aligned} \tag{3.3}$$

From (iii) and (iv) in Lemma A.3, it holds that

$$\frac{\left|\left(\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g}\right)-\left(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}+\frac{\mathrm{tr}\Sigma^{(g)^2}}{N_g}\right)\right|}{\sigma_g}\xrightarrow{p}0 \tag{3.4}$$

under (C1), (A0) and (A1) or (C2), (A0) and (A1). From (3.3) and (3.4), it holds that

$$\frac{\left|\max\left\{\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g},0\right\}-\left(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}+\frac{\mathrm{tr}\Sigma^{(g)^2}}{N_g}\right)\right|}{\sigma_g}\xrightarrow{p}0 \tag{3.5}$$

under (C1), (A0) and (A1) or (C2), (A0) and (A1). By assigning the truncated estimator (3.2) to the portion of σ_g that may be negative, we propose

$$\widehat{\sigma_g^2}=4\max\left\{\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g},0\right\}+\frac{4}{N_{g'}}\widehat{\mathrm{tr}\Sigma^{(g)}\Sigma^{(g')}}.$$

From (ii) in Lemma A.3 and (3.5), under (C1), (A0) and (A1) or (C2), (A0) and (A1),

$$\frac{\widehat{\sigma_g^2}}{\sigma_g^2}\xrightarrow{p}1. \tag{3.6}$$

By replacing the unknown values in (3.1) with their estimators $\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}}$ and $\widehat{\sigma}_g^2$, we can propose $\widehat{e(g'|g)} = \Phi\left(\widehat{w}_g\right)$ $(g, g' = 1, 2, g \neq g')$, where

$$\widehat{w}_g = -\frac{\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}} + (-1)^g c}{\widehat{\sigma}_g}.$$

The following lemma provides the asymptotic properties of the estimator $\widehat{e(g'|g)}$.

Lemma 3.1. *Under* (C1), (A0) *and* (A1) *or* (C2), (A0) *and* (A1), *it holds that*

$$\widehat{e(g'|g)} - \Phi\left(w_g\right) = o_p(1).$$

Proof. First, we show statement when $\lim_{N_1,N_2,p\to\infty} |w_g| = \infty$. Then, from (i) in Lemma A.3,

$$\frac{\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}} + (-1)^g c}{\boldsymbol{\delta}'\boldsymbol{\delta} + (-1)^g c} \xrightarrow{p} 1 \tag{3.7}$$

under (C1), (A0) and (A1) or (C2), (A0) and (A1). From (3.6) and (3.7),

$$\frac{|\widehat{w}_g - w_g|}{|w_g|} \xrightarrow{p} 0.$$

For$^\forall \varepsilon \in (0, \infty)$,

$$P(|\Phi(\widehat{w}_g) - \Phi(w_g)| > \varepsilon) = J_1 + J_2,$$

where

$$\begin{aligned} J_1 &= P(\{|\widehat{w}_g - w_g| > \xi|w_g|\} \cap \{|\Phi(\widehat{w}_g) - \Phi(w_g)| > \varepsilon\}), \\ J_2 &= P(\{|\widehat{w}_g - w_g| \leq \xi|w_g|\} \cap \{|\Phi(\widehat{w}_g) - \Phi(w_g)| > \varepsilon\}). \end{aligned}$$

Here, ξ is some positive constant that satisfies $\xi \in (0, 1)$. Then, $J_1 \to 0$ under (C1), (A0) and (A1) or (C2), (A0) and (A1). Now, we evaluate J_2 when $w > 0$. It can be expressed as

$$\begin{aligned} |\Phi(\widehat{w}_g) - \Phi(w_g)| &\leq \xi|w_g|\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(w_g - \xi|w_g|)^2\right] \\ &= \frac{\xi}{\sqrt{2\pi}}|w_g|\exp\left[-\frac{1}{2}(1-\xi)^2|w_g|^2\right]. \end{aligned}$$

The right-hand side of the above inequality converges to 0. Thus, $J_2 \to 0$ under (C1), (A0) and (A1) or (C2), (A0) and (A1). Similarly, we can prove that $J_2 \to 0$ when $w_g \leq 0$. Thus, we get $|\Phi(\widehat{w}_g) - \Phi(w_g)| = o_p(1)$.

Next, we show statement when $\lim_{N_1,N_2,p\to\infty} |w_g| < \infty$. From (i)-(iv) in Lemma A.3, under (C1), (A0) and (A1) or (C2), (A0) and (A1),

$$-\frac{\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}} + (-1)^g c}{\sigma_g} \xrightarrow{p} w_g^*, \qquad \frac{\widehat{\sigma_g^2}}{\sigma_g^2} \xrightarrow{p} 1,$$

where $w_g^* = \lim_{N_1,N_2,p\to\infty} w_g$. From the above results, we get $\widehat{w}_g \xrightarrow{p} w_g^*$. Then, by using the continuous mapping theorem, we get $\Phi(\widehat{w}_g) \xrightarrow{p} \Phi\left(w_g^*\right)$. Thus, the proof is complete. □

Using Lemma 3.1 and (3.1), we obtain the following theorem.

Theorem 3.1. *Under* (C1),(A0),(A1) *and* (A2') *or* (C2),(A0)-(A3)*, it holds that*

$$\widehat{e(g'|g)} - e(g'|g) = o_p(1).$$

Proof. From Lemma 3.1 and (3.1), under (C1),(A0),(A1) and (A2') or (C2),(A0)-(A3), it holds that

$$\begin{aligned} |e(g'|g) - \Phi(\widehat{w}_g)| &= |(e(g'|g) - \Phi(w_g)) - (\Phi(\widehat{w}_g) - \Phi(w_g))| \\ &\leq |e(g'|g) - \Phi(w_g)| + |\Phi(\widehat{w}_g) - \Phi(w_g)| \\ &= o_p(1). \end{aligned}$$

Thus, the proof is complete. □

We assume $\lim_{N_1,N_2,p\to\infty} |w_g| < \infty$. By applying Lebesgue's dominated convergence theorem to Theorem 3.1 since

$$\left|\widehat{e(g'|g)} - e(g'|g)\right| < 1 \ a.s.,$$

we get the following corollary.

Corollary 3.1. *Under* (C1), (A0), (A1), (A2') *and* $\lim_{N_1,N_2,p\to\infty} |w_g| < \infty$ *or* (C2), (A0)-(A3) *and* $\lim_{N_1,N_2,p\to\infty} |w_g| < \infty$*, it holds that*

$$\mathrm{E}\left[\widehat{e(g'|g)}\right] = e(g'|g) + o(1).$$

In many practical problems, certain types of misclassification probabilities are generally regarded as more serious than others, e.g., medical applications associated with the diagnosis of diseases. In such cases, it might be desirable to determine the cut-off c to obtain a specified probability of error, or at least to approximate a specified probability. Then, one might base the choice of c on the misclassification probability. This method, denoted in what follows by **M**, proposes that the cut-off point c be set such that

$$\mathbf{M}: \quad e(g'|g) = \alpha,$$

where α is a value derived experimentally. From the results of Theorem 3.1, the **M**-based cut-off point for the Euclidean distance discriminant rule using W is given by

$$\hat{c}_g = (-1)^{-g-1}\left(\widehat{\sigma}_g z_\alpha + \widehat{\boldsymbol{\delta}'\boldsymbol{\delta}}\right),$$

where z_α is the α-percentile of $\mathcal{N}(0,1)$ and $\alpha \in (0,1)$. Then the following theorem holds.

Corollary 3.2. *Let us consider the classification rule*

$$W(\boldsymbol{x}) > (\text{resp.} \leq) \hat{c}_g \;\Rightarrow\; \boldsymbol{x} \in \Pi_1 (\text{resp.} \Pi_2).$$

Then, under assumptions (C1),(A0),(A1) *and* (A2') *or assumptions* (C2),(A0)-(A3)*, and* $\boldsymbol{\delta}'\boldsymbol{\delta}/\sigma_g < \infty$*, it holds that* $e(g'|g) \to \alpha$*.*

Proof. Under assumptions (C1),(A0),(A1) and (A2') or assumptions (C2),(A0)-(A3) and $\boldsymbol{\delta}'\boldsymbol{\delta}/\sigma_g < \infty$, it holds that

$$\frac{\hat{c}_g}{\sigma_g} \xrightarrow{p} (-1)^{-g-1}\left(z_\alpha - \frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\sigma_g}\right).$$

The above result and Theorem 2.1 imply that $e(g'|g) \to \alpha$. □

4. Numerical results

Now, we investigate the numerical performance of the approximation which is based on (3.1) and the consistent estimator $\Phi(\widehat{w}_g)$, via Monte Carlo simulation.

First, we investigate the accuracy of the asymptotic approximations

$$\text{(YHS)} \ : \ e(2|1) \approx \Phi\left(-\frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\sigma_1}\right), \quad \text{(AY)} \ : \ e(2|1) \approx \Phi\left(-\frac{\boldsymbol{\delta}'\boldsymbol{\delta}}{\delta_1}\right).$$

Here, the approximation (YHS) represents our proposed method based on (3.1), and the approximation (AY) represents the method proposed by Aoshima and Yata (2014). The misclassification probability $e(2|1)$ is calculated via simulation with 100,000 replications, where in each step, the data sets are generated as

$$\begin{aligned}
&\text{(Case I)} \quad : \boldsymbol{x}_1^{(g)}, \boldsymbol{x}_2^{(g)}, \ldots, \boldsymbol{x}_{N_g}^{(g)} \overset{i.i.d.}{\sim} \mathcal{N}_p(\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}) \ (g = 1, 2), \\
&\text{(Case II)} \quad : \boldsymbol{x}_1^{(g)}, \boldsymbol{x}_2^{(g)}, \ldots, \boldsymbol{x}_{N_g}^{(g)} \overset{i.i.d.}{\sim} t_p(\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}, \nu) \ (g = 1, 2),
\end{aligned}$$

where $t_p(\boldsymbol{\mu}, \Sigma, \nu)$ denotes a p-variate t-distribution with mean $\boldsymbol{\mu}$, covariance matrix Σ and degrees of freedom ν, $\boldsymbol{\mu}^{(1)} = \mathbf{0}$ and $\boldsymbol{\mu}^{(2)} = (1, \ldots, 1, 0, \ldots, 0)'$ the first $\lceil \sqrt{\mathrm{tr}\Sigma^{(1)^2}} \rceil$ elements of which are 1 or $\boldsymbol{\mu}^{(2)} = (\sqrt{10/p}, \ldots, \sqrt{10/p})'$. Here,

$$\Sigma^{(1)} = B\left((0.3)^{|i-j|^{1/3}}\right)B, \ \Sigma^{(2)} = 1.2B\left((0.3)^{|i-j|^{1/3}}\right)B,$$

where

$$B = \mathrm{diag}\left(\left\{0.5 + \frac{1}{p+1}\right\}^{\frac{1}{2}}, \left\{0.5 + \frac{2}{p+1}\right\}^{\frac{1}{2}}, \ldots, \left\{0.5 + \frac{p}{p+1}\right\}^{\frac{1}{2}}\right).$$

We set $p = 100, 250, 500, 1000$, $(N_1, N_2) = (20, 40)$, $(40, 60)$, $(60, 120)$ and $\nu = 10$. Then we compare the true value $e(2|1)$, the approximation (YHS) and the approximation (AY) on the basis of these settings. The results are shown in Table 1 and 2. By comparing the approximations listed in Table 1 and 2, it can be seen that the approximation (YHS) is closer to the true value $e(2|1)$ than (AY) in most cases. In addition, the approximation (YHS) has stably good result when varying the population distribution and the value of the mean vector $\boldsymbol{\mu}^{(2)}$.

Next we investigate the bias and mean squared error (MSE) of the consistent estimator $\Phi\left(\widehat{w}_1\right)$ on the basis of the same settings. For comparison, we consider the leave-one-out cross-validation method (CV), which is a popular

method for estimating prediction errors for small samples. For $j = 1, \ldots, N_1$, consider the set

$$X_1^{(-j)} = (\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{j-1}^{(1)}, \boldsymbol{x}_{j+1}^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}).$$

This set represents the leave-one-out learning set, which is a collection of data with observation $\boldsymbol{x}_j^{(1)}$ removed. In a prediction problem, it calculates the probability of misclassification for a sample using all other observations in the sample. Using the learning set, we define the discriminant function by

$$W^{(-j)} = \|\boldsymbol{x}_j^{(1)} - \overline{\boldsymbol{x}}^{(2)}\|^2 - \|\boldsymbol{x}^{(1)} - \overline{\boldsymbol{x}}_{(-j)}^{(1)}\|^2 - \mathrm{tr}\left[\frac{S^{(2)}}{N_2} - \frac{S_{(-j)}^{(1)}}{N_1 - 1}\right],$$

where $\overline{\boldsymbol{x}}_{(-j)}^{(1)}$ and $S_{(-j)}^{(1)}$ are calculated using procedures based on (1.1) and the learning set $X_1^{(-j)}$. Then the CV estimator of $e(2|1)$ is given by

$$CV(2|1) = \frac{1}{N_1} \sum_{j=1}^{N_1} I_{\{W^{(-j)}<0\}}(W^{(-j)}),$$

where the function $I_A(x)$ is the indicator function defined as

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

The biases and MSEs of the estimators $CV(2|1)$ and $\Phi(\widehat{w}_1)$ are listed in Table 3-6. From these tables, the both estimators have small biases, and the estimator $\Phi(\widehat{w}_1)$ has smaller MSEs than the estimator $CV(2|1)$ in all cases. Thus, through these simulation experiments, we recommend our suggested estimator in high-dimensional cases.

Finally, we apply our results to a microarray dataset analyzed by Dudoit et al. (2002). The dataset includes information on 72 patients suffering from either Π_1:acute lymphoblastic leukemia (ALL, 47 cases) or Π_2:acute myeloid leukemia (AML, 25 cases), and it was obtained using affymetrix oligonucleotide microarrays. We preprocess the dataset by using the protocol described by Dudoit et al. (2002). The preprocessed dataset comprises 3571 variables. We apply the Euclidean distance discriminant rule with cut-off 0 to this dataset. Using the estimators $\widehat{e(1|2)}$ and $\widehat{e(2|1)}$, we calculate the estimator of misclassification probabilities. The estimate of $e(1|2)$ and $e(2|1)$ is 0.026 and 0.022, respectively.

5. Concluding remarks

We considered the classification problem for high-dimensional data. For high-dimensional data classification, owing to the small number of observations and large number of dimensions, the Fisher linear discriminant rule provides sub-optimal performance corresponding to the singularity and instability of the pooled sample covariance matrix. In such cases, the Euclidean distance-based classifier is often employed. In this paper, we proposed consistent and asymptotically unbiased estimators of misclassification probabilities in high-dimensional settings. Our proposed method has the advantage of establishing under variance heterogeneity and nonnormality. In addition, we performed numerical simulations, which confirmed that this estimator provides accurate approximations.

Appendix

In this section, we state some results on the moments of a random vector $\overline{\boldsymbol{z}}^{(g)} = N_g^{-1}\sum_{j=1}^{N_g}\boldsymbol{z}_j^{(g)}$, the variance of W, and the variances of unbiased estimators $\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}}$, $\widehat{\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}}$, $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ and $\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}$.

Lemma A. 1 (Some results on the moments of a random vector $\overline{\boldsymbol{z}}^{(g)}$)**.** *Let $\boldsymbol{z}_j^{(g)}(j=1,\ldots,N_g)$ be i.i.d. random vectors that satisfy (C1) or (C2). Then for any $p\times p$ positive semidefinite matrices $A=(a_{ij})$ and $B=(b_{ij})$, it holds that*

$$\text{(i)}\quad \mathrm{E}[\overline{z}_i^{(g)^4}] = \frac{\kappa_{4i}^{(g)}+3N_g}{N_g^3},$$

$$\text{(ii)}\quad \mathrm{E}[(\overline{\boldsymbol{z}}^{(g)\prime}A\overline{\boldsymbol{z}}^{(g)})^2] \le \frac{\kappa_{4\max}^{(g)}+2N_g}{N_g^3}\mathrm{tr}A^2+\frac{1}{N_g^2}(\mathrm{tr}A)^2,$$

$$\text{(iii)}\quad \mathrm{E}[(\boldsymbol{z}_j^{(g)\prime}A\boldsymbol{z}_j^{(g)})^2] \le (\kappa_{4\max}^{(g)}+2)\mathrm{tr}A^2+(\mathrm{tr}A)^2,$$

$$\text{(iv)}\quad \mathrm{E}[(\boldsymbol{z}_j^{(g)\prime}A\boldsymbol{z}_k^{(g)})^4] \le (\kappa_{4\max}^{(g)}+3)\left\{(\kappa_{4\max}^{(g)}+2)\mathrm{tr}A^4+(\mathrm{tr}A^2)^2\right\}.$$

Proof. The proof of Lemma A.1 is routine and hence omitted here.

Lemma A. 2 (The variance of W)**.** *The variance of W is*

$$\mathrm{Var}[W] = \sigma_g^2+\sum_{g=1}^{2}\frac{2\mathrm{tr}\Sigma^{(g)^2}}{N_g(N_g-1)}+\frac{4\boldsymbol{\delta}'\Sigma^{(g')}\boldsymbol{\delta}}{N_{g'}},$$

where

$$\sigma_g^2 \;=\; 4\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta} + \frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2} + \frac{4}{N_{g'}}\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}.$$

Proof. Let $\boldsymbol{x} \in \Pi_g$; then, W can be expressed as

$$\begin{aligned} W \;=\;& (-1)^{g-1}\boldsymbol{\delta}'\boldsymbol{\delta} + 2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\boldsymbol{\delta} \\ & +2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\{(\overline{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)}) - (\overline{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})\} \\ & -\frac{1}{N_1(N_1-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_1}(\boldsymbol{x}_j^{(1)}-\boldsymbol{\mu}^{(1)})'(\boldsymbol{x}_k^{(1)}-\boldsymbol{\mu}^{(1)}) \\ & +\frac{1}{N_2(N_2-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_2}(\boldsymbol{x}_j^{(2)}-\boldsymbol{\mu}^{(2)})'(\boldsymbol{x}_k^{(2)}-\boldsymbol{\mu}^{(2)}) - 2\boldsymbol{\delta}'(\overline{\boldsymbol{x}}^{(g')}-\boldsymbol{\mu}^{(g')}). \end{aligned}$$

It is easy to show that $\mathrm{E}[W] = (-1)^{g-1}\boldsymbol{\delta}'\boldsymbol{\delta}$. Let $T = W - (-1)^{g-1}\boldsymbol{\delta}'\boldsymbol{\delta}$ and decompose T as $T = T_1 + T_2$, where

$$\begin{aligned} T_1 \;=\;& 2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\boldsymbol{\delta} + 2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\{(\overline{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)}) - (\overline{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})\}, \\ T_2 \;=\;& -\frac{1}{N_1(N_1-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_1}(\boldsymbol{x}_j^{(1)}-\boldsymbol{\mu}^{(1)})'(\boldsymbol{x}_k^{(1)}-\boldsymbol{\mu}^{(1)}) \\ & +\frac{1}{N_2(N_2-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_2}(\boldsymbol{x}_j^{(2)}-\boldsymbol{\mu}^{(2)})'(\boldsymbol{x}_k^{(2)}-\boldsymbol{\mu}^{(2)}) - 2\boldsymbol{\delta}'(\overline{\boldsymbol{x}}^{(g')}-\boldsymbol{\mu}^{(g')}). \end{aligned}$$

It can be shown that

$$\begin{aligned} \mathrm{Var}\,[T_1] \;=\;& 4\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta} + \frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2} + \frac{4}{N_{g'}}\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}, \\ \mathrm{Var}\,[T_2] \;=\;& \frac{2}{N_1(N_1-1)}\mathrm{tr}\Sigma^{(1)^2} + \frac{2}{N_2(N_2-1)}\mathrm{tr}\Sigma^{(2)^2} + \frac{4}{N_{g'}}\boldsymbol{\delta}'\Sigma^{(g')}\boldsymbol{\delta} \end{aligned}$$

and $\mathrm{Cov}(T_1, T_2) = 0$. From the above results, the proof of Lemma A.2 is complete. □

Lemma A. 3 (The variance of some estimators)**.** *We assume that* (C1)*,* (A0)*,* (A1)*,* (A2') *or* (C2)*,* (A0)-(A3)*. Then it holds that*

$$
\begin{aligned}
\text{(i)}\quad & \mathrm{Var}[\widehat{\boldsymbol{\delta}'\boldsymbol{\delta}}] = \frac{2}{N_1(N_1-1)}\mathrm{tr}\Sigma^{(1)^2} + \frac{2}{N_2(N_2-1)}\mathrm{tr}\Sigma^{(2)^2} \\
& \qquad + \frac{4}{N_1 N_2}\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)} + \frac{4}{N_1}\boldsymbol{\delta}'\Sigma^{(1)}\boldsymbol{\delta} + \frac{4}{N_2}\boldsymbol{\delta}'\Sigma^{(2)}\boldsymbol{\delta}, \\
\text{(ii)}\quad & \mathrm{Var}[\widehat{\mathrm{tr}\Sigma^{(1)}\Sigma^{(2)}}] = O\left(\left(\frac{1}{N_g}+\frac{1}{N_{g'}}\right)\mathrm{tr}(\Sigma^{(g)}\Sigma^{(g')})^2\right. \\
& \qquad \left. + \frac{1}{N_g N_{g'}}(\mathrm{tr}\Sigma^{(g)}\Sigma^{(g')})^2\right), \\
\text{(iii)}\quad & \mathrm{Var}[\widehat{\mathrm{tr}\Sigma^{(g)^2}}] = O\left(\frac{1}{N_g}\mathrm{tr}\Sigma^{(g)^4} + \frac{1}{N_g^2}(\mathrm{tr}\Sigma^{(g)^2})^2\right), \\
\text{(iv)}\quad & \mathrm{Var}[\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}] = O\left(\frac{1}{N_g}\left(\frac{1}{N_g}+\frac{1}{N_{g'}}\right)^2(\mathrm{tr}\Sigma^{(g)^2})^2\right. \\
& \qquad + \frac{1}{N_g}\left(\frac{1}{N_g}+\frac{1}{N_{g'}}\right)\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)^2} \\
& \qquad \left. + \frac{1}{N_g}(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta})^2\right) \\
& \qquad + o\left(\left(\frac{1}{N_g}+\frac{1}{N_{g'}}\right)^2(\mathrm{tr}\Sigma^{(g)^2})^2\right. \\
& \qquad \left. + \left(\frac{1}{N_g}+\frac{1}{N_{g'}}\right)\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)^2}\right).
\end{aligned}
$$

Proof. For the proof of (i), see e.g. Chen and Qin (2010). The proof of (ii) follows the same approach. Note that the estimator $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ is the same as that is proposed by Chen et al. (2010). For the details of (iii), see e.g. Chen et al. (2010). We give the proof of (iv). Let $\boldsymbol{y}_j^{(g)} = \boldsymbol{x}_j^{(g)} - \boldsymbol{\mu}^{(g)}$. From the definition of $\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}}$, this statistic can be expressed as $\widehat{\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}} = \sum_{\alpha=1}^{12} W_\alpha$,

where

$$
\begin{aligned}
W_1 &= \frac{1}{N_g(N_g-1)(N_g-2)} \sum_{\substack{j,k,\ell=1 \\ j\neq k, k\neq \ell, \ell\neq j}}^{N_g} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(g)}, \\
W_2 &= -\frac{1}{N_g(N_g-1)(N_g-2)(N_g-3)} \sum_{\substack{j,k,\ell,m=1 \\ j\neq k\neq \ell\neq m \\ \ell\neq j\neq m\neq k}}^{N_g} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_\ell^{(g)\prime}\boldsymbol{y}_m^{(g)}, \\
W_3 &= -\frac{2}{N_g N_{g'}(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \sum_{\ell=1}^{N_{g'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(g')}, \\
W_4 &= \frac{2}{N_g N_{g'}(N_g-1)(N_g-2)} \sum_{\substack{j,k,\ell=1 \\ j\neq k, k\neq \ell, \ell\neq j}}^{N_g} \sum_{m=1}^{N_{g'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_\ell^{(g)\prime}\boldsymbol{y}_m^{(g')}, \\
W_5 &= \frac{2}{N_g N_{g'}(N_{g'}-1)} \sum_{j=1}^{N_g} \sum_{\substack{k,\ell=1 \\ k\neq \ell}}^{N_{g'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g')}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(g')}, \\
W_6 &= -\frac{2}{N_g(N_g-1)N_{g'}(N_{g'}-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \sum_{\substack{\ell,m=1 \\ \ell\neq m}}^{N_{g'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(g')}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_m^{(g')}, \\
W_7 &= \frac{2}{N_g(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}, \\
W_8 &= -\frac{2}{N_g(N_g-1)(N_g-2)} \sum_{\substack{j,k,\ell=1 \\ j\neq k, k\neq \ell, \ell\neq j}}^{N_g} \left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(g)}, \\
W_9 &= -\frac{2}{N_g N_{g'}} \sum_{j=1}^{N_g} \sum_{k=1}^{N_{g'}} \left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g')}, \\
W_{10} &= \frac{2}{N_g N_{g'}(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \sum_{\ell=1}^{N_{g'}} \left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(g')},
\end{aligned}
$$

$$
\begin{aligned}
W_{11} &= \frac{1}{N_g}\sum_{j=1}^{N_g}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right),\\
W_{12} &= -\frac{1}{N_g(N_g-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_g}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(g')}\right).
\end{aligned}
$$

The variances of W_α(for $\alpha=1,\ldots,12$) are derived as

$$
\begin{aligned}
\mathrm{Var}[W_1] &= O\left(\frac{1}{N_g^3}(\mathrm{tr}\Sigma^{(g)^2})^2+\frac{1}{N_g^2}\mathrm{tr}\Sigma^{(g)^4}\right),\\
\mathrm{Var}[W_2] &= O\left(\frac{1}{N_g^4}(\mathrm{tr}\Sigma^{(g)^2})^2\right),\\
\mathrm{Var}[W_3] &= O\left(\frac{1}{N_g^2N_{g'}}(\mathrm{tr}\Sigma^{(g)^2})^2+\frac{1}{N_gN_{g'}}\mathrm{tr}\Sigma^{(g)^3}\Sigma^{(g')}\right),\\
\mathrm{Var}[W_4] &= O\left(\frac{1}{N_g^3N_{g'}}(\mathrm{tr}\Sigma^{(g)^2})^2\right),\\
\mathrm{Var}[W_5] &= O\left(\frac{1}{N_gN_{g'}^2}(\mathrm{tr}\Sigma^{(g)}\Sigma^{(g')})^2+\frac{1}{N_{g'}^2}\mathrm{tr}(\Sigma^{(g)}\Sigma^{(g')})^2\right),\\
\mathrm{Var}[W_6] &= O\left(\frac{1}{N_g^2N_{g'}^2}(\mathrm{tr}\Sigma^{(g)}\Sigma^{(g')})^2\right),\\
\mathrm{Var}[W_7] &= O\left(\frac{1}{N_g^2}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)^2}+\frac{1}{N_g^3}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[W_8] &= O\left(\frac{1}{N_g^3}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)^2}+\frac{1}{N_g^3}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[W_9] &= O\left(\frac{1}{N_gN_{g'}}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)^2}+\frac{1}{N_{g'}}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\left(\mathrm{tr}(\Sigma^{(g)}\Sigma^{(g')})^2\right)^{1/2}\right),\\
\mathrm{Var}[W_{10}] &= O\left(\frac{1}{N_g^3N_{g'}}\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}\mathrm{tr}\Sigma^{(g)}\Sigma^{(g')}\right),\\
\mathrm{Var}[W_{11}] &= O\left(\frac{1}{N_g}(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta})^2\right),\\
\mathrm{Var}[W_{12}] &= O\left(\frac{1}{N_g^2}(\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta})^2\right).
\end{aligned}
$$

From the above results, the proof is complete. □

Acknowledgments. The authors thank Professor Makoto Aoshima and Professor Yasunori Fujikoshi for their extensive and insightful suggestions, encouragement, and reference material.

References

[1] Aoshima, M. and Yata, K. (2014). A distance-based, misclassification probability adjusted classifier for multiclass, high-dimensional data. *Ann. Inst. Statist. Math.*, **66**, 983-1010.

[2] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Anal.*, **30**, 356-399.

[3] Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, **38**, 808-835.

[4] Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Am. Statist. Assoc.*, **105**, 810-819.

[5] Dudoit, S., Fridlyand, J. and Speed, P.T. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Amer. Statist. Assoc.*, **97**, 77-87.

[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. 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. Stoch. Equ.*, **6**, 269-280.

[8] Himeno, T. and Yamada, T. (2014). Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Anal.*, **130**, 27-44.

[9] Konishi, S. and Honda, M. (1990). Comparison of procedures for estimation of error probabilities in discriminant analysis under nonnormal populations. *J. Statist. Comput. Simul.*, **36**, 105-115.

[10] Kubokawa, T., Hyodo, M. and Muni S. Srivastava. (2013). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension. *J. Multivariate Anal.*, **115**, 496–515.

[11] 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.

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

[13] Okamoto, M. (1968). Correction to "An asymptotic expansion for the distribution of the linear discriminant function". *Ann. Math. Stat.*, **39**, 1358-1359.

[14] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistic. *Handbook of Statistics* 2 (P. R. Krishnaiah and L. N. Kanal, Eds.), North-Holland Publishing Company, 61-100.

[15] Srivastava, M.S., Yanagihara, H. and Kubokawa, T. (2014). Tests for covariance matrices in high dimension with less sample size. *J. Multivariate Anal.*, **130**, 289–309.

Table 1: Comparison of approximations where $\boldsymbol{\mu}^{(2)} = (\sqrt{10/p}, \ldots, \sqrt{10/p})'$

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	$e(2\|1)$	Case I	0.2700	0.2626	0.2611
		Case II	0.2586	0.2480	0.2436
	approx	YHS	0.2744	0.2641	0.2604
		AY	0.0883	0.0273	0.0092
250	$e(2\|1)$	Case I	0.3071	0.2886	0.2817
		Case II	0.2908	0.2731	0.2673
	approx	YHS	0.3076	0.2903	0.2835
		AY	0.1984	0.1141	0.0696
500	$e(2\|1)$	Case I	0.3354	0.3118	0.2996
		Case II	0.3229	0.2973	0.2838
	approx	YHS	0.3349	0.3113	0.3009
		AY	0.2751	0.1977	0.1485
1000	$e(2\|1)$	Case I	0.3653	0.3356	0.3229
		Case II	0.3574	0.3219	0.3090
	approx	YHS	0.3646	0.3363	0.3220
		AY	0.3365	0.2742	0.2308

Table 2: Comparison of approximations where $\boldsymbol{\mu}^{(2)} = (1, \ldots, 1, 0, \ldots, 0)'$

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	$e(2\|1)$	Case I	0.1360	0.1125	0.1020
		Case II	0.1257	0.1018	0.0918
	approx	YHS	0.1369	0.1120	0.1026
		AY	0.0395	0.0062	0.0011
250	$e(2\|1)$	Case I	0.1152	0.0840	0.0740
		Case II	0.1042	0.0783	0.0698
	approx	YHS	0.1125	0.0837	0.0730
		AY	0.0376	0.0057	0.0010
500	$e(2\|1)$	Case I	0.1008	0.0677	0.0551
		Case II	0.0891	0.0618	0.0516
	approx	YHS	0.0969	0.0658	0.0544
		AY	0.0365	0.0054	0.0009
1000	$e(2\|1)$	Case I	0.0877	0.0537	0.0407
		Case II	0.0823	0.0494	0.0406
	approx	YHS	0.0861	0.0525	0.0406
		AY	0.0382	0.0059	0.0010

Table 3: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (\sqrt{10/p}, \ldots, \sqrt{10/p})'$ in Case I

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	Bias	$\Phi(\widehat{w}_1)$	0.0062	0.0018	-0.0008
		$CV(2\|1)$	0.0020	-0.0006	-0.0025
	MSE	$\Phi(\widehat{w}_1)$	0.0046	0.0019	0.0012
		$CV(2\|1)$	0.0073	0.0033	0.0022
250	Bias	$\Phi(\widehat{w}_1)$	0.0042	0.0030	0.0023
		$CV(2\|1)$	0.0007	0.0002	0.0004
	MSE	$\Phi(\widehat{w}_1)$	0.0055	0.0022	0.0013
		$CV(2\|1)$	0.0084	0.0037	0.0024
500	Bias	$\Phi(\widehat{w}_1)$	0.0029	0.0013	0.0021
		$CV(2\|1)$	0.0010	-0.0006	0.0003
	MSE	$\Phi(\widehat{w}_1)$	0.0063	0.0025	0.0015
		$CV(2\|1)$	0.0095	0.0042	0.0026
1000	Bias	$\Phi(\widehat{w}_1)$	0.0026	0.0027	0.0006
		$CV(2\|1)$	0.0019	0.0013	-0.0007
	MSE	$\Phi(\widehat{w}_1)$	0.0069	0.0030	0.0018
		$CV(2\|1)$	0.0102	0.0047	0.0030

Table 4: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (1, \ldots, 1, 0, \ldots, 0)'$ in Case I

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	Bias	$\Phi(\widehat{w}_1)$	-0.0005	-0.0010	0.0003
		$CV(2\|1)$	0.0027	0.0000	0.0010
	MSE	$\Phi(\widehat{w}_1)$	0.0028	0.0011	0.0007
		$CV(2\|1)$	0.0048	0.0020	0.0012
250	Bias	$\Phi(\widehat{w}_1)$	-0.0030	-0.0001	-0.0007
		$CV(2\|1)$	0.0004	0.0007	-0.0004
	MSE	$\Phi(\widehat{w}_1)$	0.0024	0.0009	0.0005
		$CV(2\|1)$	0.0043	0.0016	0.0010
500	Bias	$\Phi(\widehat{w}_1)$	-0.0032	-0.0011	0.0000
		$CV(2\|1)$	0.0002	-0.0007	0.0001
	MSE	$\Phi(\widehat{w}_1)$	0.0022	0.0007	0.0004
		$CV(2\|1)$	0.0039	0.0014	0.0008
1000	Bias	$\Phi(\widehat{w}_1)$	-0.0002	-0.0002	0.0007
		$CV(2\|1)$	0.0028	0.0002	0.0004
	MSE	$\Phi(\widehat{w}_1)$	0.0020	0.0005	0.0003
		$CV(2\|1)$	0.0037	0.0012	0.0006

Table 5: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (\sqrt{10/p}, \ldots, \sqrt{10/p})'$ in Case II

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	Bias	$\Phi(\widehat{w}_1)$	0.0143	0.0149	0.0158
		$CV(2\|1)$	-0.0017	-0.0015	-0.0007
	MSE	$\Phi(\widehat{w}_1)$	0.0052	0.0024	0.0016
		$CV(2\|1)$	0.0071	0.0033	0.0021
250	Bias	$\Phi(\widehat{w}_1)$	0.0171	0.0167	0.0156
		$CV(2\|1)$	0.0031	0.0008	-0.0004
	MSE	$\Phi(\widehat{w}_1)$	0.0062	0.0027	0.0017
		$CV(2\|1)$	0.0084	0.0037	0.0024
500	Bias	$\Phi(\widehat{w}_1)$	0.0131	0.0142	0.0169
		$CV(2\|1)$	0.0011	-0.0004	0.0018
	MSE	$\Phi(\widehat{w}_1)$	0.0068	0.0030	0.0019
		$CV(2\|1)$	0.0093	0.0042	0.0026
1000	Bias	$\Phi(\widehat{w}_1)$	0.0075	0.0149	0.0134
		$CV(2\|1)$	-0.0016	0.0023	-0.0001
	MSE	$\Phi(\widehat{w}_1)$	0.0071	0.0034	0.0021
		$CV(2\|1)$	0.0101	0.0047	0.0029

Table 6: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (1, \ldots, 1, 0, \ldots, 0)'$ in Case II

			(N_1, N_2)		
p			(20,40)	(40,80)	(60,120)
100	Bias	$\Phi(\widehat{w}_1)$	0.0074	0.0086	0.0099
		$CV(2\|1)$	0.0008	0.0009	0.0021
	MSE	$\Phi(\widehat{w}_1)$	0.0034	0.0015	0.0010
		$CV(2\|1)$	0.0046	0.0019	0.0012
250	Bias	$\Phi(\widehat{w}_1)$	0.0062	0.0049	0.0031
		$CV(2\|1)$	0.0019	-0.0004	-0.0016
	MSE	$\Phi(\widehat{w}_1)$	0.0030	0.0011	0.0007
		$CV(2\|1)$	0.0041	0.0016	0.0009
500	Bias	$\Phi(\widehat{w}_1)$	0.0071	0.0045	0.0032
		$CV(2\|1)$	0.0035	0.0011	0.0006
	MSE	$\Phi(\widehat{w}_1)$	0.0028	0.0009	0.0005
		$CV(2\|1)$	0.0038	0.0013	0.0007
1000	Bias	$\Phi(\widehat{w}_1)$	0.0035	0.0040	0.0007
		$CV(2\|1)$	0.0006	0.0018	-0.0002
	MSE	$\Phi(\widehat{w}_1)$	0.0025	0.0007	0.0004
		$CV(2\|1)$	0.0035	0.0011	0.0006