Estimation of misclassification probability for multi-class classification based on the Euclidean distance in high-dimensional data

Masashi Hyodo[*], Yuki Yamada[†]and Hiroshi Furukawa[‡]

Tokyo University of Science

April 8, 2015

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 problem in discriminant analysis is estimating the probability of misclassification. Recently, Yamada et al. (2015) proposed the asymptotically unbiased and consistent estimator of misclassification probabilities for two class classification under high-dimensional settings where the number of parameters exceeds the sample size. Their proposed method has the advantage of establishing under variance heterogeneity and nonnormality. In this paper, we extend their discussion to the case of multiple groups. One of the main contributions of this paper is to establish the asymptotic multivariate normality for several discriminant functions. By using this asymptotic multivariate normality and new estimator of bilinear form, we also obtain the asymptotically unbiased and consistent estimator of misclassification probabilities for multi-class classification. Finally, we numerically justify the high accuracy of our proposed estimator in finite sample applications, inclusive of high-dimensional scenarios.

Key words and phrases: multi-class classification, asymptotic normality, Euclidean distance-based classifier, high-dimensional data, probability of misclassification.

[*]Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, 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, E-Mail:y.y.rfpfg.t.o.d.2006.s.m.e@gmail.com

[‡]Graduate School of Engineering, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, JAPAN

1 Introduction

In this paper, we focus on the multi-class classification which 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 $q+1$ populations, $\Pi_1, \ldots, \Pi_q$ and Π_{q+1}. Suppose that $\boldsymbol{x} \in \Pi_g$ $(g = 1, \ldots, q+1)$; then, we assume that

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

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

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

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 to explore the discrimination procedure that can accommodate $p > \max\{N_1, N_2, \ldots, N_{q+1}\}$ cases, with the main focus on the performance accuracy in the asymptotic framework that allows p to grow together with $N_1, N_2, \ldots, N_{q+1}$. Recently, Aoshima and Yata (2014) have been considered the Euclidean distance-based classifier for the high-dimensional multi-class problem with different class covariance matrices. Let $\boldsymbol{x}$ be an observation vector into one of the several populations $\Pi_1, \Pi_2, \ldots, \Pi_{q+1}$. The Euclidean distance-based discriminant function is defined as

$$W_{hg}(\boldsymbol{x}) = \|\boldsymbol{x} - \overline{\boldsymbol{x}}^{(h)}\|^2 - \|\boldsymbol{x} - \overline{\boldsymbol{x}}^{(g)}\|^2 - \mathrm{tr}\left[\frac{1}{N_h}S^{(h)} - \frac{1}{N_g}S^{(g)}\right], \qquad (1.1)$$

where $S^{(i)}$ $(i = 1, 2, \ldots, q+1)$ denote the sample covariance matrices. Then, to classify an individual $\boldsymbol{x}$ to Π_k, the regions of classification, $\mathcal{R}_g$ $(g = 1, \ldots, q+1)$, are given by

$$\mathcal{R}_g = \{\boldsymbol{x} | W_{hg}(\boldsymbol{x}) > c_g - c_h, h = 1, 2, \ldots, q+1, h \neq g\},$$

where c_g and c_h are some constants. The classification rule is given by $\boldsymbol{x} \in \mathcal{R}_g \Rightarrow \boldsymbol{x} \in \Pi_g$. The misclassification probability in this case is given by

$$e_g = 1 - \Pr(\boldsymbol{x} \in \mathcal{R}_g | \boldsymbol{x} \in \Pi_g) \ (g = 1, \ldots, q+1).$$

In this study, we focus on the misclassification probability e_g. Our main objective is to derive limiting value of misclassification probability and propose its consistent and asymptotically unbiased estimator under high-dimensional frameworks. It is generally difficult to obtain an exact value of misclassification probability. So, there are much works for asymptotic approximations for misclassification probability of the Fisher linear discriminant rule. The approximations are the ones under a framework such that N_1 and N_2 are large and p is fixed has been studied. For a review of these results, see, e.g., Okamoto (1963, 1968), Siotani (1982). Further, the asymptotic approximation under a framework that N_1, N_2 and p are all large have also been studied (see, e.g., Lachenbruch (1968) and Fujikoshi and Seo (1998)). Moreover, Fujikoshi (2000) obtained an explicit formula of error bounds for an approximation of misclassification probability. Recently, Aoshima and Yata (2014) showed the asymptotic normality of the Euclidean distance discriminant function under the under high-dimensional asymptotic framework. Yamada et al. (2015) obtained the asymptotically unbiased and consistent estimator of misclassification probability for two-class classification based on the Euclidean distance. In this paper, we extend their discussion to the case of multiple groups. One of the main contributions of this paper is to establish the asymptotic multivariate normality for several discriminant functions. By using asymptotic multivariate normality, we propose the consistent and asymptotically unbiased estimator of misclassification probability of e_g. As a by product, we also derive the unbiased estimator of $\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}$, where $\boldsymbol{\delta}_{gh} = \boldsymbol{\mu}_g - \boldsymbol{\mu}_h$.

The remaining part of the paper is organized as follows. In Section 2, we show the asymptotic multivariate normality for several Euclidean discriminant functions. In Section 3, we derived the asymptotically unbiased and consistent estimator of misclassification probability of e_g. In Section 4 summaries the results of numerical experiments justifying the validity of the suggested estimators for the data along with a number of high-dimensional scenarios. We conclude in Section 5, and give some auxiliary lemmas in Appendix.

2 Asymptotic normality of Euclidean distance-based classifier

In this section, we show the asymptotic multivariate normality for q-dimensional vector $\boldsymbol{w}_j$, the component is several Euclidean discriminant functions, i.e.

$$\boldsymbol{w}_g = (W_{1g}(\boldsymbol{x}), \ldots, W_{g-1g}(\boldsymbol{x}), W_{g+1g}(\boldsymbol{x}), \ldots, W_{q+1g}(\boldsymbol{x}))'.$$

We assume (C1) and the high-dimensional asymptotic framework

$$\text{(A0)} \qquad q : \text{fix},\ p, N_1, \ldots, N_{q+1} \to \infty,\ N_1 \asymp N_h\ (h = 2, \ldots, q+1),$$

$$\text{(A1)} \qquad \lim_{p\to\infty} \frac{\mathrm{tr}\Sigma^{(g)^4}}{(\mathrm{tr}\Sigma^{(g)^2})^2} = 0, \quad 0 < \lim_{p\to\infty} \frac{\mathrm{tr}\Sigma^{(h)}\Sigma^{(g)}}{\mathrm{tr}\Sigma^{(g)^2}} < \infty \quad (g, h = 1, 2, \ldots, q+1),$$

$$\text{(A2')} \qquad \lim_{p,N_g,N_h\to\infty} \frac{\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}}{\delta^2_{hg}} = 0 \ \ (g, h = 1, 2, \ldots, q+1)$$

or (C2) and the high-dimensional asymptotic framework (A0), (A1),

$$\text{(A2)} \qquad \lim_{p,N_g,N_h\to\infty} \frac{\boldsymbol{\delta}'_{gh}\Sigma^{(h)}\boldsymbol{\delta}_{gh}}{N_h\sigma^2_{hg}} = 0 \quad (g, h = 1, 2, \ldots, q+1,\ g \neq h),$$

$$\text{(A3)} \qquad \lim_{p,N_g,N_h\to\infty} \frac{\max\{\gamma^{(g)^2}_{(g,h),1}, \ldots, \gamma^{(g)^2}_{(g,h),p}\}}{\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}} = 0,\ \boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} \asymp \frac{\mathrm{tr}\Sigma^{(g)^2}}{N_g}$$
$$(g, h = 1, 2, \ldots, q+1,\ g \neq h),$$

where

$$\begin{aligned}
\delta^2_{hg} &= \frac{4\mathrm{tr}\Sigma^{(g)^2}}{N_g} + \frac{4\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}}{N_h} + \frac{2\mathrm{tr}\Sigma^{(g)^2}}{N_g(N_g-1)} + \frac{2\mathrm{tr}\Sigma^{(h)^2}}{N_h(N_h-1)}, \\
\sigma^2_{hg} &= 4\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} + \frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2} + \frac{4}{N_h}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)},\ \ (g, h = 1, \ldots, q+1,\ g \neq h), \\
\sigma_{hg,h'g} &= 4\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'} + \frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2}\ \ (g, h, h' = 1, \ldots, q+1,\ g \neq h \neq h' \neq g).
\end{aligned}$$

Here, $\gamma^{(g)}_{(g,h),i}$ denote the i-th element of $\Gamma^{(g)'}\boldsymbol{\delta}_{gh}$. Suppose that

$$\boldsymbol{v}_g = \mathrm{diag}(\sigma^{-1}_{1g}, \ldots, \sigma^{-1}_{g-1g}, \sigma^{-1}_{g+1g}, \ldots, \sigma^{-1}_{q+1g})(\boldsymbol{w}_g - \mathrm{E}[\boldsymbol{w}_g|\boldsymbol{x} \in \Pi_g]).$$

The following theorem provides the asymptotic normality of $\boldsymbol{v}_g$.

Theorem 2.1 *Let $\boldsymbol{x} \in \Pi_g\,(g = 1, \ldots, q+1)$. Under (C1), (A0), (A1) and (A2') or (C2), (A0)-(A3), it holds that*

$$\boldsymbol{v}_g \xrightarrow{d} \mathcal{N}(\mathbf{0}, R),$$

where

$$R = \lim_{p,N_1,\ldots,N_{q+1}\to\infty} \begin{pmatrix} 1 & \rho_{1g,2g} & \rho_{1g,3g} & \cdots & \rho_{1g,qg} & \rho_{1g,q+1g} \\ \rho_{2g,1g} & 1 & \rho_{2g,3g} & \cdots & \rho_{2g,qg} & \rho_{2g,q+1g} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{q+1g,1g} & \rho_{q+1g,2g} & \rho_{q+1g,3g} & \cdots & \rho_{q+1g,qg} & 1 \end{pmatrix}.$$

Here, $\rho_{hg,h'g} = \sigma_{hg,h'g}/(\sigma_{hg}\sigma_{h'g})$ $(g, h, h' = 1, 2, \ldots, q+1,\ g \neq h \neq h' \neq g)$.

Proof. Let β_i $(i = 1, \ldots, q+1, i \neq g)$ denotes q non-random values which satisfy

$$0 < \lim_{p, N_1, \ldots, N_{q+1} \to \infty} |\beta_i| < \infty,$$

and introduce the random variable

$$T = \sum_{\substack{h=1 \\ h \neq g}}^{q+1} \beta_h T_{hg}.$$

which is defined as the linear combination of

$$T_{hg} = \frac{W_{hg}(\boldsymbol{x}) - \boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}{\sigma_{hg}} \ (h = 1, \ldots, q+1, h \neq g).$$

Then T_{hg} is decomposed as $T_{hg} = T_{hg}^{(1)} + T_{hg}^{(2)}$, where

$$\begin{aligned}
\sigma_{hg} T_{hg}^{(1)} &= 2(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\boldsymbol{\delta}_{gh} + 2(\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\{(\overline{\boldsymbol{x}}^{(g)} - \boldsymbol{\mu}^{(g)}) - (\overline{\boldsymbol{x}}^{(h)} - \boldsymbol{\mu}^{(h)})\}, \\
\sigma_{hg} T_{jk}^{(2)} &= -\frac{1}{N_g(N_g - 1)} \sum_{\substack{j,k=1 \\ j \neq k}}^{N_g} (\boldsymbol{x}_j^{(g)} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_k^{(g)} - \boldsymbol{\mu}^{(g)}) \\
&\quad + \frac{1}{N_h(N_h - 1)} \sum_{\substack{j,k=1 \\ j \neq k}}^{N_h} (\boldsymbol{x}_j^{(h)} - \boldsymbol{\mu}^{(h)})'(\boldsymbol{x}_k^{(h)} - \boldsymbol{\mu}^{(h)}) - 2\boldsymbol{\delta}'_{gh}(\overline{\boldsymbol{x}}^{(h)} - \boldsymbol{\mu}^{(h)}).
\end{aligned}$$

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

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

where

$$\begin{aligned}
\psi_i &= \left(\sum_{\substack{h=1 \\ h \neq g}}^{q+1} \frac{\beta_h}{\sigma_{hg}} \right) \frac{2}{N_g} (\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_i^{(g)} - \boldsymbol{\mu}^{(g)}) \ (i = 1, \ldots, N_g), \\
\psi_i &= -\frac{2\beta_h}{\sigma_{hg} N_h} (\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_{i-N_g}^{(h)} - \boldsymbol{\mu}^{(h)}) \\
&\quad \left(i = N_g + \sum_{\substack{h'=0 \\ h' \neq g}}^{h-1} N_{h'} + 1, \ldots, N_g + \sum_{\substack{h'=0 \\ h' \neq g}}^{h} N_{h'} \right).
\end{aligned}$$

Here, $N = \sum_{j=1}^{q+1} N_j$ and $N_0 = 0$.

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. To apply martingale central limit theorem, we need to show that

$$\sum_{i=1}^{N} \mathrm{E}[\psi_i^2|\mathcal{F}_{i-1}] \xrightarrow{p} \sigma_g^2 \qquad \text{and} \qquad \sum_{i=1}^{N} \mathrm{E}[\psi_i^4] \to 0, \tag{2.1}$$

where

$$\sigma_g^2 = \sum_{\substack{h=1 \\ h \neq g}}^{q+1} \beta_h^2 + \sum_{\substack{h,h'=1 \\ g \neq h \neq h' \neq g}}^{q+1} \beta_h \beta_{h'} \rho_{hg,h'g}.$$

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

$$\begin{aligned}\sum_{i=1}^{N} \mathrm{E}[\psi_i^2|\mathcal{F}_{i-1}] - \sigma_g^2 &= \left(\sum_{h=1, h\neq g}^{q+1} \frac{\beta_h}{\sigma_{hg}}\right)^2 \frac{4}{N_g} \left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\Sigma^{(g)}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)}) - \mathrm{tr}\Sigma^{(g)^2}\right) \\ &\quad + \sum_{h=1, h\neq g}^{q+1} \frac{4\beta_h^2}{\sigma_{hg}^2 N_h} \left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\Sigma^{(h)}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)}) - \mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}\right) + o(1).\end{aligned}$$

Hence, it holds that

$$\sum_{i=1}^{N} \mathrm{E}[\psi_i^2|\mathcal{F}_{i-1}] - \sigma_g^2 = o_p(1)$$

since

$$\begin{aligned}\mathrm{E}\left[\left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\Sigma^{(g)}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)}) - \mathrm{tr}\Sigma^{(g)^2}\right)^2\right] &\leq (2 + \kappa_{4\max}^{(g)})\mathrm{tr}\Sigma^{(g)^4}, \\ \mathrm{E}\left[\left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'\Sigma^{(h)}(\boldsymbol{x} - \boldsymbol{\mu}^{(g)}) - \mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}\right)^2\right] &\leq (2 + \kappa_{4\max}^{(g)})\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2.\end{aligned}$$

This proves the first part of (2.1).

Next, we show the second part of (2.1). Under (C1), (A0), (A1) and (A2'), it holds that

$$\sum_{i=1}^{N} \mathrm{E}[\psi_i^4] = o(1),$$

since

$$\begin{aligned}\mathrm{E}\left[\frac{\left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_i^{(g)} - \boldsymbol{\mu}^{(g)})\right)^4}{N_g^4}\right] &= O\left(\frac{(\mathrm{tr}\Sigma^{(g)^2})^2 + \mathrm{tr}\Sigma^{(g)^4}}{N_g^4}\right), \\ \mathrm{E}\left[\frac{\left((\boldsymbol{x} - \boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_{i-N_g}^{(h)} - \boldsymbol{\mu}^{(h)})\right)^4}{N_h^4}\right] &= O\left(\frac{(\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)})^2 + \mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2}{N_h^4}\right).\end{aligned}$$

This proves the second part of (2.1) and finishes the proof of asymptotic normality of T under (C1),(A0),(A1) and (A2').

Next, we show the asymptotic normality of T under (C1),(A0)-(A3). Let $\overline{\boldsymbol{y}}^{(g)} = \overline{\boldsymbol{x}}^{(g)} - \boldsymbol{\mu}^{(g)}$, $\overline{\boldsymbol{y}}^{(g)} = \Gamma^{(g)}\overline{\boldsymbol{z}}^{(g)}$. Then T_{hg} can be factorized as $\sum_{i=1}^{p}\epsilon_i^{(h)} + o_p(1)$, where

$$\sigma_{hg}\epsilon_i^{(h)} = 2\gamma_{(g,h),i}^{(g)} z_i + 2\lambda_i^{(g)} z_i \overline{z}_i^{(g)} - 2\lambda_i^{(g)^{1/2}} z_i \boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h)}.$$

Here, $H^{(g)}$ is an orthogonal matrix such that $H^{(g)}\Lambda^{(g)}H^{(g)\prime} = \Sigma^{(g)}$ where, $\Lambda^{(g)} = \mathrm{diag}(\lambda_1^{(g)}, \ldots, \lambda_p^{(g)})$ and $\lambda_i^{(g)}$ is i-th eigenvalues of $\Sigma^{(g)}$. Note that

$$T = \sum_{i=1}^{p}\sum_{\substack{h=1\\h\neq g}}^{q+1}\beta_h\epsilon_i^{(h)} + o_p(1).$$

Suppose that

$$\epsilon_i = \sum_{h=1,h\neq g}^{q+1}\beta_h\epsilon_i^{(h)}$$

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

$$\sum_{i=1}^{p}\theta_{gi}^2 \xrightarrow{p} \sigma_g^2 \qquad \text{and} \qquad \sum_{i=1}^{p}\mathrm{E}[\epsilon_i^4] \to 0, \tag{2.2}$$

where

$$\sigma_g^2 = \sum_{\substack{h=1\\h\neq g}}^{q+1}\beta_h^2 + \sum_{\substack{h,h'=1\\g\neq h\neq h'\neq g}}^{q+1}\beta_h\beta_{h'}\rho_{hg,h'g},\ \ \theta_{gi}^2 = \mathrm{E}[\epsilon_i^2|\mathcal{F}_{i-1}].$$

We first show the first part of (2.2). Note that

$$\theta_{gi}^2 \ = \ \sum_{\substack{h=1\\h\neq g}}^{q+1}\beta_h^2\theta_i^{(h)} + \sum_{\substack{h,h'=1\\g\neq h\neq h'\neq g}}^{q+1}\beta_h\beta_{h'}\theta_i^{(h,h')},$$

where

$$\begin{aligned}
\theta_i^{(h)} \ &= \ \frac{4}{\sigma_{hg}^2}\left(\gamma_{(g,h),i}^{(g)^2} + \frac{\lambda_i^{(g)^2}}{N_g} + \lambda_i^{(g)}\overline{\boldsymbol{y}}^{(h)\prime}\boldsymbol{h}_i^{(g)}\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h)} - 2\gamma_{(g,h),i}^{(g)}\lambda_i^{(g)^{1/2}}\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h)}\right),\\
\theta_i^{(h,h')} \ &= \ \frac{4}{\sigma_{hg}\sigma_{h'g}}\Bigg(\gamma_{(g,h),i}^{(g)}\gamma_{(g,h'),i}^{(g)} + \frac{\lambda_i^{(g)^2}}{N_g} + \lambda_i^{(g)}\overline{\boldsymbol{y}}^{(h)\prime}\boldsymbol{h}_i^{(g)}\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h')}\\
&\quad -\lambda_i^{(g)^{1/2}}\gamma_{(g,h),i}^{(g)}\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h')} - \lambda_i^{(g)^{1/2}}\gamma_{(g,h'),i}^{(g)}\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h)}\Bigg).
\end{aligned}$$

Then we get

$$\mathrm{E}\left[\sum_{i=1}^{p}\theta_{gi}^2\right] = \sigma_g^2.$$

It is sufficient to show $\mathrm{Var}[A_1] = o(1)$, $\mathrm{Var}[A_2] = o(1)$ and $\mathrm{Var}[A_3] = o(1)$, where

$$A_1 = \frac{\overline{\boldsymbol{y}}^{(h)\prime}\Sigma^{(g)}\overline{\boldsymbol{y}}^{(h)}}{\sigma_{hg}^2}, \ A_2 = \frac{\boldsymbol{\delta}_{gh}'\Sigma^{(g)}\overline{\boldsymbol{y}}^{(h)}}{\sigma_{hg}^2}, \ A_3 = \frac{\overline{\boldsymbol{y}}^{(h)\prime}\Sigma^{(g)}\overline{\boldsymbol{y}}^{(h')}}{\sigma_{hg}\sigma_{h'g}}.$$

The variance of A_1 is evaluated as

$$\begin{aligned}\mathrm{Var}[A_1] &= \mathrm{E}[A_1^2] - \frac{1}{\sigma_{hg}^4 N_h^2}\{\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})\}^2 \\ &\leq \frac{1}{\sigma_{hg}^4 N_h^2}\left(\frac{\kappa_{4\max}^{(h)}}{N_h} + 2\right)\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2 = o(1).\end{aligned}$$

By applying Cauchy-Schwarz inequality,

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

So, the variance of A_2 is evaluated as

$$\mathrm{Var}[A_2] = \mathrm{E}[A_2^2] = \frac{\boldsymbol{\delta}_{gh}'\Sigma^{(h)}\Sigma^{(g)}\Sigma^{(h)}\boldsymbol{\delta}_{gh}}{\sigma_{hg}^4 N_h} \leq \frac{\boldsymbol{\delta}_{gh}'\Sigma^{(h)}\boldsymbol{\delta}_{gh}\sqrt{\mathrm{tr}(\Sigma^{(h)}\Sigma^{(g)})^2}}{\sigma_{hg}^4 N_h} = o(1).$$

The variance of A_3 is evaluated as

$$\mathrm{Var}[A_3] = \mathrm{E}[A_3^2] = \frac{\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}\Sigma^{(g)}\Sigma^{(h')}}{\sigma_{hg}^2\sigma_{h'g}^2 N_h N_{h'}} \leq \frac{\sqrt{\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h')})^2}}{\sigma_{hg}^2\sigma_{h'g}^2 N_h N_{h'}} = o(1).$$

Hence completes the proof for the first part of (2.2).

Next, we show the second part of (2.2). We decompose ϵ_i into $\sum_{h=1,h\neq g}^{q+1}\beta_h\epsilon_i^{(h)}$, where $\epsilon_i^{(h)} = \epsilon_{i1}^{(h)} + \epsilon_{i2}^{(h)} + \epsilon_{i3}^{(h)}$. Here,

$$\epsilon_{i1}^{(h)} = \frac{2\gamma_{(g,h),i}^{(g)}z_i}{\sigma_{hg}}, \ \epsilon_{i2}^{(h)} = \frac{2\lambda_i^{(g)}z_i\bar{z}_i^{(g)}}{\sigma_{hg}}, \ \epsilon_{i3}^{(h)} = -\frac{2\lambda_i^{(g)\frac{1}{2}}z_i\boldsymbol{h}_i^{(g)\prime}\overline{\boldsymbol{y}}^{(h)}}{\sigma_{hg}}.$$

By applying Hölder's inequality, we obtain

$$\sum_{i=1}^{p}\mathrm{E}[\epsilon_i^4] = \sum_{i=1}^{p}\mathrm{E}\left[\left(\sum_{h=1,h\neq g}^{q+1}\beta_h\epsilon_i^{(h)}\right)^4\right] \leq 27q^3\sum_{h=1,h\neq g}^{q+1}\beta_h^4\sum_{i=1}^{p}\mathrm{E}[\epsilon_{i1}^{(h)^4} + \epsilon_{i2}^{(h)^4} + \epsilon_{i3}^{(h)^4}].$$

So, we need to show $\sum_{i=1}^p \mathrm{E}[\epsilon_{i\ell}^{(h)^4}] = o(1)$ for $\ell = 1, 2, 3$. Note that

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

This proves the second part of (2.2) and finishes the proof of asymptotic normality of T.□

By using Theorem 2.1, we obtain the limiting values of probabilities of misclassification in following corollary.

Corollary 2.1 *We assume that*

$$
\text{(A4) } 0 < \lim_{N_g, N_h, p \to \infty} \frac{c_g - c_h - \boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}{\sigma_{hg}} < \infty \ (h = 1, \cdots, q+1, \ h \neq g).
$$

Let

$$
\boldsymbol{c}_g = \lim_{p, N_1, \ldots, N_{q+1} \to \infty}\left(\frac{c_g - c_1 - \boldsymbol{\delta}'_{g1}\boldsymbol{\delta}_{g1}}{\sigma_{1g}}, \ldots, \frac{c_g - c_{q+1} - \boldsymbol{\delta}'_{gq+1}\boldsymbol{\delta}_{gq+1}}{\sigma_{q+1g}}\right)'.
$$

Under (C1), (A0), (A1), (A2') and (A4) or (C2), (A0)-(A4), it holds that

$$
e_g = 1 - F(\boldsymbol{c}_g|R) + o(1), \tag{2.3}
$$

where

$$
F(\boldsymbol{x}|B) = \int_{x_1}^\infty \int_{x_2}^\infty \cdots \int_{x_q}^\infty \frac{1}{(2\pi)^{q/2}|B|^{1/2}} e^{-\frac{1}{2}\boldsymbol{\eta}' B^{-1}\boldsymbol{\eta}} d\boldsymbol{\eta}.
$$

3 Estimation of probability of misclassification

The limiting values (2.3) include the unknown values $\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}$ and σ_{hg}. We use unbiased estimator of $\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}$:

$$
\widehat{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}} = (\overline{\boldsymbol{x}}^{(g)} - \overline{\boldsymbol{x}}^{(h)})'(\overline{\boldsymbol{x}}^{(g)} - \overline{\boldsymbol{x}}^{(h)}) - \frac{\mathrm{tr}S^{(g)}}{N_g} - \frac{\mathrm{tr}S^{(h)}}{N_h}.
$$

The unbiased estimator $\widehat{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}$ is used in two sample test (e.g. Chen and Qin (2010), Aoshima and Yata (2011).) Now consider the estimator of σ_{jk}. We prepare the unbiased estimators of $\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}$, $\mathrm{tr}\Sigma^{(g)^2}$ and $\boldsymbol{\delta}'\Sigma^{(g)}\boldsymbol{\delta}$ as follows:

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

where

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

The unbiased estimator $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ is proposed in Himeno and Yamada (2013), and they showed the consistency of this estimator. Also note that $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ is the same as the one of Chen et al. (2010). Yamada et al. (2015) derived the unbiased estimator $\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh}$, and investigate the leading term of variance of these estimators. In this paper, we obtain the unbiased estimator of $\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}$ and investigate the leading term of variance of this estimator (see, Appendix). By using these estimators $\mathrm{tr}\widehat{\Sigma^{(g)}\Sigma^{(h)}}$, $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$ and $\boldsymbol{\delta}'_{gh}\Sigma^{(j)}\boldsymbol{\delta}_{gh'}$, Yamada et al. (2015) proposed the truncated estimator of σ_{hg}:

$$\widehat{\sigma^2_{hg}} = 4\max\left\{\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh}+\frac{\widehat{\mathrm{tr}\Sigma^{(g)^2}}}{N_g},0\right\}+\frac{4}{N_h}\mathrm{tr}\widehat{\Sigma^{(g)}\Sigma^{(h)}}.$$

This estimator satisfies $\widehat{\sigma^2_{hg}} > 0$ $a.e.$ and $\widehat{\sigma^2_{hg}}/\sigma^2_{hg} \xrightarrow{p} 1$ under assumptions (C1), (A0) and (A1) or (C2), (A0) and (A1). By using $\widehat{\sigma^2_{hg}}$, $\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh'}$ and $\widehat{\mathrm{tr}\Sigma^{(g)^2}}$, we obtain the estimator of $\rho_{hg,h'g}$ as $\widehat{\rho_{hg,h'g}} = \widehat{\sigma_{hg,h'g}}/(\widehat{\sigma_{hg}}\widehat{\sigma_{h'g}})$, where

$$\widehat{\sigma_{hg,h'g}} = 4\boldsymbol{\delta}'_{hg}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{h'g}+\frac{4}{N_g}\widehat{\mathrm{tr}\Sigma^{(g)^2}}.$$

Replacing the unknown values in R, we obtain $q \times q$ matrix $\widehat{R} = (\rho_{hg,h'g})$. We use the following ridge type estimator

$$\widehat{R}_* = \widehat{R} + I_{\{\widehat{R} | \mathrm{Ch}_{min}(\widehat{R}) \le 0\}}(\widehat{R}) q I_q,$$

so that the estimator of R may be negative semidefinite matrix. Here, $\mathrm{Ch}_{min}(A)$ denotes the smallest eigenvalue of matrix A and the function $I_A(\boldsymbol{x})$ is the indicator function defined as

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

Note that $\widehat{R}_* > O$, and under assumptions (C1), (A0) and (A1) or (C2), (A0) and (A1), it holds that

$$\widehat{R}_* \xrightarrow{p} R. \tag{3.1}$$

Further, under assumptions (C1), (A0) and (A1) or (C2), (A0) and (A1), it holds that

$$\frac{\widehat{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}}{\widehat{\sigma_{hg}}} \xrightarrow{p} \frac{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}{\sigma_{hg}}. \tag{3.2}$$

Based on consistency (3.1) and (3.2), we propose the estimator of e_g as

$$\widehat{e_g} = 1 - F\left(\widehat{\boldsymbol{c}}_g | \widehat{R}_*\right). \tag{3.3}$$

Note that $F(\boldsymbol{c}|Y)$ is a function that is continuous on a set $\mathcal{A} = \{(\boldsymbol{c}, Y) | \boldsymbol{c} \in \mathbb{R}^q, Y > O\}$ and $\Pr((\widehat{\boldsymbol{c}}_g, \widehat{R}_*) \in \mathcal{A}) = 1$. Thus, from (3.3) and multivariate continuous mapping theorem, we obtain $F(\widehat{\boldsymbol{c}}_g | \widehat{R}_*) \xrightarrow{p} F(\boldsymbol{c}_g | R)$. From this result and Corollary 2.1, we obtain the following theorem.

Theorem 3.1 *Under assumptions (C1), (A0), (A1), (A2') and (A4) or (C2), (A0)-(A4), it holds that* $\widehat{e_g} \xrightarrow{p} e_g$.

By applying Lebesgue's dominated convergence theorem to Theorem 3.1 since

$$|\widehat{e_g} - e_g| < 2 \quad a.e.,$$

we get the following corollary.

Corollary 3.1 *Under assumptions (C1), (A0), (A1), (A2') and (A4) or (C2), (A0)-(A4), it holds that*

$$\mathrm{E}\left[\widehat{e_g}\right] = e_g + o(1).$$

4 Numerical Results

Now, we investigate numerical performances of the consistent estimator $\widehat{e}_g$ by Monte Carlo simulation. First, we investigate the accuracy of the asymptotic approximations

$$\text{(HYF)} \ : \ e_1 \approx 1 - F\left(\widetilde{\boldsymbol{c}}_1 | \widetilde{R}\right), \quad \text{(AY)} \ : \ e_1 \approx \sum_{h=2}^{3} \Phi\left(-\frac{\boldsymbol{\delta}_{1h}'\boldsymbol{\delta}_{1h}}{\delta_{h1}}\right),$$

where

$$\widetilde{\boldsymbol{c}}_1 \ = \ \left(-\frac{\boldsymbol{\delta}_{12}'\boldsymbol{\delta}_{12}}{\sigma_{21}}, -\frac{\boldsymbol{\delta}_{13}'\boldsymbol{\delta}_{13}}{\sigma_{31}}\right)', \ \widetilde{R} = \begin{pmatrix} 1 & \rho_{21,31} \\ \rho_{31,21} & 1 \end{pmatrix}.$$

Here, the approximation (HYF) represents our proposed method based on (2.3), and the approximation (AY) represents the method proposed by Aoshima and Yata (2014). The misclassification probability e_1 is calculated via simulation with 100,000 replications, where in each step, the data sets are generated as

$$\begin{aligned} &\text{(Case I)} \quad : \boldsymbol{x}_1^{(g)}, \boldsymbol{x}_2^{(g)}, \ldots, \boldsymbol{x}_{N_g}^{(g)} \overset{i.i.d.}{\sim} \mathcal{N}_p(\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}) \ (g = 1, 2, 3), \\ &\text{(Case II)} \quad : \boldsymbol{x}_1^{(g)}, \boldsymbol{x}_2^{(g)}, \ldots, \boldsymbol{x}_{N_g}^{(g)} \overset{i.i.d.}{\sim} t_p(\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}, \nu) \ (g = 1, 2, 3), \end{aligned}$$

where $t_p(\boldsymbol{\mu}, \Sigma, \nu)$ denotes p-variate t-distribution with mean $\boldsymbol{\mu}$, covariance matrix Σ and degrees of freedom ν, $\boldsymbol{\mu}^{(1)} = \boldsymbol{0}$, $\boldsymbol{\mu}^{(2)} = (1, \ldots, 1, 0, \ldots, 0)'$ whose first $\left\lceil \sqrt{\mathrm{tr}\Sigma^{(1)2}} \right\rceil$ elements are 1 and $\boldsymbol{\mu}^{(2)} = (0, \ldots, 0, 1, \ldots, 1)'$ whose last $\left\lceil \sqrt{\mathrm{tr}\Sigma^{(1)2}} \right\rceil$ elements are 1 or $\boldsymbol{\mu}^{(1)} = \boldsymbol{0}$, $\boldsymbol{\mu}^{(2)} = (\sqrt{50/p}, \ldots, \sqrt{50/p})'$ and $\boldsymbol{\mu}^{(3)} = -\boldsymbol{\mu}^{(2)}$. Here,

$$\Sigma^{(1)} = B\left(\rho^{|i-j|^{1/3}}\right)B, \ \Sigma^{(2)} = 1.2B\left(\rho^{|i-j|^{1/3}}\right)B, \ \Sigma^{(3)} = 2.4B\left(\rho^{|i-j|^{1/3}}\right)B,$$

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 chose $p = 100, 250, 500, 1000$, $(N_1, N_2, N_3) = (20, 40, 60)$, $(40, 80, 120)$, $(60, 120, 180)$ and $\nu = 30$. Then we compare the true value e_1, the approximation (HYF) and the approximation (AY) on these settings. In comparison of the approximations in Table 1 and 2, it is seen that approximation (HYF) is closer to the true value e_1 than (AY) in most cases.

Next, we investigate the bias and mean squared error (MSE) of the consistent estimator $\widehat{e}_1$ on 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. Set for $j = 1, \ldots, N_1$

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

The set $X_1^{(-j)}$ represents the leave-one-out learning set, which is the collection of data with observation $\boldsymbol{x}_j^{(1)}$ removed. It calculates the rate of misclassification when predicting for each specimen using a learning set containing all other observations in the sample. We define the discriminant function using the learning set by

$$W_{h1}^{(-j)}(\boldsymbol{x}_j^{(1)}) = \left\|\boldsymbol{x}_j^{(1)} - \overline{\boldsymbol{x}}^{(h)}\right\|^2 - \left\|\boldsymbol{x}_j^{(1)} - \overline{\boldsymbol{x}}_{(-j)}^{(1)}\right\|^2 - \mathrm{tr}\left[\frac{1}{N_h}S^{(h)} - \frac{1}{N_1 - 1}S_{(-j)}^{(1)}\right], \ (h = 2, 3).$$

where $\bar{\boldsymbol{x}}_{(-j)}^{(1)}$ and $S_{(-j)}^{(1)}$ are calculated like procedures given around (1.1) based on the learning set $X_1^{(-j)}$. Then the CV estimator of e_1 is given by

$$CV(1) = 1 - \frac{1}{N_1}\sum_{j=1}^{N_1} I_{\left\{\boldsymbol{x}_j | W_{21}^{(-j)}(\boldsymbol{x}_j^{(1)})>0, \ W_{31}^{(-j)}(\boldsymbol{x}_j^{(1)})>0\right\}}(\boldsymbol{x}_j),$$

The biases and MSEs of the estimators $CV(1)$ and $\widehat{e}_1$ are given in Table 3 and 4 for Case I and in Table 5 and 6 for Case II. These tables show that $\widehat{e}_1$ has smaller MSEs than $CV(1)$ in most cases.

5 Concluding Remarks

We considered the multi-class classification problem for high-dimensional data. In this paper, we showed the asymptotic multivariate normality for several Euclidean distance-based discriminant functions under high-dimensional settings where the number of parameters exceeds the sample size. By using this results and new estimator of bilinear form, we also proposed the asymptotically unbiased and consistent estimator. Our theoretical results have the advantage of establishing under variance heterogeneity and nonnormality. We confirmed that proposed estimator have good performances in high-dimensional situation by numerical simulations.

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 $W_{hg}(\boldsymbol{x})$, and variances of unbiased estimators $\widehat{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}$, $\mathrm{tr}\widehat{\Sigma^{(g)}\Sigma^{(h)}}$, $\widehat{\mathrm{tr}\Sigma^{(g)2}}$, $\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh}$ and $\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh'}$.

Lemma A. 1 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

$$
\begin{aligned}
&\text{(i)} \quad \mathrm{E}[\overline{z}_i^{(g)^4}] = \frac{\kappa_{4i}^{(g)}+3N_g}{N_g^3},\\
&\text{(ii)} \quad \mathrm{E}[(\overline{\boldsymbol{z}}^{(g)\prime}A\overline{\boldsymbol{z}}^{(g)})^2] \le \frac{\kappa_{4\max}^{(g)}+2N_g}{N_g^3}\mathrm{tr}A^2+\frac{1}{N_g^2}(\mathrm{tr}A)^2,\\
&\text{(iii)} \quad \mathrm{E}[(\boldsymbol{z}_j^{(g)\prime}A\boldsymbol{z}_j^{(g)})^2] = \sum_{i=1}^{p}\kappa_{4i}^{(g)}a_{ii}^2+2\mathrm{tr}A^2+(\mathrm{tr}A)^2,\\
&\text{(iv)} \quad \mathrm{E}[(\boldsymbol{z}_j^{(g)\prime}A\boldsymbol{z}_k^{(g)})^4] \le (\kappa_{4\max}^{(g)}+3)\left\{(\kappa_{4\max}^{(g)}+2)\mathrm{tr}A^4+(\mathrm{tr}A^2)^2\right\}.
\end{aligned}
$$

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

Lemma A. 2 The variance of $W_{hg}(\boldsymbol{x})$ is

$$
\mathrm{Var}[W_{hg}(\boldsymbol{x})] = \sigma_{gh}^2+\frac{2\mathrm{tr}\Sigma^{(g)^2}}{N_g(N_g-1)}+\frac{2\mathrm{tr}\Sigma^{(h)^2}}{N_h(N_h-1)}+\frac{4\boldsymbol{\delta}'_{gh}\Sigma^{(h)}\boldsymbol{\delta}_{gh}}{N_h},
$$

where

$$
\sigma_{hg}^2 \;=\; 4\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}+\frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2}+\frac{4}{N_h}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}.
$$

Proof. $W_{hg}(\boldsymbol{x})$ is decomposed as

$$
W_{hg}(\boldsymbol{x}) = \boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}+W_{hg}^{(1)}+W_{hg}^{(2)},
$$

where

$$
\begin{aligned}
W_{hg}^{(1)} &= 2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\boldsymbol{\delta}_{gh}+2(\boldsymbol{x}-\boldsymbol{\mu}^{(g)})'\{(\overline{\boldsymbol{x}}^{(g)}-\boldsymbol{\mu}^{(g)})-(\overline{\boldsymbol{x}}^{(h)}-\boldsymbol{\mu}^{(h)})\},\\
W_{hg}^{(2)} &= -\frac{1}{N_g(N_g-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_g}(\boldsymbol{x}_j^{(g)}-\boldsymbol{\mu}^{(g)})'(\boldsymbol{x}_k^{(g)}-\boldsymbol{\mu}^{(g)})\\
&\quad +\frac{1}{N_h(N_h-1)}\sum_{\substack{j,k=1\\ j\neq k}}^{N_h}(\boldsymbol{x}_j^{(h)}-\boldsymbol{\mu}^{(h)})'(\boldsymbol{x}_k^{(h)}-\boldsymbol{\mu}^{(h)})-2\boldsymbol{\delta}'_{gh}(\overline{\boldsymbol{x}}^{(h)}-\boldsymbol{\mu}^{(h)}).
\end{aligned}
$$

It can be shown that

$$\begin{aligned}
\mathrm{Var}\left[W_{hg}^{(1)}\right] &= 4\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} + \frac{4}{N_g}\mathrm{tr}\Sigma^{(g)^2} + \frac{4}{N_h}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}(=\sigma_g^2), \\
\mathrm{Var}\left[W_{hg}^{(2)}\right] &= \frac{2}{N_g(N_g-1)}\mathrm{tr}\Sigma^{(g)^2} + \frac{2}{N_h(N_h-1)}\mathrm{tr}\Sigma^{(h)^2} + \frac{4}{N_h}\boldsymbol{\delta}'_{gh}\Sigma^{(h)}\boldsymbol{\delta}_{gh}
\end{aligned}$$

and $\mathrm{Cov}(W_{hg}^{(1)}, W_{hg}^{(2)}) = 0$. From the above results, the proof of Lemma A.2 is complete. □

Lemma A. 3 Under (C1), (A0), (A1), (A2') or (C2), (A0)-(A3), it holds that

$$\begin{aligned}
&\text{(i)} \quad \mathrm{Var}[\widehat{\boldsymbol{\delta}'_{gh}\boldsymbol{\delta}_{gh}}] = \frac{2}{N_g(N_g-1)}\mathrm{tr}\Sigma^{(g)^2} + \frac{2}{N_h(N_h-1)}\mathrm{tr}\Sigma^{(h)^2} \\
&\qquad + \frac{4}{N_gN_h}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)} + \frac{4}{N_g}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} + \frac{4}{N_h}\boldsymbol{\delta}'_{gh}\Sigma^{(h)}\boldsymbol{\delta}_{gh}, \\
&\text{(ii)} \quad \mathrm{Var}[\mathrm{tr}\widehat{\Sigma^{(g)}\Sigma^{(h)}}] = O\left(\left(\frac{1}{N_g} + \frac{1}{N_h}\right)\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2 \right. \\
&\qquad \left. + \frac{1}{N_gN_h}(\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)})^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}[\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh}] = O\left(\frac{1}{N_g}\left(\frac{1}{N_g} + \frac{1}{N_h}\right)^2(\mathrm{tr}\Sigma^{(g)^2})^2 \right. \\
&\qquad \left. + \frac{1}{N_g}\left(\frac{1}{N_g} + \frac{1}{N_h}\right)\mathrm{tr}\Sigma^{(g)^2}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} + \frac{1}{N_g}(\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh})^2\right) \\
&\qquad + o\left(\left(\frac{1}{N_g} + \frac{1}{N_h}\right)^2(\mathrm{tr}\Sigma^{(g)^2})^2 \right. \\
&\qquad \left. + \left(\frac{1}{N_g} + \frac{1}{N_h}\right)\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\mathrm{tr}\Sigma^{(g)^2}\right), \\
&\text{(v)} \quad \mathrm{Var}[\boldsymbol{\delta}'_{gh}\widehat{\Sigma^{(g)}}\boldsymbol{\delta}_{gh'}] = O\left(\frac{1}{N_g}\left(\frac{1}{N_g} + \frac{1}{N_h}\right)\left(\frac{1}{N_g} + \frac{1}{N_{h'}}\right)(\mathrm{tr}\Sigma^{(g)^2})^2 \right. \\
&\qquad + \frac{1}{N_g}\left(\frac{1}{N_g} + \frac{1}{N_h} + \frac{1}{N_{h'}}\right)\mathrm{tr}\Sigma^{(g)^2}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh} \\
&\qquad \left. + \frac{1}{N_g}(\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'})^2\right) \\
&\qquad + o\left(\left(\frac{1}{N_g} + \frac{1}{N_h}\right)\left(\frac{1}{N_g} + \frac{1}{N_{h'}}\right)(\mathrm{tr}\Sigma^{(g)^2})^2 \right. \\
&\qquad \left. + \left(\frac{1}{N_g} + \frac{1}{N_h} + \frac{1}{N_{h'}}\right)\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\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 proof of (iii), see e.g. Chen et al. (2010). For the proof of (iv), see e.g. Yamada et al. (2015). We give the proof of (v). Let $\boldsymbol{y}_j^{(g)} = \boldsymbol{x}_j^{(g)} - \boldsymbol{\mu}^{(g)}$. From the definition of $\widehat{\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}}$, this statistic can be expressed as $\widehat{\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}} = \sum_{\alpha=1}^{18} Y_\alpha$, where

$$
\begin{aligned}
Y_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)}, \\
Y_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)}, \\
Y_3 &= -\frac{1}{N_g N_h(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \sum_{\ell=1}^{N_h} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(h)}, \\
Y_4 &= -\frac{1}{N_g N_{h'}(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \sum_{\ell=1}^{N_{h'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_\ell^{(h')}, \\
Y_5 &= \frac{1}{N_g(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \left(\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}^{(h)}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}, \\
Y_6 &= \frac{1}{N_g(N_g-1)} \sum_{\substack{j,k=1 \\ j\neq k}}^{N_g} \left(\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}^{(h')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}, \\
Y_7 &= \frac{1}{N_g N_h(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_h} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_\ell^{(g)\prime}\boldsymbol{y}_m^{(h)}, \\
Y_8 &= \frac{1}{N_g N_{h'}(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_{h'}} \boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_\ell^{(g)\prime}\boldsymbol{y}_m^{(h')}, \\
Y_9 &= -\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} \left(\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}^{(h)}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(g)}, \\
Y_{10} &= -\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} \left(\boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}^{(h')}\right)' \boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(g)}, \\
Y_{11} &= \frac{1}{N_g N_h N_{h'}} \sum_{j=1}^{N_h} \sum_{k=1}^{N_g} \sum_{\ell=1}^{N_{h'}} \boldsymbol{y}_j^{(h)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(h')},
\end{aligned}
$$

$$
\begin{aligned}
Y_{12} &= -\frac{1}{N_g N_h}\sum_{j=1}^{N_g}\sum_{k=1}^{N_h}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h')}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(h)},\\
Y_{13} &= -\frac{1}{N_g N_h N_{h'}(N_g-1)}\sum_{j=1}^{N_h}\sum_{\substack{k,\ell=1\\k\neq\ell}}^{N_g}\sum_{m=1}^{N_{h'}}\boldsymbol{y}_j^{(h)\prime}\boldsymbol{y}_k^{(g)}\boldsymbol{y}_\ell^{(g)\prime}\boldsymbol{y}_m^{(h')},\\
Y_{14} &= \frac{1}{N_g N_h(N_g-1)}\sum_{\substack{j,k=1\\j\neq k}}^{N_g}\sum_{\ell=1}^{N_h}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h')}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(h)},\\
Y_{15} &= -\frac{1}{N_g N_{h'}}\sum_{j=1}^{N_g}\sum_{k=1}^{N_{h'}}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h)}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\boldsymbol{y}_k^{(h')},\\
Y_{16} &= \frac{1}{N_g}\sum_{j=1}^{N_g}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h)}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_j^{(g)\prime}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h')}\right),\\
Y_{17} &= \frac{1}{N_g N_{h'}(N_g-1)}\sum_{\substack{j,k=1\\j\neq k}}^{N_g}\sum_{\ell=1}^{N_{h'}}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h)}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\boldsymbol{y}_\ell^{(h')},\\
Y_{18} &= -\frac{1}{N_g(N_g-1)}\sum_{\substack{j,k=1\\j\neq k}}^{N_g}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h)}\right)'\boldsymbol{y}_j^{(g)}\boldsymbol{y}_k^{(g)\prime}\left(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu}^{(h')}\right).
\end{aligned}
$$

The variances of Y_α(for $\alpha = 1,\dots,18$) are derived as

$$
\begin{aligned}
\mathrm{Var}[Y_1] &= O\left(\frac{1}{N_g^3}(\mathrm{tr}\Sigma^{(g)^2})^2+\frac{1}{N_g^2}\mathrm{tr}\Sigma^{(g)^4}\right),\\
\mathrm{Var}[Y_2] &= O\left(\frac{1}{N_g^4}(\mathrm{tr}\Sigma^{(g)^2})^2\right),\\
\mathrm{Var}[Y_3] &= O\left(\frac{1}{N_g^2 N_h}(\mathrm{tr}\Sigma^{(g)^2})^2+\frac{1}{N_g N_h}\mathrm{tr}\Sigma^{(g)^3}\Sigma^{(h)}\right),\\
\mathrm{Var}[Y_4] &= O\left(\frac{1}{N_g^2 N_{h'}}(\mathrm{tr}\Sigma^{(g)^2})^2+\frac{1}{N_g N_{h'}}\mathrm{tr}\Sigma^{(g)^3}\Sigma^{(h')}\right),\\
\mathrm{Var}[Y_5] &= O\left(\frac{1}{N_g^2}\boldsymbol{\delta}_{gh}'\Sigma^{(g)}\boldsymbol{\delta}_{gh}\mathrm{tr}\Sigma^{(g)^2}+\frac{1}{N_g^3}\boldsymbol{\delta}_{gh}'\Sigma^{(g)}\boldsymbol{\delta}_{gh}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[Y_6] &= O\left(\frac{1}{N_g^2}\boldsymbol{\delta}_{gh'}'\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\mathrm{tr}\Sigma^{(g)^2}+\frac{1}{N_g^3}\boldsymbol{\delta}_{gh'}'\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[Y_7] &= O\left(\frac{1}{N_g^3 N_h}(\mathrm{tr}\Sigma^{(g)^2})^2\right),\\
\mathrm{Var}[Y_8] &= O\left(\frac{1}{N_g^3 N_{h'}}(\mathrm{tr}\Sigma^{(g)^2})^2\right),
\end{aligned}
$$

$$\begin{aligned}
\mathrm{Var}[Y_9] &= O\left(\frac{1}{N_g^3}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\mathrm{tr}\Sigma^{(g)^2} + \frac{1}{N_g^3}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[Y_{10}] &= O\left(\frac{1}{N_g^3}\boldsymbol{\delta}'_{gh'}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\mathrm{tr}\Sigma^{(g)^2} + \frac{1}{N_g^3}\boldsymbol{\delta}'_{gh'}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\left(\mathrm{tr}\Sigma^{(g)^4}\right)^{1/2}\right),\\
\mathrm{Var}[Y_{11}] &= O\left(\frac{1}{N_gN_hN_{h'}}(\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)})(\mathrm{tr}\Sigma^{(g)}\Sigma^{(h')}) + \frac{1}{N_hN_{h'}}\mathrm{tr}(\Sigma^{(g)^2}\Sigma^{(h)}\Sigma^{(h')})\right),\\
\mathrm{Var}[Y_{12}] &= O\left(\frac{1}{N_gN_h}\boldsymbol{\delta}'_{gh'}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\mathrm{tr}\Sigma^{(g)^2} + \frac{1}{N_h}\boldsymbol{\delta}'_{gh'}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\left(\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h)})^2\right)^{1/2}\right),\\
\mathrm{Var}[Y_{13}] &= O\left(\frac{1}{N_g^2N_hN_{h'}}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h')}\right),\\
\mathrm{Var}[Y_{14}] &= O\left(\frac{1}{N_g^3N_h}\boldsymbol{\delta}'_{gh'}\Sigma^{(g)}\boldsymbol{\delta}_{gh'}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h)}\right),\\
\mathrm{Var}[Y_{15}] &= O\left(\frac{1}{N_gN_{h'}}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\mathrm{tr}\Sigma^{(g)^2} + \frac{1}{N_{h'}}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\left(\mathrm{tr}(\Sigma^{(g)}\Sigma^{(h')})^2\right)^{1/2}\right),\\
\mathrm{Var}[Y_{16}] &= O\left(\frac{1}{N_g}(\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'})^2\right),\\
\mathrm{Var}[Y_{17}] &= O\left(\frac{1}{N_g^3N_{h'}}\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh}\mathrm{tr}\Sigma^{(g)}\Sigma^{(h')}\right),\\
\mathrm{Var}[Y_{18}] &= O\left(\frac{1}{N_g^2}(\boldsymbol{\delta}'_{gh}\Sigma^{(g)}\boldsymbol{\delta}_{gh'})^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 rate adjusted classifier for multiclass, high-dimensional data, *Ann. Inst. Stat. 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] Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample sizes and dimensionality are large. *J. Multivar. Anal.*, **73**, 1-17.

[6] Fujikoshi, Y. and Seo, T. (1998). Asymptotic aproximations for EPMC's of the linear and the quadratic discriminant functions when the sample sizes and the dimension are large. *Random Operators and Stochastic Equations*, **6**, 269-280.

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

[8] 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.

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

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

[11] 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.

[12] Yamada, Y., Hyodo, M. and Seo, T. (2015). Estimation of misclassification probability for a distance-based classifier in high-dimensional data. Technical Report, No.15-06, *Hiroshima statistical research group*, *Hiroshima University.*

Table 1: Comparison of approximations where $\boldsymbol{\mu}^{(2)} = (\sqrt{50/p}, \ldots, \sqrt{50/p})'$

p			(N_1, N_2, N_3) (20,40,80)	(40,80,120)	(60,120,180)
100	e_1	Case I	0.1462	0.1407	0.1385
		Case II	0.1442	0.1400	0.1373
	approx	HYF	0.1455	0.1406	0.1390
		AY	0.0000	0.0000	0.0000
250	e_1	Case I	0.1889	0.1766	0.1718
		Case II	0.1858	0.1732	0.1674
	approx	HYF	0.1880	0.1767	0.1729
		AY	0.0000	0.0000	0.0000
500	e_1	Case I	0.2226	0.2011	0.1926
		Case II	0.2195	0.1945	0.1883
	approx	HYF	0.2210	0.2003	0.1931
		AY	0.0038	0.0000	0.0000
1000	e_1	Case I	0.2650	0.2276	0.2155
		Case II	0.2563	0.2229	0.2069
	approx	HYF	0.2631	0.2274	0.2140
		AY	0.0408	0.0036	0.0004

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

p			(N_1, N_2, N_3) (20,40,80)	(40,80,120)	(60,120,180)
100	e_1	Case I	0.2006	0.1890	0.1836
		Case II	0.1973	0.1843	0.1782
	approx	HYF	0.2014	0.1897	0.1856
		AY	0.0007	0.0000	0.0000
250	e_1	Case I	0.1500	0.1372	0.1306
		Case II	0.1460	0.1333	0.1259
	approx	HYF	0.1494	0.1355	0.1308
		AY	0.0006	0.0000	0.0000
500	e_1	Case I	0.1175	0.1026	0.0973
		Case II	0.1151	0.0991	0.0957
	approx	HYF	0.1163	0.1011	0.0960
		AY	0.0007	0.0000	0.0000
1000	e_1	Case I	0.0881	0.0722	0.0662
		Case II	0.0899	0.0737	0.0663
	approx	HYF	0.0869	0.0712	0.0661
		AY	0.0007	0.0000	0.0000

Table 3: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (\sqrt{50/p}, \ldots, \sqrt{50/p})'$ in Case I

			(N_1, N_2, N_3)		
p			(20,40,60)	(40,80,120)	(60,120,180)
100	Bias	$\widehat{e}_1$	0.0008	0.0007	0.0008
		$CV(1)$	0.0006	0.0006	0.0007
	MSE	$\widehat{e}_1$	0.0040	0.0020	0.0014
		$CV(1)$	0.0065	0.0031	0.0021
250	Bias	$\widehat{e}_1$	0.0002	0.0005	0.0013
		$CV(1)$	0.0012	0.0009	0.0015
	MSE	$\widehat{e}_1$	0.0048	0.0024	0.0016
		$CV(1)$	0.0079	0.0038	0.0025
500	Bias	$\widehat{e}_1$	-0.0003	-0.0007	0.0003
		$CV(1)$	0.0020	-0.0002	0.0006
	MSE	$\widehat{e}_1$	0.0052	0.0025	0.0017
		$CV(1)$	0.0087	0.0041	0.0027
1000	Bias	$\widehat{e}_1$	-0.0013	0.0004	-0.0014
		$CV(1)$	0.0028	0.0017	-0.0007
	MSE	$\widehat{e}_1$	0.0057	0.0026	0.0017
		$CV(1)$	0.0095	0.0045	0.0029

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

			(N_1, N_2, N_3)		
p			(20,40,80)	(40,80,120)	(60,120,180)
100	Bias	$\widehat{e}_1$	-0.0011	-0.0007	-0.0017
		$CV(1)$	-0.0009	-0.0008	-0.0018
	MSE	$\widehat{e}_1$	0.0036	0.0016	0.0010
		$CV(1)$	0.0067	0.0031	0.0021
250	Bias	$\widehat{e}_1$	0.0020	-0.0020	-0.0004
		$CV(1)$	0.0025	-0.0021	-0.0006
	MSE	$\widehat{e}_1$	0.0029	0.0013	0.0008
		$CV(1)$	0.0055	0.0025	0.0016
500	Bias	$\widehat{e}_1$	0.0006	0.0008	0.0005
		$CV(1)$	0.0002	0.0002	0.0000
	MSE	$\widehat{e}_1$	0.0023	0.0009	0.0006
		$CV(1)$	0.0045	0.0020	0.0012
1000	Bias	$\widehat{e}_1$	0.0030	-0.0004	0.0012
		$CV(1)$	0.0021	-0.0016	0.0005
	MSE	$\widehat{e}_1$	0.0018	0.0007	0.0004
		$CV(1)$	0.0037	0.0015	0.0009

Table 5: Comparison of Biases and MSEs where $\boldsymbol{\mu}^{(2)} = (\sqrt{50/p}, \ldots, \sqrt{50/p})'$ in Case II

			(N_1, N_2, N_3)		
p			(20,40,60)	(40,80,120)	(60,120,180)
100	Bias	$\widehat{e}_1$	0.0033	0.0012	0.0005
		$CV(1)$	0.0001	-0.0018	-0.0022
	MSE	$\widehat{e}_1$	0.0043	0.0022	0.0015
		$CV(1)$	0.0064	0.0031	0.0021
250	Bias	$\widehat{e}_1$	0.0026	0.0038	0.0055
		$CV(1)$	-0.0005	0.0001	0.0016
	MSE	$\widehat{e}_1$	0.0052	0.0026	0.0018
		$CV(1)$	0.0078	0.0037	0.0025
500	Bias	$\widehat{e}_1$	0.0016	0.0049	0.0044
		$CV(1)$	-0.0015	0.0017	0.0000
	MSE	$\widehat{e}_1$	0.0056	0.0028	0.0019
		$CV(1)$	0.0086	0.0041	0.0027
1000	Bias	$\widehat{e}_1$	0.0058	0.0044	0.0067
		$CV(1)$	0.0037	-0.0003	-0.0021
	MSE	$\widehat{e}_1$	0.0061	0.0029	0.0020
		$CV(1)$	0.0094	0.0044	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, N_3)		
p			(20,40,80)	(40,80,120)	(60,120,180)
100	Bias	$\widehat{e}_1$	0.0027	0.0048	0.0069
		$CV(1)$	-0.0014	0.0000	0.0022
	MSE	$\widehat{e}_1$	0.0039	0.0018	0.0012
		$CV(1)$	0.0065	0.0031	0.0020
250	Bias	$\widehat{e}_1$	0.0047	0.0028	0.0052
		$CV(1)$	0.0022	-0.0003	0.0024
	MSE	$\widehat{e}_1$	0.0031	0.0013	0.0009
		$CV(1)$	0.0055	0.0024	0.0016
500	Bias	$\widehat{e}_1$	0.0039	0.0036	0.0011
		$CV(1)$	0.0018	0.0015	-0.0004
	MSE	$\widehat{e}_1$	0.0025	0.0011	0.0006
		$CV(1)$	0.0045	0.0020	0.0012
1000	Bias	$\widehat{e}_1$	0.0012	-0.0004	0.0010
		$CV(1)$	-0.0010	-0.0016	0.0003
	MSE	$\widehat{e}_1$	0.0019	0.0007	0.0005
		$CV(1)$	0.0036	0.0015	0.0009