Hiroshima Math. J.
49 (2019), 175–193

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

Hiroki Watanabe, Masashi Hyodo, Yuki Yamada and Takashi Seo

(Received August 24, 2016)
(Revised January 23, 2019)

Abstract. We estimate the misclassification probability of a Euclidean distance-based classifier in high-dimensional data. We discuss two types of estimator: a plug-in type estimator based on the normal approximation of misclassification probability (newly proposed), and an estimator based on the well-known leave-one-out cross-validation method. Both estimators perform consistently when the dimension exceeds the total sample size, and the underlying distribution need not be multivariate normality. We also numerically determine the mean squared errors (MSEs) of these estimators in finite sample applications of high-dimensional scenarios. The newly proposed plug-in type estimator gives smaller MSEs than the estimator based on leave-one-out cross-validation in simulation.

1. Introduction

We discuss a discrimination problem that allocates a given object $\mathbf{x}$ to one of two populations, G_1 and G_2. Here $\mathbf{x}$ is a continuous random vector (such as an observation vector) represented by a set of features $(x_1, x_2, \dots, x_p)$.

We assume a training data set $(\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1}, \mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_1})$, where $\mathbf{x}_{\ell j}$ is a p-dimensional continuous observation vector from the ℓ-th population G_ℓ, and we calculate

$$\forall_{\ell \in \{1,2\}} \bar{\mathbf{x}}_\ell = n_\ell^{-1} \sum_{j=1}^{n_\ell} \mathbf{x}_{\ell j}, \qquad \mathbf{S}_\ell = (n_\ell - 1)^{-1} \sum_{j=1}^{n_\ell} (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)^\top.$$

When $p \le n_1 + n_2 - 2$ and the population covariance matrices are equal, the data are often distinguished by Fisher's linear discriminant rule. In this paper, we examine a discrimination procedure that accommodates $p > \max\{n_1, n_2\}$ and heteroscedastic covariance matrices. Recently, Chan and Hall [3] and

---

The second author is supported by Grant-in-Aid for Young Scientists (B) (17K14238) from the Japan Society for the Promotion of Science.

2000 *Mathematics Subject Classification*. Primary 62H30, 62H12; Secondary 62E20.

Key words and phrases. asymptotic approximations, expected probability of misclassification, linear discriminant function.

Aoshima and Yata [1] studied a Euclidean distance-based classifier for the high-dimensional multiclass problem with different class covariance matrices. They introduced the following Euclidean distance discriminant function:

$$W = \{2\mathbf{x} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)\}^\top (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) + n_1^{-1}\,\mathrm{tr}(\mathbf{S}_1) - n_2^{-1}\,\mathrm{tr}(\mathbf{S}_2). \tag{1.1}$$

They also obtained a distance discriminant rule that assigns a new observation $\mathbf{x}$ to G_1 if $W > 0$, and to G_2 otherwise.

Here, we estimate the misclassification probability of the distance discriminant rule. The performance accuracy of the discriminant rule is represented by the resulting pair of misclassification error probabilities, defined as

$$e(2|1) = \Pr(W \le 0 \,|\, \mathbf{x} \sim G_1), e(1|2) = \Pr(W > 0 \,|\, \mathbf{x} \sim G_2).$$

Here, the notation "$\mathbf{x} \sim G_\ell$" means that $\mathbf{x}$ is generated from G_ℓ. Our main objective is to propose a consistent and asymptotically unbiased estimator of the misclassification probability in high-dimensional settings. To this end, we show the consistencies of two estimators in these settings. The first estimator is based on the well-known leave-one-out cross-validation (CV) method; the second is a plug-in estimator based on the normal approximation. We also compare the mean squared errors (MSEs) of these estimators in simulation studies.

The remainder of this paper is organized as follows. In Section 2, we derive plug-in estimator based on the normal approximation, and show the consistency of this estimator and the estimator based on leave-one-out CV in high-dimensional settings. In Section 3, we numerically validate the proposed estimators in several high-dimensional scenarios, in which p far exceeds the sample size. The paper concludes with Section 4. Some auxiliary lemmas and proofs are presented in the Appendix.

2. Estimators of misclassification probability

2.1. Statistical model. Assuming fixed $g, g' \in \{1,2\}$ and $g' \neq g$, we let $\mathbf{x} = \boldsymbol{\Sigma}_g^{1/2}\mathbf{z} + \boldsymbol{\mu}_g$. We further assume that $\forall_{\ell \in \{1,2\}, j \in \{1,2,\dots,n_\ell\}} \mathbf{x}_{\ell j} = \boldsymbol{\Sigma}_\ell^{1/2}\mathbf{z}_{\ell j} + \boldsymbol{\mu}_\ell$. Here, $\boldsymbol{\Sigma}_\ell$ is positive-semi-definite, and the random vectors $\mathbf{z}$, $\mathbf{z}_{11}, \mathbf{z}_{12}, \dots, \mathbf{z}_{1n_1}$, $\mathbf{z}_{21}, \mathbf{z}_{22}, \dots, \mathbf{z}_{2n_2}$ are independent and identically distributed (i.i.d.) random vectors such that $\mathrm{E}(\mathbf{z}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{z}) = \mathbf{I}_p$. We denote $\mathbf{z} = (z_1, z_2, \dots, z_p)^\top$, and consider two cases, (C1) and (C2), as follows.

(C1) $\mathrm{E}(z_i^4) = \kappa_4 + 3 < \infty$, $\mathrm{E}(z_{i_1}^2 z_{i_2}^2) = 1$, and $\mathrm{E}(z_{i_1} z_{i_2} z_{i_3} z_{i_4}) = 0$ $(i_1 \neq i_2, i_3, i_4)$.

(C2) $z_1, z_2, \dots, z_p$ are mutually independent, and $\mathrm{E}(z_i^4) = \kappa_4 + 3 < \infty$.

The condition (C1) means that each $\{z_i\}_{i=1}^p$ has a kind of pseudo-independence among its components. Obviously, if (C2) holds, then (C1) is trivially true. Note that (C1) and (C2) include multivariate normal populations.

2.2. Normal approximation of misclassification. In this subsection, we discuss the normal approximation of the misclassification probability. The misclassification probability is approximated as follows:

$$e(g'|g) \approx \Phi(-\mu/\sigma_g), \tag{2.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. From Lemma A.2 (see Appendix), we have

$$\mu = \mathrm{E}\{(-1)^{g+1}W\} = \boldsymbol{\delta}^\top \boldsymbol{\delta},$$

$$\sigma_g^2 = \mathrm{var}(W) = 4\{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + n_g^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^2) + n_{g'}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2) + n_{g'}^{-1}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_{g'}\boldsymbol{\delta}\} + 2\sum_{\ell=1}^{2}\{n_\ell(n_\ell-1)\}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2),$$

where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

The normal approximation is justified under some assumptions. For each $\ell \in \{1,2\}$, let n_ℓ be a function of p, i.e., $n_\ell = n_\ell(p)$. For any $\ell \in \{1,2\}$, let $\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}$ and $\mathrm{tr}\{(\boldsymbol{\Sigma}_\ell^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_\ell^{1/2}) \odot (\boldsymbol{\Sigma}_\ell^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_\ell^{1/2})\}$ be a function of p. For any $\ell, \ell' \in \{1,2\}$ and any $r \in \{1,2\}$, let $\mathrm{tr}\{(\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_{\ell'})^r\}$ be a function of p. Then we use the following conditions:

(A0) For all $\ell \in \{1,2\}$, $\lim_{p\to\infty} n_\ell(p) = \infty$.

(A1) For all $\ell \in \{1,2\}$, $\mathrm{tr}(\boldsymbol{\Sigma}_\ell^4)/\{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)\}^2 = o(1)$, $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)/\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2) \in (0,\infty)$.

(A2) $\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_{g'}\boldsymbol{\delta} = o(n_{g'}\sigma_g^2)$.

(A3) $\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} = o(\delta_g^2)$.

(A4) $\mathrm{tr}\{(\boldsymbol{\Sigma}_g^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_g^{1/2}) \odot (\boldsymbol{\Sigma}_g^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_g^{1/2})\} = o(\sigma_g^4)$.

Here, "$A \odot B$" denotes Hadamard product of same size matrices A and B, for a function $f(\cdot)$, "$f(p) \in (0,\infty)$ as $p\to\infty$" implies $\liminf_{p\to\infty} f(p) > 0$ and $\limsup_{p\to\infty} f(p) < \infty$, and

$$\delta_g^2 = 4\{n_g^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^2) + n_{g'}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\} + 2\sum_{\ell=1}^{2}\{n_\ell(n_\ell-1)\}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2).$$

The following theorem represents the asymptotic normality of $(-1)^{g+1}W$.

Theorem 2.1. *We assume* (A0)–(A2). *Then* (i) *and* (ii) *hold.*

(i) *Under* (C1) *and* (A3), $\{(-1)^{g+1}W - \mu\}/\delta_g \rightsquigarrow \mathcal{N}(0,1)$ *as* $p\to\infty$.

(ii) *Under* (C2) *and* (A4), $\{(-1)^{g+1}W - \mu\}/\sigma_g \rightsquigarrow \mathcal{N}(0,1)$ *as* $p\to\infty$.

Here, $\rightsquigarrow$ *denotes that the convergence in distribution.*

PROOF. Statement (i) has been demonstrated by Aoshima and Yata [1]. The proof of statement (ii) is given in Appendix B.

Note that under (A2) and (A3), $\sigma_g = \delta_g + o(\delta_g)$. Thus we obtain the following corollary.

COROLLARY 2.1. *Under* (C1) *and* (A0)–(A3), $\{(-1)^{g+1}W - \mu\}/\sigma_g \rightsquigarrow \mathcal{N}(0,1)$ *as* $p \to \infty$.

From Theorem 2.1 and Corollary 2.1, we propose the following proposition. This result represents the accuracy of approximation (2.1).

PROPOSITION 2.1. *We assume* (A0)–(A2) *and* $\mu/\sigma_g = O(1)$. *Then* (i) *and* (ii) *hold.*

(i) $$e(g'|g) - \Phi(-\mu/\delta_g) = \begin{cases} o(1) & \text{under (C1) and (A3).} \\ O(1) & \text{under (C2) and (A4).} \end{cases}$$

(ii) $$e(g'|g) - \Phi(-\mu/\sigma_g) = \begin{cases} o(1) & \text{under (C1) and (A3).} \\ o(1) & \text{under (C2) and (A4).} \end{cases}$$

REMARK 2.1. *We assume* (C1) *or* (C2). *Under* $\mu/\sigma_g \to \infty$, $e(g'|g) = o(1)$.

From Remark 2.1, we assume a sufficient condition that guarantees a non-zero limit value of the misclassification probability, i.e., $\mu/\sigma_g = O(1)$.

2.3. Estimator of misclassification probability and its consistency. Based on Proposition 2.1, we approximate the misclassification probability as $\Phi(-\mu/\sigma_g)$. To estimate the unknown values in μ and σ_g, we apply unbiased estimators.

Let $\ell, \ell' \in \{1,2\}$ and $\ell \neq \ell'$. Preliminarily, we introduce the unbiased estimators of μ, $\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)$, $\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)$ and $\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_\ell\boldsymbol{\delta}$ as follows:

$$\hat{\mu} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - n_1^{-1}\,\mathrm{tr}(\mathbf{S}_1) - n_2^{-1}\,\mathrm{tr}(\mathbf{S}_2),$$

$$\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2}) = \mathrm{tr}(\mathbf{S}_1\mathbf{S}_2),$$

$$\mathrm{tr}(\widehat{\boldsymbol{\Sigma}_\ell^2}) = \frac{(n_\ell - 1)[(n_\ell - 1)(n_\ell - 2)\,\mathrm{tr}(\mathbf{S}_\ell^2) + \{\mathrm{tr}(\mathbf{S}_\ell)\}^2 - n_\ell K_\ell]}{n_\ell(n_\ell - 2)(n_\ell - 3)},$$

$$\widehat{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}_\ell\boldsymbol{\delta}} = (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}_{\ell'})^\top\mathbf{S}_\ell(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}_{\ell'}) - \frac{2U_\ell}{(n_\ell - 1)(n_\ell - 2)} - \frac{\mathrm{tr}(\mathbf{S}_1\mathbf{S}_2)}{n_{\ell'}} + \frac{2n_\ell K_\ell - (n_\ell - 1)\{\mathrm{tr}(\mathbf{S}_\ell)\}^2 - (n_\ell - 1)^2\,\mathrm{tr}(\mathbf{S}_\ell^2)}{n_\ell(n_\ell - 2)(n_\ell - 3)},$$

where

$$K_\ell = (n_\ell - 1)^{-1} \sum_{j=1}^{n_\ell} \{(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)^\top (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)\}^2,$$

$$U_\ell = \sum_{j=1}^{n_\ell} (\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}}_{\ell'})^\top (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)^\top (\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell).$$

The unbiased estimator $\hat{\mu}$ has been used in the L^2 norm based on two-sample test (see for example Chen and Qin [4] or Aoshima and Yata [2]). The unbiased estimator $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)}$ was proposed by Himeno and Yamada [6]. The unbiased estimator $\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}$ is newly derived in the present paper. To show the consistency of the plug-in estimator based on the normal approximation, we investigate the leading variance term of these estimators (see Lemma A.3 in Appendix A).

These estimators provide the following estimator of σ_g^2:

$$\hat{\sigma}_g^2 = 4\{\max(0, \widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}}) + n_g^{-1}\, \widehat{\mathrm{tr}(\boldsymbol{\Sigma}_g^2)} + n_{g'}^{-1}\, \widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)} + n_{g'}^{-1} \max(0, \widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_{g'} \boldsymbol{\delta}})\}$$
$$+ 2 \sum_{\ell=1}^{2} \{n_\ell (n_\ell - 1)\}^{-1}\, \widehat{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)}.$$

Replacing the unknown values μ and σ_g^2 by their estimators, we propose $\widehat{e(g'|g)} = \Phi(-\hat{\mu}/\hat{\sigma}_g)$. The consistency of the estimator $\widehat{e(g'|g)}$ is demonstrated in the following proposition.

Proposition 2.2. *We assume* (A0)–(A2) *and* $\mu/\sigma_g = O(1)$. *Then*

$$\widehat{e(g'|g)} = \begin{cases} e(g'|g) + o_p(1) & \text{under (C1) and (A3).} \\ e(g'|g) + o_p(1) & \text{under (C2) and (A4).} \end{cases}$$

Proof. See Appendix C.

From $|\widehat{e(g'|g)} - e(g'|g)| < 1$ and Proposition 2.2, we obtain the following corollary.

Corollary 2.2. *We assume* (A0)–(A2) *and* $\mu/\sigma_g = O(1)$. *Then*

$$\mathrm{E}\{\widehat{e(g'|g)}\} = \begin{cases} e(g'|g) + o(1) & \text{under (C1) and (A3).} \\ e(g'|g) + o(1) & \text{under (C2) and (A4).} \end{cases}$$

2.4. Leave-one-out cross-validation method and its consistency. In this subsection, we consider the leave-one-out CV method, which is popularly used for estimating prediction errors in small samples. For $j \in \{1, 2, \ldots, n_g\}$, we consider the set

$$\mathbb{X}_g^{(-j)} = (\mathbf{x}_{g1}, \mathbf{x}_{g2}, \ldots, \mathbf{x}_{gj-1}, \mathbf{x}_{gj+1}, \ldots, \mathbf{x}_{gn_g}).$$

This set denotes the leave-one-out learning set, which is a collection of data with observation $\mathbf{x}_{gj}$ removed. In a prediction problem, CV calculates the probability of misclassifying a sample from all other observations in the sample. We define the discriminant function by

$$W_g^{(-j)} = \{2\mathbf{x}_{gj} - (\bar{\mathbf{x}}_{g(-j)} + \bar{\mathbf{x}}_{g'})\}^\top (\bar{\mathbf{x}}_{g(-j)} - \bar{\mathbf{x}}_{g'}) \\ + \{(n_g - 1)^{-1} \operatorname{tr}(\mathbf{S}_{g(-j)}) - n_{g'}^{-1} \operatorname{tr}(\mathbf{S}_{g'})\},$$

where $\bar{\mathbf{x}}_{g(-j)}$ and $\mathbf{S}_{g(-j)}$ are calculated by the procedures in (1.1) using the learning set $\mathbb{X}_g^{(-j)}$. The CV-based estimator is then given by

$$c(g'|g) = n_g^{-1} \sum_{j=1}^{n_g} I(W_g^{(-j)} < 0),$$

where the function $I(A)$ is the indicator function defined as

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true}, \\ 0 & \text{if } A \text{ is false}. \end{cases}$$

By straightforward calculation, we obtain

$$\mathrm{E}\{c(g'|g)\} = \Pr(W_g^{(-1)} < 0),$$

$$\operatorname{var}\{c(g'|g)\} = \Pr(W_g^{(-1)} < 0, W_g^{(-2)} < 0) - \{\Pr(W_g^{(-1)} < 0)\}^2 \\ + n_g^{-1}\{\Pr(W_g^{(-1)} < 0) - \Pr(W_g^{(-1)} < 0, W_g^{(-2)} < 0)\}.$$

Note that $c(g'|g)$ is consistent when

$$\Pr(W_g^{(-1)} < 0) - \Pr(W_g < 0) = o(1),$$

$$\Pr(W_g^{(-1)} < 0, W_g^{(-2)} < 0) - \{\Pr(W_g^{(-1)} < 0)\}^2 = o(1).$$

To confirm this statement in a high-dimensional setting, we must investigate the distribution of $W_g^{(-1)}$ and the joint distribution of $(W_g^{(-1)}, W_g^{(-2)})^\top$. The joint asymptotic normality of the random vector $(W_g^{(-1)}, W_g^{(-2)})^\top$ is given by the following lemmas:

LEMMA 2.1. *Under* (C1) *and* (A0)–(A3),

$$((W_g^{(-1)} - \mu)/\delta_g, (W_g^{(-2)} - \mu)/\delta_g)^\top \rightsquigarrow \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2).$$

PROOF. For the proof, see Appendix D.

LEMMA 2.2. *Under* (C2), (A0)–(A2), *and* (A4),

$$((W_g^{(-1)} - \mu)/\sigma_g, (W_g^{(-2)} - \mu)/\sigma_g)^\top \rightsquigarrow \mathcal{N}_2(\mathbf{0}, \mathbf{I}_2).$$

Proof. For the proof, see Appendix E.

From Lemmas 2.1 and 2.2, we obtain the following proposition.

Proposition 2.3. *We assume* (A0)–(A2), *and* $\mu/\sigma_g = O(1)$. *Then*

$$c(g'|g) = \begin{cases} e(g'|g) + o_p(1) & \text{under (C1) and (A3).} \\ e(g'|g) + o_p(1) & \text{under (C2) and (A4).} \end{cases}$$

3. Numerical results

In Monte Carlo simulations, we investigated the numerical performance of the approximation based on Proposition 2.1, and compared the consistencies of the estimators $\widehat{e(2|1)}$ and $c(2|1)$.

3.1. Accuracy of normal approximations. First, we investigate the accuracy of the normal approximations:

$$(\mathrm{I}) : e(2|1) \approx \Phi(-\mu/\delta_1), \qquad (\mathrm{II}) : e(2|1) \approx \Phi(-\mu/\sigma_1).$$

Approximation (I) was proposed in Aoshima and Yata [1], and approximation (II) is newly proposed in the present paper. The asymptotic property of these approximations is shown in (i) and (ii) of Proposition 2.1. The misclassification probability $e(2|1)$ was calculated in 100,000 replications of the Monte Carlo simulations. In each step, the data-sets were generated as

$$\forall_{j \in \{1,\dots,n_1\}} \mathbf{x}_{1j} = \boldsymbol{\Sigma}_1^{1/2} \mathbf{z}_{1j} + \boldsymbol{\mu}_1, \qquad \forall_{j \in \{1,\dots,n_2\}} \mathbf{x}_{2j} = \boldsymbol{\Sigma}_2^{1/2} \mathbf{z}_{2j} + \boldsymbol{\mu}_2,$$

where $\boldsymbol{\mu}_1 = \mathbf{0}$. In $\boldsymbol{\mu}_2$, the first $\lfloor \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)} \rfloor$ elements are $\sqrt{3} n_1^{-1/4}$, and all other elements are 0. Moreover,

$$\boldsymbol{\Sigma}_1 = \mathbf{B}(0.3^{|i-j|})\mathbf{B}, \qquad \boldsymbol{\Sigma}_2 = 1.2\mathbf{B}(0.3^{|i-j|})\mathbf{B}.$$

Here,

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

We considered the following four distributions of $\mathbf{z}_{gj} = (z_{gij})$. Note that the fourth moment of z_{gij} exists.

(A) Standard normal distribution: $z_{gij} \sim \mathcal{N}(0, 1)$,

(B) Standardized chi-squared distribution with 10 degrees of freedom:

$$z_{gij} = (u_{gij} - 10)/\sqrt{20} \qquad \text{for } u_{gij} \sim \chi^2_{10},$$

(C) Standardized t distribution with 10 degrees of freedom:

$$z_{gij} = u_{gij}/\sqrt{5/4} \qquad \text{for } u_{gij} \sim t_{10},$$

(D) Standardized skew normal distribution:

$$z_{gij} = \{1 - 9/(5\pi)\}^{-1/2}(u_{gij} - 3/\sqrt{5\pi}) \qquad \text{for } u_{gij} \sim \mathscr{SN}(3).$$

Setting $p \in \{50, 100, 200, 400, 800\}$ and $(n_1, n_2) \in \{(20, 40), (30, 30), (40, 20), (40, 80), (60, 60), (80, 40)\}$, we compared the $e(2|1)$ values calculated by the simulation, approximation (I), and approximation (II). The results are shown in Table 1. Comparing the tabulated approximations, we observe that in most

Table 1. Comparison of approximations

			(n_1, n_2)					
p			$(20, 40)$	$(30, 30)$	$(40, 20)$	$(40, 80)$	$(60, 60)$	$(80, 40)$
50	$e(2\|1)$	(A)	0.2071	0.2354	0.2696	0.2270	0.2529	0.2846
		(B)	0.2057	0.2332	0.2686	0.2246	0.2530	0.2847
		(C)	0.2025	0.2332	0.2685	0.2272	0.2507	0.2856
		(D)	0.2038	0.2355	0.2678	0.2251	0.2555	0.2842
	approx	(I)	0.1212	0.1590	0.2117	0.1199	0.1579	0.2102
		(II)	0.2072	0.2351	0.2682	0.2268	0.2542	0.2830
100	$e(2\|1)$	(A)	0.1908	0.2243	0.2616	0.2072	0.2385	0.2739
		(B)	0.1898	0.2185	0.2598	0.2071	0.2360	0.2694
		(C)	0.1915	0.2238	0.2584	0.2096	0.2369	0.2710
		(D)	0.1884	0.2234	0.2623	0.2087	0.2386	0.2708
	approx	(I)	0.1283	0.1662	0.2186	0.1269	0.1651	0.2171
		(II)	0.1922	0.2224	0.2595	0.2084	0.2382	0.2712
200	$e(2\|1)$	(A)	0.1689	0.2049	0.2478	0.1842	0.2178	0.2562
		(B)	0.1685	0.2033	0.2439	0.1861	0.2151	0.2529
		(C)	0.1686	0.2024	0.2456	0.1859	0.2154	0.2547
		(D)	0.1695	0.2045	0.2451	0.1846	0.2162	0.2530
	approx	(I)	0.1218	0.1597	0.2123	0.1205	0.1585	0.2108
		(II)	0.1709	0.2031	0.2439	0.1842	0.2162	0.2534
400	$e(2\|1)$	(A)	0.1634	0.1982	0.2393	0.1745	0.2092	0.2471
		(B)	0.1623	0.1986	0.2427	0.1742	0.2061	0.2495
		(C)	0.1641	0.1996	0.2402	0.1747	0.2079	0.2481
		(D)	0.1624	0.1949	0.2438	0.1727	0.2072	0.2493
	approx	(I)	0.1286	0.1666	0.2189	0.1273	0.1654	0.2174
		(II)	0.1639	0.1976	0.2412	0.1740	0.2075	0.2479
800	$e(2\|1)$	(A)	0.1468	0.1849	0.2299	0.1569	0.1893	0.2350
		(B)	0.1492	0.1821	0.2289	0.1537	0.1909	0.2331
		(C)	0.1500	0.1850	0.2299	0.1561	0.1927	0.2355
		(D)	0.1490	0.1835	0.2296	0.1543	0.1905	0.2345
	approx	(I)	0.1220	0.1598	0.2125	0.1207	0.1587	0.2110
		(II)	0.1484	0.1832	0.2293	0.1560	0.1908	0.2343

Table 2. Comparison of MSE $\times 10^3$ of estimators

p			(n_1, n_2)					
			$(20, 40)$	$(30, 30)$	$(40, 20)$	$(40, 80)$	$(60, 60)$	$(80, 40)$
50	(A)	$\widehat{e(2\|1)}$	**3.961**	**3.600**	**4.413**	**1.982**	**1.768**	**2.130**
		$c(2\|1)$	6.515	5.493	5.956	3.361	2.786	2.980
	(B)	$\widehat{e(2\|1)}$	**3.939**	**3.636**	**4.394**	**1.996**	**1.784**	**2.127**
		$c(2\|1)$	6.486	5.517	5.911	3.373	2.833	2.990
	(C)	$\widehat{e(2\|1)}$	**4.024**	**3.637**	**4.435**	**2.012**	**1.819**	**2.132**
		$c(2\|1)$	6.501	5.481	5.948	3.366	2.787	2.982
	(D)	$\widehat{e(2\|1)}$	**3.939**	**3.569**	**4.432**	**1.985**	**1.776**	**2.120**
		$c(2\|1)$	6.476	5.449	5.977	3.371	2.807	2.965
100	(A)	$\widehat{e(2\|1)}$	**3.748**	**3.555**	**4.459**	**1.963**	**1.786**	**2.156**
		$c(2\|1)$	6.296	5.445	6.047	3.293	2.772	2.988
	(B)	$\widehat{e(2\|1)}$	**3.786**	**3.520**	**4.448**	**1.982**	**1.803**	**2.177**
		$c(2\|1)$	6.308	5.375	5.960	3.266	2.780	2.981
	(C)	$\widehat{e(2\|1)}$	**3.758**	**3.665**	**4.464**	**1.984**	**1.802**	**2.163**
		$c(2\|1)$	6.280	5.389	5.955	3.316	2.770	2.991
	(D)	$\widehat{e(2\|1)}$	**3.796**	**3.556**	**4.388**	**1.984**	**1.777**	**2.163**
		$c(2\|1)$	6.352	5.389	5.926	3.321	2.771	2.953
200	(A)	$\widehat{e(2\|1)}$	**3.453**	**3.337**	**4.264**	**1.842**	**1.690**	**2.105**
		$c(2\|1)$	6.023	5.179	5.831	3.123	2.621	2.912
	(B)	$\widehat{e(2\|1)}$	**3.376**	**3.509**	**4.303**	**1.885**	**1.730**	**2.120**
		$c(2\|1)$	5.997	5.180	5.766	3.106	2.643	2.870
	(C)	$\widehat{e(2\|1)}$	**3.432**	**3.365**	**4.327**	**1.852**	**1.707**	**2.109**
		$c(2\|1)$	5.950	5.161	5.842	3.122	2.619	2.883
	(D)	$\widehat{e(2\|1)}$	**3.431**	**3.374**	**4.275**	**1.866**	**1.718**	**2.120**
		$c(2\|1)$	5.977	5.168	5.766	3.082	2.640	2.893
400	(A)	$\widehat{e(2\|1)}$	**3.703**	**3.274**	**4.346**	**1.781**	**1.696**	**2.155**
		$c(2\|1)$	5.922	5.176	5.890	3.043	2.624	2.935
	(B)	$\widehat{e(2\|1)}$	**3.283**	**3.279**	**4.270**	**1.793**	**1.714**	**2.150**
		$c(2\|1)$	5.910	5.194	5.836	3.039	2.627	2.925
	(C)	$\widehat{e(2\|1)}$	**3.273**	**3.285**	**4.263**	**1.776**	**1.708**	**2.121**
		$c(2\|1)$	5.894	5.187	5.876	3.050	2.635	2.903
	(D)	$\widehat{e(2\|1)}$	**3.312**	**3.359**	**4.406**	**1.785**	**1.706**	**2.138**
		$c(2\|1)$	5.956	5.234	5.834	3.055	2.622	2.922
800	(A)	$\widehat{e(2\|1)}$	**2.972**	**3.034**	**4.118**	**1.616**	**1.620**	**2.079**
		$c(2\|1)$	5.628	4.928	5.723	2.853	2.525	2.840
	(B)	$\widehat{e(2\|1)}$	**2.961**	**3.106**	**4.124**	**1.6378**	**1.627**	**2.090**
		$c(2\|1)$	5.575	4.953	5.672	2.868	2.540	2.864
	(C)	$\widehat{e(2\|1)}$	**2.917**	**3.053**	**4.117**	**1.620**	**1.603**	**2.083**
		$c(2\|1)$	5.538	4.955	5.707	2.835	2.513	2.846
	(D)	$\widehat{e(2\|1)}$	**2.927**	**3.060**	**4.128**	**1.618**	**1.607**	**2.092**
		$c(2\|1)$	5.575	4.958	5.725	2.832	2.504	2.862

cases, approximation (II) more closely approaches $e(2|1)$ than approximation (I). In addition, approximation (II) exhibits high stability when we vary the population distribution.

3.2. Accuracy of the estimators. Next, we computed the MSEs of the proposed estimator $\widehat{e(2|1)}$ and the previous estimator $c(2|1)$. The MSEs of both estimators are listed in Table 2. In all cases, the estimator $\widehat{e(2|1)}$ gives a smaller MSE than the estimator $c(2|1)$. Based on these simulation experiments, we therefore recommend estimator $\widehat{e(2|1)}$.

4. Conclusion

We proposed consistent and asymptotically unbiased estimators of misclassification probabilities in high-dimensional settings. Our proposed estimator was obtained by using a normal approximation of the misclassification probability. We confirmed the consistency of the proposed estimator under variance heterogeneity and non-normality (Proposition 2.2). We also showed the consistency of an estimator based on the leave-one-out CV method (Proposition 2.3). The MSEs of the two estimators were compared in numerical simulations. The estimator based on the normal approximation proved more accurate than the estimator based on leave-one-out CV.

Appendix

A. Preliminary. In this Appendix, we state some preliminary results.

LEMMA A.1. *Let* $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ *be i.i.d. random vectors that satisfy* (C1) *or* (C2), *and* $\bar{\mathbf{z}} = n^{-1}\sum_{j=1}^{n}\mathbf{z}_j$. *Then for any* $p \times p$ *real symmetric matrix* $\mathbf{A}$, *it holds that*

(i) $\forall_{i\in\{1,2,\dots,p\}}\mathrm{E}\{(\boldsymbol{e}_i^\top\bar{\mathbf{z}}_g)^4\} = n^{-3}(\kappa_4 + 3n)$,

(ii) $\mathrm{E}\{(\bar{\mathbf{z}}_g^\top\mathbf{A}\bar{\mathbf{z}}_g)^2\} = n^{-3}\kappa_4 \operatorname{tr}(\mathbf{A}\odot\mathbf{A}) + n^{-2}\{\operatorname{tr}(\mathbf{A})\}^2 + 2n^{-2}\operatorname{tr}(\mathbf{A}^2)$,

(iii) $\mathrm{E}\{(\mathbf{z}_1^\top\mathbf{A}\mathbf{z}_1)^2\} = \kappa_4 \operatorname{tr}(\mathbf{A}\odot\mathbf{A}) + \{\operatorname{tr}(\mathbf{A})\}^2 + 2\operatorname{tr}(\mathbf{A}^2)$,

(iv) $\mathrm{E}\{(\mathbf{z}_1^\top\mathbf{A}\mathbf{z}_2)^4\} = 6\operatorname{tr}(\mathbf{A}^4) + 3\{\operatorname{tr}(\mathbf{A}^2)\}^2 + 6\kappa_4\operatorname{tr}(\mathbf{A}^2\odot\mathbf{A}^2) + \kappa_4^2\operatorname{tr}\{(\mathbf{A}\odot\mathbf{A})^2\}$.

PROOF. The proof is routine and hence omitted here.

LEMMA A.2. *The variance of* W *is* σ_g^2.

PROOF. Define $(g, g') \in \{(1,2),(2,1)\}$. Let $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}_g$ and $\mathbf{y}_{\ell j} = \mathbf{x}_{\ell j} - \boldsymbol{\mu}_\ell$ for $\ell \in \{g, g'\}$. Then, $(-1)^{g+1}W$ can be expressed as $(-1)^{g+1}W = \mu + W_1 + W_2$, where

$$W_1 = 2(-1)^{g+1}\boldsymbol{\delta}^\top \mathbf{y} + 2(\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})^\top \mathbf{y},$$

$$W_2 = (-1)^g 2\boldsymbol{\delta}^\top \bar{\mathbf{y}}_{g'} + \{n_{g'}(n_{g'}-1)\}^{-1} \sum_{j_1,j_2=1,j_1\neq j_2}^{n_{g'}} \mathbf{y}_{g'j_1}^\top \mathbf{y}_{g'j_2} - \{n_g(n_g-1)\}^{-1} \sum_{j_1,j_2=1,j_1\neq j_2}^{n_g} \mathbf{y}_{gj_1}^\top \mathbf{y}_{gj_2}.$$

Here, $\bar{\mathbf{y}}_\ell = \bar{\mathbf{x}}_\ell - \boldsymbol{\mu}_\ell$. Since $\mathrm{E}(W_1) = \mathrm{E}(W_2) = 0$, we obtain $\mathrm{E}(W) = (-1)^{g-1}\mu$. Also, it can be shown that

$$\mathrm{var}(W_1) = 4\{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + n_g^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^2) + n_{g'}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\},$$

$$\mathrm{var}(W_2) = 2[\{n_g(n_g-1)\}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^2) + \{n_{g'}(n_{g'}-1)\}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_{g'}^2) + 2n_{g'}^{-1}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_{g'}\boldsymbol{\delta}],$$

and $\mathrm{cov}(W_1, W_2) = 0$. □

Lemma A.3 (The variance of some estimators). *We assume* (C1) *or* (C2). *Then* (i)–(iv) *hold.*

(i) *Under* (A0)–(A2), $\mathrm{var}(\hat{\mu}) = o(\sigma_g^2)$,

(ii) *Under* (A0) *and* (A1), $\mathrm{var}(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}) = o(n_{g'}^2\sigma_g^4)$,

(iii) *Under* (A0) *and* (A1), $\mathrm{var}(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)}) = o(n_\ell^2\sigma_\ell^4)$,

(iv) *Under* (A0) *and* (A1), $\mathrm{var}(\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}) = o(\sigma_\ell^4)$.

Proof. (i) is obtained by Section 6.1 in Chen and Qin [4]. (iii) is obtained by Lemma 1 in Himeno and Yamada [6]. (ii) is obtained by same route as (iii). We present only the proof of (iv). Let $\mathbf{y}_{\ell j} = \mathbf{x}_{\ell j} - \boldsymbol{\mu}_\ell$ and $\mathbf{y}_{\ell' j} = \mathbf{x}_{\ell' j} - \boldsymbol{\mu}_{\ell'}$. The statistic $\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}$ can be expressed as $\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}} = \sum_{\alpha=1}^{12} A_\alpha$, where

$$A_1 = \{n_\ell(n_\ell-1)(n_\ell-2)\}^{-1} \sum_{\substack{j_1,j_2,j_3=1\\ j_1\neq j_2, j_2\neq j_3, j_3\neq j_1}}^{n_\ell} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_2} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_3},$$

$$A_2 = -\{n_\ell(n_\ell-1)(n_\ell-2)(n_\ell-3)\}^{-1} \sum_{\substack{j_1,j_2,j_3,j_4=1\\ j_1\neq j_2\neq j_3\neq j_4\\ j_3\neq j_1\neq j_4\neq j_2}}^{n_\ell} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_2} \mathbf{y}_{\ell j_3}^\top \mathbf{y}_{\ell j_4},$$

$$A_3 = -2\{n_\ell(n_\ell-1)\}^{-1} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n_\ell} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_2} \mathbf{y}_{\ell j_1}^\top \bar{\mathbf{y}}_{\ell'},$$

$$A_4 = 2\{n_\ell(n_\ell-1)(n_\ell-2)\}^{-1} \sum_{\substack{j_1,j_2,j_3=1\\ j_1\neq j_2, j_2\neq j_3, j_3\neq j_1}}^{n_\ell} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_2} \mathbf{y}_{\ell j_3}^\top \bar{\mathbf{y}}_{\ell'},$$

$$A_5 = 2\{n_\ell n_{\ell'}(n_{\ell'} - 1)\}^{-1} \sum_{j_1=1}^{n_\ell} \sum_{\substack{j_2, j_3=1 \\ j_2 \neq j_3}}^{n_{\ell'}} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell' j_2} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell' j_3},$$

$$A_6 = -2\{n_\ell(n_\ell - 1)n_{\ell'}(n_{\ell'} - 1)\}^{-1} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{n_\ell} \sum_{\substack{j_3, j_4=1 \\ j_3 \neq j_4}}^{n_{\ell'}} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell' j_3} \mathbf{y}_{\ell j_2}^\top \mathbf{y}_{\ell' j_4},$$

$$A_7 = 2\{n_\ell(n_\ell - 1)\}^{-1} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j_1} \mathbf{y}_{\ell j_1}^\top \mathbf{y}_{\ell j_2},$$

$$A_8 = -2\{n_\ell(n_\ell - 1)(n_\ell - 2)\}^{-1} \sum_{\substack{j_1, j_2, j_3=1 \\ j_1 \neq j_2, j_2 \neq j_3, j_3 \neq j_1}}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j_1} \mathbf{y}_{\ell j_2}^\top \mathbf{y}_{\ell j_3},$$

$$A_9 = -2n_\ell^{-1} \sum_{j=1}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j} \mathbf{y}_{\ell j}^\top \bar{\mathbf{y}}_{\ell'},$$

$$A_{10} = 2\{n_\ell(n_\ell - 1)\}^{-1} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j_1} \mathbf{y}_{\ell j_2}^\top \bar{\mathbf{y}}_{\ell'},$$

$$A_{11} = n_\ell^{-1} \sum_{j=1}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j} \mathbf{y}_{\ell j}^\top (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}),$$

$$A_{12} = -\{n_\ell(n_\ell - 1)\}^{-1} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{n_\ell} (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \mathbf{y}_{\ell j_1} \mathbf{y}_{\ell j_2}^\top (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}).$$

The expectations of A_α are derived as $\mathrm{E}(A_\alpha) = 0$ $(\alpha \neq 11)$ and $\mathrm{E}(A_{11}) = \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}$. The variances of A_α are derived as follows:

$$\mathrm{var}(A_1) = O\left(\frac{\{\mathrm{tr}\boldsymbol{\Sigma}_\ell^2)\}^2}{n_\ell^3} + \frac{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^4)}{n_\ell^2}\right), \qquad \mathrm{var}(A_2) = O\left(\frac{\{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)\}^2}{n_\ell^4}\right),$$

$$\mathrm{var}(A_3) = O\left(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)\ \mathrm{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})}{n_\ell^2 n_{\ell'}} + \frac{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^4)}\sqrt{\mathrm{tr}\{(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})^2\}}}{n_\ell n_{\ell'}}\right),$$

$$\mathrm{var}(A_4) = O\left(\frac{\mathrm{tr}(\boldsymbol{\Sigma}_\ell^2)\ \mathrm{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})}{n_\ell^3 n_{\ell'}}\right),$$

$$\mathrm{var}(A_5) = O\left(\frac{\{\mathrm{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})\}^2}{n_\ell n_{\ell'}^2} + \frac{\mathrm{tr}\{(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})^2\}}{n_{\ell'}^2}\right),$$

$$\operatorname{var}(A_6) = O\left(\frac{\{\operatorname{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})\}^2}{n_\ell^2 n_{\ell'}^2}\right),$$

$$\operatorname{var}(A_7) = O\left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2)\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_\ell^2} + \frac{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_\ell^4)}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_\ell}\right),$$

$$\operatorname{var}(A_8) = O\left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}_\ell^2)\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_\ell^3}\right),$$

$$\operatorname{var}(A_9) = O\left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_\ell n_{\ell'}} + \frac{\sqrt{\operatorname{tr}\{(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})^2\}}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_{\ell'}}\right),$$

$$\operatorname{var}(A_{10}) = O\left(\frac{\operatorname{tr}(\boldsymbol{\Sigma}_\ell \boldsymbol{\Sigma}_{\ell'})\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta}}{n_\ell^2 n_{\ell'}}\right), \qquad \operatorname{var}(A_{11}) = O\left(\frac{(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta})^2}{n_\ell}\right),$$

$$\operatorname{var}(A_{12}) = O\left(\frac{(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_\ell \boldsymbol{\delta})^2}{n_\ell^2}\right).$$

Thus $\operatorname{var}(A_\alpha) = o(\sigma_g^4)$ for all $\alpha \in \{1, 2, \ldots, 12\}$. □

B. Proof of Theorem 2.1. Under conditions (C2) and (A0)–(A2), W_2 in the proof of Lemma A.2 is negligible. Thus $\{(-1)^{g+1} W - \mu\}/\sigma_g = \sum_{i=1}^p \epsilon_i + o_p(1)$, where $\epsilon_i = 2\{(-1)^{g+1}\boldsymbol{\delta} + (\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})\}^\top \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{e}_i z_i/\sigma_g$. Here,

$$\boldsymbol{e}_i = (0 \ldots 0\overset{i}{\check{1}}0 \ldots 0)^\top.$$

Defining $\mathscr{F}_0 = \sigma\{\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2\}$ and $\mathscr{F}_{i-1} = \sigma\{\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, z_1, z_2, \ldots, z_{i-1}\}$ $(2 \le i)$, it is straightforward to show that $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{E}(\epsilon_i | \mathscr{F}_{i-1}) = 0$. Thus, ϵ_i is a martingale difference sequence. To show the asymptotic normality of $\sum_{i=1}^p \epsilon_i$, we adapt the martingale difference central-limit theorem (see Shiryaev [7] or Hall and Heyde [5]). Now let $\sigma_{g\cdot i}^2 = \mathrm{E}(\epsilon_i^2 | \mathscr{F}_{i-1})$. To apply the martingale central-limit theorem, we need to show that (a): $\sum_{i=1}^p \sigma_{g\cdot i}^2 = 1 + o_p(1)$ and (b): $\sum_{i=1}^p \mathrm{E}(\epsilon_i^4) = o(1)$.

To show (a), we evaluate $\sigma_{g\cdot i}^2 = 4[\{(-1)^{g+1}\boldsymbol{\delta} + (\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})\}^\top \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{e}_i]^2/\sigma_g^2$, and

$$\sum_{i=1}^p \sigma_{g\cdot i}^2 = 4\sigma_g^{-2}\{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + 2(-1)^{g+1} R_1 + R_2\}.$$

Here, $R_1 = \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g(\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})$ and $R_2 = (\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})^\top \boldsymbol{\Sigma}_g(\bar{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'})$. Since $\mathrm{E}(R_1) = 0$ and $\mathrm{E}(R_2) = \mathrm{tr}(\boldsymbol{\Sigma}_g^2)/n_g + \mathrm{tr}(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})/n_{g'}$, we obtain

$$\mathrm{E}\left(\sum_{i=1}^{p} \sigma_{g\cdot i}^2\right) = 4\sigma_g^{-2}\{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + n_g^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^2) + n_{g'}^{-1}\,\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)\} = 1 + o(1).$$

To check (a), we need to show that $\mathrm{var}(R_1) = o(\sigma_g^4)$ and $\mathrm{var}(R_2) = o(\sigma_g^4)$. From Lemma A.1, these variances are given as follows:

$$\mathrm{var}(R_1) = O(n_g^{-1}\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_g^4)}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + n_{g'}^{-1}\sqrt{\mathrm{tr}\{(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})^2\}}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}),$$
$$\mathrm{var}(R_2) = O(n_g^{-2}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^4) + n_{g'}^{-2}\,\mathrm{tr}\{(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})^2\}).$$

Hence, under (A1), $\mathrm{var}(R_1) = o(\sigma_g^4)$ and $\mathrm{var}(R_2) = o(\sigma_g^4)$. Thus, under (A1), (a) holds.

To show (b), we decompose ϵ_i into the sum of three parts, $\epsilon_i = 2\{(-1)^{g+1}\epsilon_{1i} + \epsilon_{i2} - \epsilon_{i3}\}/\sigma_g$, where $\epsilon_{i1} = \boldsymbol{\delta}^\top \Sigma_g^{1/2}\boldsymbol{e}_i z_i$, $\epsilon_{i2} = \bar{\mathbf{y}}_g^\top \Sigma_g^{1/2}\boldsymbol{e}_i z_i$, and $\epsilon_{i3} = \bar{\mathbf{y}}_{g'}^\top \Sigma_g^{1/2}\boldsymbol{e}_i z_i$. Then, we need to show that $\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i\ell}^4) = o(\sigma_g^4)$ for $\ell \in \{1, 2, 3\}$. These expectations are given as follows:

$$\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i1}^4) = O(\mathrm{tr}\{(\boldsymbol{\Sigma}_g^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g^{1/2}) \odot (\boldsymbol{\Sigma}_g^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g^{1/2})\}),$$
$$\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i2}^4) = O(n_g^{-2}\,\mathrm{tr}(\boldsymbol{\Sigma}_g^4)), \qquad \sum_{i=1}^{p} \mathrm{E}(\epsilon_{i3}^4) = O(n_{g'}^{-2}\,\mathrm{tr}\{(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})^2\}).$$

Thus, $\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i1}^4) = o(\sigma_g^4)$ under (A4). Also, under (A1), $\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i2}^4) = o(\sigma_g^4)$ and $\sum_{i=1}^{p} \mathrm{E}(\epsilon_{i3}^4) = o(\sigma_g^4)$. These results complete the proof. □

C. Proof of Proposition 2.2. We assume (C1) or (C2). From Lemma A.3, under (A0)–(A2),

$$\hat{\mu} = \mu + o_p(\sigma_g), \tag{A. 1}$$

$$\frac{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}}{n_{g'}} = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}{n_{g'}} + o_p(\sigma_g^2), \qquad \frac{\widehat{\mathrm{tr}(\boldsymbol{\Sigma}_g^2)}}{n_g} = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_g^2)}{n_g} + o_p(\sigma_g^2). \tag{A. 2}$$

We also note that $|\max(0, \widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}}) - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}| \le |\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}} - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}|$ a.s. From this result and (iv) in Lemma A.3, we get

$$\mathrm{E}\{(\max(0, \widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}}) - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta})^2\} \le \mathrm{var}(\widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}}) = o(\sigma_g^4).$$

Hence,

$$\max(0, \widehat{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}}) = \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + o_p(\sigma_g^2). \tag{A. 3}$$

From (A. 1), $\hat{\mu} = \mu + o_p(\sigma_g)$. From (A. 2) and (A. 3), $\hat{\sigma}_g^2 = \sigma_g^2 + o_p(\sigma_g^2)$. Thus, under (A0)–(A2),

$$\hat{w}_g = w_g + o_p(w_g), \tag{A. 4}$$

where $w_g = -\mu/\sigma_g$ and $\hat{w}_g = -\hat{\mu}/\hat{\sigma}_g$.

We note that $|e(g'|g) - \Phi(\hat{w}_g)| \le |e(g'|g) - \Phi(w_g)| + |\Phi(\hat{w}_g) - \Phi(w_g)|$. From Proposition 2.1, $|e(g'|g) - \Phi(w_g)| = o(1)$. Hence, it is sufficient to show that $|\Phi(\hat{w}_g) - \Phi(w_g)| = o_p(1)$. From (A. 4), we obtain $\hat{w}_g = w_g + o_p(1)$. By the continuous mapping theorem, we then get $|\Phi(\hat{w}_g) - \Phi(w_g)| = o_p(1)$. □

D. Proof of Lemma 2.1. We assume (C1) or (C2). Let $k, k' \in \{1, 2\}$ and $k \ne k'$. Then we decompose $W_g^{(-k)} - \mu$ as $W_{g1}^{(-k)} + W_{g2}^{(-k)}$, where

$$\begin{aligned}
W_{g1}^{(-k)} &= 2\{(-1)^{g+1}\boldsymbol{\delta} - \bar{\mathbf{y}}_{g'} + (n_g - 2)/(n_g - 1)\bar{\mathbf{y}}_{g(-1,-2)}\}^\top \mathbf{y}_{gk}, \\
W_{g2}^{(-k)} &= 2(n_g - 1)^{-1}\mathbf{y}_{g1}^\top \mathbf{y}_{g2} - 2(n_g - 1)^{-1}\bar{\mathbf{y}}_{g(-1,-2)}^\top \mathbf{y}_{gk'} + 2(-1)^g \boldsymbol{\delta}^\top \bar{\mathbf{y}}_{g'} \\
&\quad + \{n_{g'}(n_{g'} - 1)\}^{-1} \sum_{j_1, j_2 = 1, j_1 \ne j_2}^{n_{g'}} \mathbf{y}_{g'j_1}^\top \mathbf{y}_{g'j_2} \\
&\quad - \{(n_g - 1)(n_g - 2)\}^{-1} \sum_{j_1, j_2 = 1, j_1 \ne j_2, j_1, j_2 \ne 1, 2}^{n_g} \mathbf{y}_{gj_1}^\top \mathbf{y}_{gj_2}.
\end{aligned}$$

Then it holds that $W_g^{(-k)} - \mu = W_{g1}^{(-k)} + o_p(\sigma_g)$ under (A0)–(A2).

For non-random constants c_1 and c_2, we define $T = c_1 W_{g1}^{(-1)} + c_2 W_{g1}^{(-2)}$. Then

$$\{c_1(W_g^{(-1)} - \mu) + c_2(W_g^{(-2)} - \mu)\}/\sigma_g = T/\sigma_g + o_p(1).$$

The asymptotic normality of T would imply Lemma 2.1.

Especially, under (C1) and (A0)–(A2), $T/\sigma_g = \sum_{j=1}^{n_g + n_{g'} - 2} \varepsilon_j + o_p(1)$, where

$$\varepsilon_j = \begin{cases} \dfrac{2\mathbf{y}_{gj+2}^\top (c_1 \mathbf{y}_{g1} + c_2 \mathbf{y}_{g2})}{\delta_g (n_g - 1)} & \forall j \in \{1, 2, \ldots, n_g - 2\}, \\ -\dfrac{2\mathbf{y}_{g'j-n_g+2}^\top (c_1 \mathbf{y}_{g1} + c_2 \mathbf{y}_{g2})}{\delta_g n_{g'}} & \forall j \in \{n_g - 1, n_g, \ldots, n_g + n_{g'} - 2\}. \end{cases}$$

Define

$$\begin{aligned}
\mathscr{F}_j &= \sigma\{\mathbf{y}_{g1}, \ldots, \mathbf{y}_{gj+2}\} \qquad (0 \le j \le n_g - 2), \\
\mathscr{F}_j &= \sigma\{\mathbf{y}_{g1}, \ldots, \mathbf{y}_{gn_g}, \mathbf{y}_{g'1}, \ldots, \mathbf{y}_{g'j-n_g+2}\} \qquad (n_g - 1 \le j).
\end{aligned}$$

Then it is straightforward to show that $\mathrm{E}(\varepsilon_j) = 0$ and $\mathrm{E}(\varepsilon_j|\mathscr{F}_{j-1}) = 0$. To apply the martingale central-limit theorem, we need to show that

$$\sum_{j=1}^{n_g+n_{g'}-2} \mathrm{E}(\varepsilon_j^2|\mathscr{F}_{j-1}) = c_1^2 + c_2^2 + o_p(1), \qquad \sum_{j=1}^{n_g+n_{g'}-2} \mathrm{E}(\varepsilon_j^4) = o(1). \quad \text{(A. 5)}$$

First, we check the first part in (A. 5). Note that

$$\sum_{j=1}^{n_g+n_{g'}-2} \mathrm{E}(\varepsilon_j^2|\mathscr{F}_{j-1}) - (c_1^2 + c_2^2) = \delta_g^{-2}(V_1 + V_2) + o_p(1),$$

where

$$V_1 = \frac{4(n_g-2)}{(n_g-1)^2}\{(c_1\mathbf{y}_{g1} + c_2\mathbf{y}_{g2})^\top \boldsymbol{\Sigma}_g(c_1\mathbf{y}_{g1} + c_2\mathbf{y}_{g2}) - (c_1^2 + c_2^2)\operatorname{tr}(\boldsymbol{\Sigma}_g^2)\},$$

$$V_2 = \frac{4}{n_{g'}}\{(c_1\mathbf{y}_{g1} + c_2\mathbf{y}_{g2})^\top \boldsymbol{\Sigma}_{g'}(c_1\mathbf{y}_{g1} + c_2\mathbf{y}_{g2}) - (c_1^2 + c_2^2)\operatorname{tr}(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})\}.$$

Under (A1),

$$\operatorname{var}(V_1) = O(n_g^{-2}\operatorname{tr}(\boldsymbol{\Sigma}_g^4)) = o(\delta_g^4), \qquad \operatorname{var}(V_2) = O(n_{g'}^{-2}\operatorname{tr}\{(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_{g'})^2\}) = o(\delta_g^4).$$

Thus, the first part of (A. 5) holds.

Next, we show the second part of (A. 5). Note that, under (A0),

$$\mathrm{E}(\varepsilon_j^4) = \begin{cases} O(n_g^{-2}) & \forall j \in \{1, 2, \ldots, n_g - 2\} \\ O(n_{g'}^{-2}) & \forall j \in \{n_g - 1, n_g, \ldots, n_g + n_{g'} - 2\}. \end{cases}$$

Thus the second part of (A. 5) holds. From these results, the proof is complete. $\square$

E. Proof of Lemma 2.2. Under (C2) and (A0)–(A2), the random variable T in section D can be factorized as $T/\sigma_g = \sum_{i=1}^p \xi_i$, where

$$\begin{aligned}\xi_i = {} & 2\sigma_g^{-1}c_1\{(-1)^{g+1}\boldsymbol{\delta} - \bar{\mathbf{y}}_{g'} + (n_g-2)/(n_g-1)\tilde{\mathbf{y}}_g\}^\top \boldsymbol{\Sigma}_g^{1/2}\boldsymbol{e}_i z_{gi1} \\ & + 2\sigma_g^{-1}c_2\{(-1)^{g+1}\boldsymbol{\delta} - \bar{\mathbf{y}}_{g'} + (n_g-2)/(n_g-1)\tilde{\mathbf{y}}_g\}^\top \boldsymbol{\Sigma}_g^{1/2}\boldsymbol{e}_i z_{gi2},\end{aligned}$$

here $z_{gi1} = \boldsymbol{e}_i^\top \mathbf{z}_{g1}$, $z_{gi2} = \boldsymbol{e}_i^\top \mathbf{z}_{g2}$, and $\tilde{\mathbf{y}}_g = \bar{\mathbf{y}}_{g(-1,-2)}$. The asymptotic normality of T would imply Lemma 2.2. Define

$$\mathscr{F}_0 = \sigma\{\bar{\mathbf{y}}_{g'}, \tilde{\mathbf{y}}_g\},$$

$$\mathscr{F}_{i-1} = \sigma\{\bar{\mathbf{y}}_{g'}, \tilde{\mathbf{y}}_g, z_{g11}, \ldots, z_{gi-11}, z_{g12}, \ldots, z_{gi-12}\} \qquad (2 \le i).$$

Thus, ξ_i is a martingale difference sequence. To apply the martingale central-limit theorem, we need to show that

$$\sum_{i=1}^{p} \mathrm{E}(\xi_i^2 | \mathscr{F}_{i-1}) = c_1^2 + c_2^2 + o_p(1), \qquad \sum_{i=1}^{p} \mathrm{E}(\xi_i^4) = o(1). \tag{A. 6}$$

To this end, we show the first part of (A. 6). Note that

$$\sum_{i=1}^{p} \mathrm{E}(\xi_i^2 | \mathscr{F}_{i-1}) - (c_1^2 + c_2^2) = 4(c_1^2 + c_2^2)\{2(-1)^{g+1} P_1 + P_2\}/\sigma_g^2 + o_p(1),$$

where

$$\begin{aligned} P_1 &= \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \{(n_g - 2)/(n_g - 1)\tilde{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'}\}, \\ P_2 &= \{(n_g - 2)/(n_g - 1)\tilde{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'}\}^\top \boldsymbol{\Sigma}_g \{(n_g - 2)/(n_g - 1)\tilde{\mathbf{y}}_g - \bar{\mathbf{y}}_{g'}\} \\ &\quad - \{n_g^{-1} \operatorname{tr}(\boldsymbol{\Sigma}_g^2) + n_{g'}^{-1} \operatorname{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)\}. \end{aligned}$$

These variances are evaluated as

$$\begin{aligned} \operatorname{var}(P_1) &= O\left(n_g^{-1} \sqrt{\operatorname{tr}(\boldsymbol{\Sigma}_g^4)} \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta} + n_{g'}^{-1} \sqrt{\operatorname{tr}\{(\boldsymbol{\Sigma}_g \boldsymbol{\Sigma}_{g'})^2\}} \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g \boldsymbol{\delta}\right) = o(\sigma_g^4), \\ \operatorname{var}(P_2) &= O(n_g^{-2} \operatorname{tr}(\boldsymbol{\Sigma}_g^4) + n_{g'}^{-2} \operatorname{tr}\{(\boldsymbol{\Sigma}_g \boldsymbol{\Sigma}_{g'})^2\}) = o(\sigma_g^4). \end{aligned}$$

Thus, under (A1), the first part of (A. 6) holds.

We decompose ξ_i into the sum of three parts, $\xi_i = 2\{(-1)^{g+1}\xi_{1i} + (n_g - 2)/(n_g - 1)\xi_{i2} - \xi_{i3}\}/\sigma_g$, where

$$\begin{aligned} \xi_{i1} &= \boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{e}_i (c_1 z_{gi1} + c_2 z_{gi2}), \qquad \xi_{i2} = \tilde{\mathbf{y}}_g^\top \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{e}_i (c_1 z_{gi1} + c_2 z_{gi2}), \\ \xi_{i3} &= \bar{\mathbf{y}}_{g'}^\top \boldsymbol{\Sigma}_g^{1/2} \boldsymbol{e}_i (c_1 z_{gi1} + c_2 z_{gi2}). \end{aligned}$$

Then, we need to show that $\sum_{i=1}^{p} \mathrm{E}(\xi_{i\ell}^4) = o(\sigma_g^4)$ for $\ell \in \{1, 2, 3\}$. These expectations are given as follows:

$$\begin{aligned} &\sum_{i=1}^{p} \mathrm{E}(\xi_{i1}^4) = O(\operatorname{tr}\{(\boldsymbol{\Sigma}_g^{1/2} \boldsymbol{\delta}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g^{1/2}) \odot (\boldsymbol{\Sigma}_g^{1/2} \boldsymbol{\delta}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_g^{1/2})\}), \\ &\sum_{i=1}^{p} \mathrm{E}(\xi_{i2}^4) = O(n_g^{-2} \operatorname{tr}(\boldsymbol{\Sigma}_g^4)), \qquad \sum_{i=1}^{p} \mathrm{E}(\xi_{i3}^4) = O(n_{g'}^{-2} \operatorname{tr}\{(\boldsymbol{\Sigma}_g \boldsymbol{\Sigma}_{g'})^2\}). \end{aligned}$$

Thus, under (A4), $\sum_{i=1}^{p} \mathrm{E}(\xi_{i1}^4) = o(\sigma_g^4)$. Also, under (A1), $\sum_{i=1}^{p} \mathrm{E}(\xi_{i2}^4) = o(\sigma_g^4)$ and $\sum_{i=1}^{p} \mathrm{E}(\xi_{i2}^4) = o(\sigma_g^4)$. This proves the second part of (A. 6). From these results, the proof is complete. $\square$

Acknowledgement

The author is also grateful to the Editor-in-Chief, an Associate Editor, and reviewer for many valuable comments and helpful suggestions which led to an improved version of this paper. The second author's research was supported in part by a Grant-in-Aid for Young Scientists (B) (17K14238) from the Japan Society for the Promotion of Science.

References

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

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

[3] Y.-B. Chan and P. Hall, Scale adjustments for classifiers in high-dimensional, low sample size settings. Biometrika. **96** (2009), 175–186.

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

[5] P. Hall and C. C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York 1980.

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

[7] A. N. Shiryaev, Probability, Springer-Verlag, Second Ed. 1984.

Hiroki Watanabe
Department of Applied Mathematics
Tokyo University of Science
Tokyo 162-8601 *Japan*
E-mail: watahiro919@gmail.com

Masashi Hyodo
Department of Mathematical Sciences
College of Engineering
Osaka prefecture university
Osaka 599-8531 *Japan*
E-mail: hyodo@ms.osakafu-u.ac.jp

Yuki Yamada
NEC Solution Innovators, Ltd.
Tokyo 136-8627 *Japan*
E-mail: y.y.rfpfg.t.o.d.2006.s.m.e@gmail.com

Takashi Seo
Department of Applied Mathematics
Tokyo University of Science
Tokyo 162-8601 *Japan*
E-mail: seo@rs.tus.ac.jp