Hiroshima Math. J.
48 (2018), 373–411

Estimating the probabilities of misclassification using CV when the dimension and the sample sizes are large

Tomoyuki Nakagawa

(Received November 30, 2017)
(Revised September 24, 2018)

Abstract. In this paper, we study about estimating the probabilities of misclassification in the high-dimensional data. In many cases, the cross-validation (CV) is often used for estimations of the probabilities of misclassification. CV provides a nearly unbiased estimate, using the original data when the sample sizes are large. On the other hand, the properties of CV are not well-known when the dimension is large as compared to the sample sizes. Therefore, we investigate asymptotic properties of CV when the dimension and the sample sizes tend to be large. Furthermore, we suggest the three methods for correcting the bias by using CV which is usable in the high-dimensional data. We show performances of the estimators in the simulation studies.

1. Introduction

In this paper, we consider estimating the probabilities of misclassification which are expressed by

$$P(2|1) = \Pr(\text{the rule classifies } \boldsymbol{x} \text{ to } \Pi_2 \,|\, \boldsymbol{x} \in \Pi_1),$$

$$P(1|2) = \Pr(\text{the rule classifies } \boldsymbol{x} \text{ to } \Pi_1 \,|\, \boldsymbol{x} \in \Pi_2),$$

for a classification rule constructed from a training data. For $k = 1, 2$, the training data $\boldsymbol{X}_k = (\boldsymbol{x}_{k1}, \ldots, \boldsymbol{x}_{kN_k})^\top$ consists of N_k observations where $\boldsymbol{a}^\top$ is the transpose of $\boldsymbol{a}$, and $\boldsymbol{x}_{ik}$ is ith p-variate feature vector belonging to kth population Π_k. Furthermore, the classification rule using the discriminant function $d(\boldsymbol{x})$, which is constructed from $\boldsymbol{X}_k$ $(k = 1, 2)$, is given by

$$d(\boldsymbol{x}) > c \Rightarrow \boldsymbol{x} \in \Pi_1, \qquad d(\boldsymbol{x}) \le c \Rightarrow \boldsymbol{x} \in \Pi_2,$$

where c is a cut-off point. By using the discriminant function d, the probabilities of misclassification are expressed by

---

2010 *Mathematics Subject Classification*. Primary 62H30; Secondary 62H12.

Key words and phrases. Discriminant analysis; Classification; Probability of Misclassification; Cross-Validation; asymptotic expansion; High-dimensional.

$$P(2|1) = \Pr\{d(\boldsymbol{x}) \le c \,|\, \boldsymbol{x} \in \Pi_1\},$$

$$P(1|2) = \Pr\{d(\boldsymbol{x}) > c \,|\, \boldsymbol{x} \in \Pi_2\}.$$

Various discriminant functions have been proposed in many researches. For example, Fisher's discriminant function is given by

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

where $\bar{\boldsymbol{x}}_k$ is the sample mean of $\boldsymbol{X}_k$ for $(k = 1, 2)$, and $\boldsymbol{S}$ is the pooled sample covariance matrix (Fisher, 1936). Moreover, we also consider the following discriminant function,

$$D_b(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)^\top \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_2) - b(\boldsymbol{x} - \bar{\boldsymbol{x}}_1)^\top \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1),$$

where b is a constant. D_b was introduced in Fujikoshi and Seo (1998) and includes various discriminant functions, for example D_b is the same as d_F when $b = 1$. The two discriminant functions D_b and d_F are often used to the discriminant analysis in the two normal populations.

For observation $\boldsymbol{x}$, the statistician wishes to estimate the probabilities of misclassification for a classification rule, because the probabilities of misclassification are natural risks to measure the goodness of discrimination. If we had the exact evaluation of the probabilities of misclassification for all classifiers, we could select the best classifier and can make an accurate classification. However, in general, it is hard to obtain the exact evaluation of the probabilities of misclassification, therefore it is necessary to estimate the probabilities of misclassification from the observations. Estimation methods of the probabilities of misclassification are separated between the parametric and the non-parametric methods. In the parametric methods, we assume a distribution and a classification rule and derive an approximation formula of the probabilities of misclassification. For example, an approximation formula of the probabilities of misclassification for d_F was given by Okamoto (1963) and Tonda, *et al.* (2017) etc. However, since it is necessary to assume a distribution and a classification rule, the parametric methods can only be applied to restrictively classification. On the other hand, the cross-validation (CV) has been used to estimate the probabilities of misclassification for a long time (see Lachenbruch and Mickey, 1968; Stone, 1974). CV is one of the non-parametric methods and is so useful that the method of CV does not need assumption of a distribution and a classification rule. Furthermore, CV provides a nearly unbiased estimate, using the original data when sample sizes are large (see McLachlan, 1974; Efron, 1997). Recently, the data whose the dimension is large are observed, for example, the image data and the genetic data. However, asymptotic properties of CV are not well-known in the high-dimensional case. Hence, we investigate asymptotic properties of CV when the dimension and the sample sizes are large. Furthermore, it is known that the bias of CV increases with the dimension in the simulation studies. Therefore, we suggest three methods for correcting the bias by using CV which are usable in the high-dimensional data.

This paper is organized as follows: In section 2, we investigate asymptotic properties of CV by using an asymptotic expansion in the high-dimensional case. In section 3, we suggest three methods for correcting the bias by using CV. In section 4, we show performances of the estimators in simulation studies.

2. Asymptotic properties

In this section, we investigate asymptotic properties of CV for estimating the probabilities of misclassification. Most of the asymptotic results of CV are based on the large samples (LS) framework:

$$p \text{ is fixed}, \qquad N_1, N_2 \to \infty, \qquad \frac{N}{N_k} = O(1) \quad (k = 1, 2),$$

where $N = N_1 + N_2$. Regarding to estimate the probabilities of misclassification by using CV, it is well-known that the bias is O_2 based on the LS framework (see McLachlan, 1974), where O_k means a term of the kth order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1}, (N-p)^{-1})$. However, the data whose the dimension is large as compared to the sample sizes has been observed in recently. Therefore, we consider an asymptotic theory based on the high-dimensional (HD) framework:

$$p, N_1, N_2 \to \infty, \qquad \frac{N}{N_k} = O(1) \quad (k = 1, 2), \qquad \frac{p}{N} \to c_0 \in (0, 1),$$

and $N - p - 2 > 0$.

Remark 1. The Mahalanobis distance $\Delta = \{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1)\}^{1/2}$ may tend to infinity depending on p. However, since $P(2|1) \to 0$ with $\Delta \to \infty$, we assume that $\Delta = O(1)$ even when $p \to \infty$ in this paper.

In this section, we assume that Π_k is the normal population with the mean vector $\boldsymbol{\mu}_k$ and the covariance matrix $\boldsymbol{\Sigma}$ for $k = 1, 2$, that is

$$\Pi_1 : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \qquad \Pi_2 : N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}). \tag{1}$$

Firstly, we consider the bias of the estimator using CV. The estimator $\hat{P}_{CV}$ using CV is expressed as

$$\hat{P}_{CV} = N_1^{-1} \sum_{i=1}^{N_1} 1(d^{(-i)}(\boldsymbol{x}_{1i}) \leq c),$$

where $1(\cdot)$ is the indicator function and $d^{(-i)}$ is the discriminant function constructed without $\boldsymbol{x}_{1i}$. Then we have the following theorem.

THEOREM 1. *If the expansion of the probability of misclassification $P(2|1)$ is given by*

$$P(2|1) = Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{1}{N} Q_1\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2, \tag{2}$$

where $Q_0(x_1, x_2)$ and $Q_1(x_1, x_2)$ are C^1 class functions around $(p/N_1, p/N_2)$, then it holds that

$$\mathrm{E}[\hat{P}_{CV}(2|1)] - P(2|1) = O_1.$$

PROOF. From (2), an expectation of the estimator by CV is given by

$$\begin{aligned} \mathrm{E}[\hat{P}_{CV}(2|1)] &= Q_0\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) + \frac{1}{N-1} Q_1\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) + O_2 \\ &= Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{1}{N} Q_1\left(\frac{p}{N_1}, \frac{p}{N_2}\right) \\ &\quad + \frac{p}{N_1(N_1 - 1)} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2. \end{aligned}$$

Since Q_0 is a C^1 class function, $\partial Q_0/\partial x_1$ is the continuous function and $\partial Q_0/\partial x_1(p/N_1, p/N_2)$ is bounded as $N_1, N_2, p \to \infty$. Therefore

$$\mathrm{E}[\hat{P}_{CV}(2|1)] - P(2|1) = \frac{p}{N_1(N_1 - 1)} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2 = O_1.$$

The proof of this theorem does not need to assume normality of the populations. From the proof of this theorem, the estimator by CV is an asymptotic unbiased estimator in the HD framework but the bias in HD framework is a larger order than it in the LS framework. For the classification using d_F, the following theorem was given in Tonda *et al.* (2017).

THEOREM 2. *Let $\boldsymbol{x} \in \Pi_1$, then $P(2|1)$ can be expanded as*

$$P(2|1) = \Phi(v) + \phi(v) F_1(\Delta) + O_2,$$

where $\phi(\cdot)$ is the density function of $N(0,1)$,

$$v = v(\Delta^2) = -\frac{1}{2}\left(\frac{N-p}{N-1}\right)^{1/2}\left\{\Delta^2 + \frac{(N_1-N_2)(p-1)}{N_1N_2}\right\}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}^{-1/2}.$$

Moreover $F_1(\Delta)$ is given as follows:

$$F_1(\Delta) = T_{(2)} - \frac{T_{(0)}}{2}T_{(1)},$$

where $m = N_1N_2/N$ and

$$q_1 = \frac{(N-1)m^2\Delta^2(p-1+m\Delta^2)}{N(N-p-1)^3(N-p)}, \qquad q_2 = \frac{m(N-1)(N_1-N_2)(p-1+m\Delta^2)^2}{N(N-p-1)^3(N-p)},$$

$$T_{(0)} = q_1 + q_2,$$

$$\begin{aligned}
T_{(1)} = {} & \frac{T_{(0)}^2}{4}\left(\frac{2(p-1)+4m\Delta^2}{(p-1+m\Delta^2)^2} + \frac{8}{N-p-1} + \frac{2(p-1)}{(N-p)(N-1)}\right) \\
& + q_1^2\left\{\frac{1}{m\Delta^2}\left(1 + \frac{(p-1)^2}{(N-p)(p-2)}\right) + \frac{2}{N-p-1}\right\} \\
& + q_2^2\left(\frac{2(p-1)+4m\Delta^2}{(p-1+m\Delta^2)^2} + \frac{2}{N-p-1}\right) + \frac{1}{N} \\
& - q_1T_{(0)}\left(\frac{2}{p-1+m\Delta^2} + \frac{4}{N-p-1}\right) \\
& - q_2T_{(0)}\left(\frac{2(p-1)+4m\Delta^2}{(p-1+m\Delta^2)^2} + \frac{4}{N-p-1}\right) \\
& + 2q_1q_2\left(\frac{2}{p-1+m\Delta^2} + \frac{2}{N-p-1}\right),
\end{aligned}$$

$$\begin{aligned}
T_{(2)} = {} & \frac{T_{(0)}}{8}\left(\frac{2(p-1)+8m\Delta^2}{(p-1+m\Delta^2)^2} - \frac{2(p-1)}{(N-p)(N-1)}\right) \\
& + q_1\frac{1}{p-1+m\Delta^2} - q_2\frac{m\Delta^2}{(p-1+m\Delta^2)^2}.
\end{aligned}$$

Therefore, the classification using d_F satisfies assumption (2). From Theorem 1, we obtain the following corollary.

Corollary 1. *In the case of classification using d_F, the bias of CV has order O_1.*

Secondly, we consider evaluating the mean squared error (MSE) of $\hat{P}_{CV}(2|1)$. The straightforward calculations give

$$\mathrm{MSE}(\hat{P}_{CV}(2|1)) = \mathrm{Bias}(\hat{P}_{CV}(2|1))^2 + \mathrm{Var}(\hat{P}_{CV}(2|1)),$$

$$\begin{aligned}\mathrm{Var}(\hat{P}_{CV}(2|1)) \\ &= \Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c, d^{(-2)}(\boldsymbol{x}_{12}) \le c) - \Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c)^2 \\ &\quad + \frac{1}{N_1}[\Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c) - \Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c, d^{(-2)}(\boldsymbol{x}_{12}) \le c)].\end{aligned}$$

Note that $\hat{P}_{CV}(2|1)$ has consistency if $d^{(-1)}(\boldsymbol{x}_{11})$ and $d^{(-2)}(\boldsymbol{x}_{12})$ are asymptotically independent, that is

$$\Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c, d^{(-2)}(\boldsymbol{x}_{12}) \le c) - \Pr(d^{(-1)}(\boldsymbol{x}_{11}) \le c)^2 \to 0, \tag{3}$$

as $N_1, N_2, p \to \infty$.

Example 1. *In the case of the LS framework, the classification rule using the discriminant function D_b clearly satisfies condition* (3) *from Slutsky's theorem.*

Hereafter, we show that MSE of CV for the discriminant function D_b in the HD framework.

Lemma 1. *Let $\boldsymbol{x} \in \Pi_1$, then $D_b(\boldsymbol{x})$ is expressed as*

$$D_b(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{U}) \tag{4}$$

where $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top$, $\boldsymbol{V}_1 \sim W_3(N-p, \boldsymbol{I}_3)$, $\boldsymbol{V}_2 = \boldsymbol{T}\boldsymbol{T}^\top \sim W_3(p, \boldsymbol{I}_3, \boldsymbol{\Omega})$, and $\boldsymbol{V}_1$ and $\boldsymbol{T}$ are independent, and

$$\boldsymbol{A} = \begin{pmatrix} n/N_2 & 0 & -n/\sqrt{N_2} \\ 0 & -nb/N_1 & nb/\sqrt{N_1} \\ -n/\sqrt{N_2} & nb/\sqrt{N_1} & n(1-b) \end{pmatrix},$$

$n = N - 2$.

The proof is given in the appendix.

Theorem 3. *Let $\boldsymbol{x} \in \Pi_1$ and $b = 1 + O_1$, then $P(2|1)$ is expanded as follows:*

$$P(2|1) = \Phi(v) + O_1, \tag{5}$$

where

$$v = s^{-1}(c - \eta).$$

Here η and s are given in the appendix.

The proof was given in Fujikoshi and Seo (1998) and Fujikoshi (2000). On the other hand, we can show the different way of the proof in the appendix. This theorem means that the estimator of the probabilities of misclassification using CV is an asymptotic unbiased estimator in the case of classification using D_b and the order of its bias is O_1 in the HD framework.

LEMMA 2. *The sample mean and the sample covariance matrix of Π_1 are expressed as follows:*

$$\bar{\boldsymbol{x}}_k = \frac{n_1}{N_1}\bar{\boldsymbol{x}}_k^{(-i)} + \frac{1}{N_1}\boldsymbol{x}_{ki},$$

$$n_1\boldsymbol{S}_1 = (n_1 - 1)\boldsymbol{S}_1^{(-i)} + \frac{n_1 - 1}{n_1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})^\top,$$

for $k = 1, 2$. Moreover,

$$n\boldsymbol{S} = (n - 1)\boldsymbol{S}^{(-i)} + \frac{n_1 - 1}{n_1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(i)})^\top,$$

$$\boldsymbol{S}^{-1} = \frac{n}{n-1}[\{\boldsymbol{S}^{(-i)}\}^{-1} - T^{-1}\{\boldsymbol{S}^{(-i)}\}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})^\top\{\boldsymbol{S}^{(-i)}\}^{-1}],$$

$$T = \frac{N_1(n-1)}{n_1} + (\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)})^\top\{\boldsymbol{S}^{(-i)}\}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)}),$$

where $n_1 = N_1 - 1$, $n_2 = N_2 - 1$, and $\bar{\boldsymbol{x}}_k^{(-i)}$, $\boldsymbol{S}_k^{(-i)}$ and $\boldsymbol{S}^{(-i)}$ are the sample mean, the sample covariance matrix and the pool covariance matrix without $\boldsymbol{x}_{ki}$, for example, $\bar{\boldsymbol{x}}_k^{(-i)}$, $\boldsymbol{S}_k^{(-i)}$ and $\boldsymbol{S}^{(-i)}$ for $k = 1$ are expressed as

$$\bar{\boldsymbol{x}}_1^{(-i)} = n_1^{-1}\sum_{j \neq i}^{N_1} x_{1j},$$

$$\boldsymbol{S}_1^{(-i)} = (N_1 - 2)^{-1}\sum_{j \neq i}^{N_1}(\boldsymbol{x}_{1j} - \bar{\boldsymbol{x}}_1^{(-j)})(\boldsymbol{x}_{1j} - \bar{\boldsymbol{x}}_1^{(-j)})^\top,$$

$$\boldsymbol{S}^{(-i)} = (N_1 - 2)\boldsymbol{S}_1^{(-i)} + n_2\boldsymbol{S}_2.$$

It is easy to prove Lemma 2, so we omit the proof. Suppose that $D_b^{(-i)}$ is D_b constructed without $\boldsymbol{x}_{1i}$. From Lemma 2, we have the following lemma.

LEMMA 3. *$D_b^{(-i)}(\boldsymbol{x}_{1i})$ and $D_b^{(-j)}(\boldsymbol{x}_{1j})$ are expressed by*

$$D_b^{(-i)}(\boldsymbol{x}_{1i}) = \mathrm{tr}(\boldsymbol{A}_1\boldsymbol{U}) - T_1^{-1}\boldsymbol{a}_1^\top \boldsymbol{U}\boldsymbol{A}_1\boldsymbol{U}\boldsymbol{a}_1, \tag{6}$$

$$T_1 = \frac{N_1 - 1}{N_1 - 2} + \mathrm{tr}(\boldsymbol{B}_1\boldsymbol{U}),$$

$$D_b^{(-j)}(\boldsymbol{x}_{1j}) = \mathrm{tr}(\boldsymbol{A}_2\boldsymbol{U}) - T_2^{-1}\boldsymbol{a}_2^\top \boldsymbol{U}\boldsymbol{A}_2\boldsymbol{U}\boldsymbol{a}_2, \tag{7}$$

$$T_2 = \frac{N_1 - 1}{N_1 - 2} + \mathrm{tr}(\boldsymbol{B}_2\boldsymbol{U}),$$

where $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top$, $\boldsymbol{V}_1 \sim W_4(N-p, \boldsymbol{I}_4)$, $\boldsymbol{V}_2 = \boldsymbol{T}\boldsymbol{T}^\top \sim W_4(p, \boldsymbol{I}_4, \boldsymbol{\Omega})$, and $\boldsymbol{V}_1$ and $\boldsymbol{T}$ are independent, and $\boldsymbol{a}_1 = (0, -n_1^{-1/2}, 0, 1)^\top$, $\boldsymbol{a}_2 = (0, -n_1^{-1/2}, 1, 0)^\top$, $\boldsymbol{B}_i = \boldsymbol{a}_i\boldsymbol{a}_i^\top$ $(i = 1, 2)$,

$$\boldsymbol{A}_1 = (n-1)\begin{pmatrix} N_2^{-1} & 0 & -N_2^{-1/2} & 0 \\ 0 & -b(n_1-1)n_1^{-2} & b(n_1-1)^{1/2}n_1^{-1} & b(n_1-1)^{1/2}n_1^{-2} \\ -N_2^{-1/2} & b(n_1-1)^{1/2}n_1^{-1} & 1-b & bn_1^{-1} \\ 0 & b(n_1-1)^{1/2}n_1^{-2} & bn_1^{-1} & -bn_1^{-2} \end{pmatrix},$$

$$\boldsymbol{A}_2 = (n-1)\begin{pmatrix} N_2^{-1} & 0 & 0 & -N_2^{-1/2} \\ 0 & -b(n_1-1)n_1^{-2} & b(n_1-1)^{1/2}n_1^{-2} & b(n_1-1)^{1/2}n_1^{-1} \\ 0 & b(n_1-1)^{1/2}n_1^{-2} & -bn_1^{-2} & bn_1^{-1} \\ -N_2^{-1/2} & b(n_1-1)^{1/2}n_1^{-1} & bn_1^{-1} & 1-b \end{pmatrix}.$$

The proof is given in the appendix. By using this lemma, we obtain the following theorem.

THEOREM 4. *Let $b = 1 + O_1$ then*

$$\Pr(D_b^{(-1)}(\boldsymbol{x}_{11}) \le c, D_b^{(-2)}(\boldsymbol{x}_{12}) \le c) - \Pr(D_b^{(-1)}(\boldsymbol{x}_{11}) \le c)^2 = O_1.$$

Therefore, it holds that

$$\mathrm{MSE}(\hat{P}_{CV}(2|1)) = O_1.$$

PROOF. The characteristic function $\phi(\boldsymbol{t}) = \phi(t_1, t_2)$ of the joint distribution of $D_b^{(-1)}(\boldsymbol{x}_{11})$ and $D_b^{(-2)}(\boldsymbol{x}_{12})$ is expanded as

$$\phi(\boldsymbol{t}) = \exp\left\{it_1\eta_1 - \frac{t^2}{2}\lambda_{11}\right\}\exp\left\{it_1\eta_2 - \frac{t^2}{2}\lambda_{22}\right\} + O_1.$$

By inverting the characteristic function, we obtain the following formula.

$$\Pr(\lambda_{11}^{-1/2}(D_b^{(-1)}(\boldsymbol{x}_{11}) - \eta_1) \le x_1, \lambda_{22}^{-1/2}(D_b^{(-2)}(\boldsymbol{x}_{12}) - \eta_2) \le x_2) = \Phi(x_1)\Phi(x_2) + O_1.$$

From this formula and Theorem 3,

$$\begin{aligned}
\Pr(D_b^{(-1)}(\boldsymbol{x}_{11}) \le c, D_b^{(-2)}(\boldsymbol{x}_{12}) \le c) \\
&= \Phi(\lambda_{11}^{-1/2}(c - \eta_1))\Phi(\lambda_{22}^{-1/2}(c - \eta_2)) + O_1 \\
&= \Phi(\lambda_{11}^{-1/2}(c - \eta_1))^2 + O_1, \\
\Pr(D_b^{(-1)}(\boldsymbol{x}_{11}) \le c)^2 = \Phi(s^{-1}(c - \eta))^2 + O_1.
\end{aligned}$$

Therefore, we complete the proof of this theorem.

From this theorem, the estimator of CV has consistency to $P(2|1)$ in the HD framework. On the other hand, the following theorem was given in Tonda *et al.* (2017).

THEOREM 5. *MSE of the proposed estimator tends to* 0 *as* O_1 *order in the normal populations.*

Theorems 4 and 5 mean that MSE of CV is the same order as that of the estimator in Tonda *et al.* (2017) and the two estimators have consistency to $P(2|1)$.

3. Correcting the bias of CV

In this section, we suggest three methods for correcting the bias of the estimator using CV. In the previous section, we showed that if the sample sizes are sufficiently large, CV is a good estimation method even if the dimension is large. However, the bias of estimator using CV is large for the small sample sizes and increases with the dimension. Therefore, it is necessary to correct the bias of estimator using CV in the HD framework.

3.1. Method I: Using the leave-two-out CV. The method I is one of the non-parametric methods for correcting the bias of the information criterion proposed by Yanagihara and Fujisawa (2012). In this section, we use this idea to estimate the probabilities of misclassification. The leave-two-out CV is expressed by

$$\hat{P}_{CV_2}(2|1) = \frac{1}{{}_{N_1}C_2} \sum_{i<j}^{N_1} \frac{1}{2} \sum_{k \in \{i,j\}} 1(d^{(-i,-j)}(\boldsymbol{x}_{1k}) \le c),$$

where $N_j^{(-\ell)} = N_j - \ell$, $N^{(-\ell)} = N - \ell$ and $d^{(-i,-j)}$ is the discriminant function constructed without $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{1j}$. Then

$$\mathrm{E}[\hat{P}_{CV}(2|1)] = Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{1}{N} Q_1\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{p}{N_1 N_1^{(-1)}} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2,$$

$$\mathrm{E}[\hat{P}_{CV_2}(2|1)] = Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{1}{N} Q_1\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + \frac{2p}{N_1 N_1^{(-2)}} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2,$$

$$\mathrm{E}[\hat{P}_{CV}(2|1)] - P(2|1) = \frac{p}{N_1 N_1^{(-1)}} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2,$$

$$\mathrm{E}[\hat{P}_{CV_2}(2|1) - \hat{P}_{CV}(2|1)] = \frac{p}{N_1^{(-1)} N_1^{(-2)}} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) + O_2.$$

Therefore, a new estimator is given by

$$\hat{P}_{\mathrm{I}}(2|1) = \left\{\hat{P}_{CV}(2|1) - \frac{N_1^{(-2)}}{N_1}(\hat{P}_{CV_2}(2|1) - \hat{P}_{CV}(2|1))\right\}.$$

Then it holds that

$$\mathrm{E}[\hat{P}_{\mathrm{I}}(2|1)] - P(2|1) = O_2.$$

Hence, we can correct the bias of CV by using the leave-two-out CV in the HD framework. Furthermore, the similar method for correcting the bias can be done by using the two estimators of leave-k-out CV of different k.

3.2. Method II: Leave-λ-out CV. We consider leaving out λ instead of one from a training data by CV method. This idea was proposed by Yanagihara *et al.* (2006) and Yanagihara *et al.* (2013) for correcting the bias of the information criterion. In this section, we use this idea to estimate the probabilities of misclassification. Suppose that $F_{N-1}^{(-i)}$ and F_i are the empirical distributions of $\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{1i-1}, \boldsymbol{x}_{1i+1}, \dots, \boldsymbol{x}_{1N_1}$ and $\boldsymbol{x}_{1i}$, respectively. The discriminant function $\hat{d}^{(-i;\lambda)}$ is constructed by using $(1-u_\lambda)F_{N-1}^{(-i)} + u_\lambda F_i$, where $u_\lambda = (1-\lambda)/(N_1-\lambda)$. For example, assuming the discriminant function d_θ is parameterized, the maximum likelihood estimator of parameter θ is given as follows:

$$\hat{\theta}^{(-i;\lambda)} = \arg\max_{\theta \in \Theta}\left\{\frac{1}{N_1-\lambda}\sum_{k \neq i}^{N_1} \log f(\boldsymbol{x}_{1k};\theta) + \frac{1-\lambda}{N_1-\lambda}\log f(\boldsymbol{x}_{1i};\theta)\right\},$$

where f is a probability density function of $\boldsymbol{x}_{1i}$. Then $\hat{d}^{(-i;\lambda)}$ is the same as $d_{\hat{\theta}^{(-i;\lambda)}}$. In the normal case, the estimators of mean $\bar{\boldsymbol{x}}_1^{(-i;\lambda)}$ and covariance matrix $\boldsymbol{S}^{(-i;\lambda)}$ are given by

$$\bar{\boldsymbol{x}}_1^{(-i;\lambda)} = \frac{N_1 - 1}{N_1 - \lambda}\bar{\boldsymbol{x}}_1^{(-i)} + \frac{1-\lambda}{N_1 - \lambda}\boldsymbol{x}_{1i},$$

$$\boldsymbol{S}^{(-i;\lambda)} = \frac{1}{N^{(-\lambda)}}\left\{(N^{(-3)})\boldsymbol{S}^{(-i)} + \frac{N_1^{(-1)}}{N_1^{(-\lambda)}}(1-\lambda)(\boldsymbol{x}_{1i} - \bar{\boldsymbol{x}}_1^{(-i)})(\boldsymbol{x}_{1i} - \bar{\boldsymbol{x}}_1^{(-i)})^\top\right\}. \quad (8)$$

In the case $\lambda = 1$, this method is the same as usually CV (leave-one-out CV). We define by using $\hat{d}^{(i;\lambda)}$ as

$$\hat{P}_{CV_\lambda}(2|1) = \frac{1}{N_1}\sum_{i=1}^{N_1} 1(\hat{d}^{(-i;\lambda)}(\boldsymbol{x}_{1i}) \le c).$$

This method is called leave-λ-out CV in this paper. Suppose that the expansion of $\Pr(\hat{d}^{(-1;\lambda)}(\boldsymbol{x}_{11}))$ is given by

$$\Pr(\hat{d}^{(-1;\lambda)}(\boldsymbol{x}_{11}) \le c) = Q_0^*\left(\frac{p}{N_1}, \frac{p}{N_2}, \lambda\right) + \frac{1}{N}Q_1^*\left(\frac{p}{N_1}, \frac{p}{N_2}, \lambda\right) + O_2,$$

where $Q_0^*(x_1, x_2, x_3)$ and $Q_1^*(x_1, x_2, x_3)$ are C^1 class functions around $(p/N_1, p/N_2, 1)$. Let $\lambda = 1 - \kappa/N$, then we obtain the following expansion:

$$\begin{aligned}\mathrm{E}[\hat{P}_{CV_\lambda}(2|1)] &= Q_0^*\left(\frac{p}{N_1}, \frac{p}{N_2}, \lambda\right) + \frac{1}{N}Q_1^*\left(\frac{p}{N_1}, \frac{p}{N_2}, \lambda\right) + O_2 \\ &= Q_0\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) - \frac{\kappa}{N}\frac{\partial}{\partial x_3}Q_0^*\left(\frac{p}{N_1}, \frac{p}{N_2}, 1\right) \\ &\quad + \frac{1}{N}Q_1\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) + O_2.\end{aligned}$$

Therefore, the bias of leave-λ-out CV is given by

$$\begin{aligned}&\mathrm{E}[\hat{P}_{CV_\lambda}(2|1)] - P(2|1) \\ &\quad = \frac{p}{N_1(N_1 - 1)}\frac{\partial}{\partial x_1}Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) - \kappa\frac{1}{N}\frac{\partial}{\partial x_3}Q_0^*\left(\frac{p}{N_1}, \frac{p}{N_2}, 1\right) + O_2.\end{aligned}$$

Thus, we can correct a bias by deciding κ so that the term of O_1 is 0, that is, κ is decided as follows:

$$\hat{\kappa} = \frac{pN}{N_1(n_1)}\frac{\partial}{\partial x_1}Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right)\Big/\frac{\partial}{\partial x_3}Q_0^*\left(\frac{p}{N_1}, \frac{p}{N_2}, 1\right).$$

Example 2. *In the case of d_F and $c = 0$, λ is decided as follows:*

$$\lambda = 1 - \kappa(\Delta)/N,$$

$$\kappa(\Delta) = \frac{N}{4N_1}\left\{2 - \left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\right\}. \tag{9}$$

A derivation of this κ is given in the appendix.

This method has the same computational complexity as CV and can correct the bias of CV, while we must derive λ.

3.3. Method III: Modified a cutoff point. We propose a method for correcting the bias by modifying a cut-off point c. We define $\hat{P}_{CV_c}(2|1)$ as

$$\hat{P}_{CV_c}(2|1) = \frac{1}{N_1}\sum_{j=1}^{N_1} 1\left(d^{(-j)}(\boldsymbol{x}_{1j}) \le c + \frac{c_1}{N}\right).$$

Suppose that the expansion of $P(2|1)$ is given by

$$\begin{aligned} P(2|1) &= \Pr(d(\boldsymbol{x}) \le c | x \in \Pi_1) \\ &= Q_0^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c\right) + \frac{1}{N} Q_1^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c\right) + O_2. \end{aligned}$$

Then, we obtain the following expansion:

$$\begin{aligned} \mathrm{E}[\hat{P}_{CV_c}(2|1)] &= Q_0^\dagger\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}, c + \frac{c_1}{N}\right) \\ &\quad + \frac{1}{N-1} Q_1^\dagger\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}, c + \frac{c_1}{N}\right) + O_2 \\ &= Q_0\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) + \frac{1}{N} Q_1\left(\frac{p}{N_1 - 1}, \frac{p}{N_2}\right) \\ &\quad + \frac{c_1}{N}\frac{\partial}{\partial x_3} Q_0^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c\right) + O_2. \end{aligned}$$

Therefore, the bias of $\hat{P}_{CV_c}(2|1)$ is given by

$$\begin{aligned} &\mathrm{E}[\hat{P}_{CV_c}(2|1)] - P(2|1) \\ &\quad = \frac{p}{N_1(N_1 - 1)}\frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) - c_1 \frac{1}{N}\frac{\partial}{\partial x_3} Q_0^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c\right) + O_2. \end{aligned}$$

Thus, we can correct the bias by deriving c_1 so that the term of O_1 is 0, that is, c_1 is derived as follows:

$$\hat{c}_1 = \frac{pN}{N_1(n_1)} \frac{\partial}{\partial x_1} Q_0\left(\frac{p}{N_1}, \frac{p}{N_2}\right) \Big/ \frac{\partial}{\partial x_3} Q_0^{\dagger}\left(\frac{p}{N_1}, \frac{p}{N_2}, c\right).$$

Example 3. *In the case of the classification rule using D_b, we can obtain c_1 as follows;*

$$\eta^{(-1)} = \frac{n-1}{N-p-1}\left(\varDelta^2 + \frac{p}{N_2} - \frac{bp}{n_1} + p(1-b)\right) = \eta + \eta_1 + O_2,$$

$$(s^{(-1)})^2 = 4\frac{(n-1)^2(N-1)}{(N-p-1)^3}\left(\varDelta^2 + \frac{pb^2}{n_1} + \frac{p}{N_2}\right)$$

$$= s^2 + s_1 + O_2,$$

$$c_1(\varDelta) = \frac{N}{\lambda}\left\{\frac{\lambda_1}{2}(c-\eta) - \lambda\eta_1\right\},$$

where η and s^2 are given by Theorem 3, and

$$\eta_1 = \left(\frac{1}{N-p} + \frac{n}{(N-p)^2}\right)\left(\varDelta^2 + \frac{p}{N_2} - \frac{bp}{N_1} + p(1-b)\right) - \frac{bnp}{(N-p)N_1^2},$$

$$s_1 = 4\frac{Nn^2}{(N-p)^3}\left(\frac{3}{N-p} - \frac{2}{n} - \frac{1}{N}\right)\left(\varDelta^2 + \frac{pb^2}{N_1} + \frac{p}{N_2}\right) + 4\frac{pb^2Nn^2}{N_1^2(N-p)^3}.$$

This method has the same computational complexity as CV and can correct the bias of CV, while we must derive c_1.

4. Numerical study

In this section, we investigate performances of CV and the three methods for the classification rule using d_F by the Monte Carlo method. Without loss of generality, we can assume that $\boldsymbol{\mu}_1 = \varDelta(1,\dots,1)'/2\sqrt{p}$, $\boldsymbol{\mu}_2 = -\varDelta(1,\dots,1)'/2\sqrt{p}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively. The configuration of the values of N_1, N_2, p and $\varDelta$ were $N_1, N_2 = 15, 20, 25, 30, 35$, $p/N = 1/5, 3/5$ and $\varDelta = 1.05, 1.68, 2.56, 3.29$ satisfying $N - p - 2 > 0$. The values of $\varDelta$ correspond to the values 0.30, 0.20, 0.10, 0.05 of $\varPhi(-\varDelta/2)$, respectively. An estimator of $\varDelta$ is necessary to use the methods II and III, so that $\hat{\varDelta}^2$ was given by

$$\hat{\varDelta}^2 = \frac{n-p-3}{n}D^2 - \frac{pN}{N_1N_2},$$

Table 1. (Bias of estimators) ×100

p/N	Δ	p	N_1	CV	I	II	III	TNW
1/5	1.05	6	15	0.451	−0.021	0.043	0.015	1.402
		10	25	0.300	−0.021	−0.049	−0.023	0.839
		14	35	0.223	0.028	0.040	0.021	0.604
	1.68	6	15	0.392	0.030	0.083	0.038	1.309
		10	25	0.186	−0.030	0.009	−0.033	0.753
		14	35	0.163	0.004	0.038	0.007	0.556
	2.56	6	15	0.275	0.035	0.101	0.045	0.978
		10	25	0.112	−0.041	0.014	−0.025	0.554
		14	35	0.104	0.001	0.037	0.003	0.404
	3.29	6	15	0.157	−0.015	0.058	0.013	0.662
		10	25	0.072	−0.028	0.019	−0.017	0.393
		14	35	0.075	0.004	0.039	0.009	0.293
3/5	1.05	18	15	0.807	0.043	0.166	0.275	1.086
		30	25	0.516	0.047	0.132	0.126	0.652
		56	35	0.335	0.002	0.067	0.042	0.434
	1.68	18	15	0.912	0.040	0.303	0.282	1.301
		30	25	0.516	−0.024	0.168	0.069	0.758
		56	35	0.396	0.024	0.156	0.061	0.554
	2.56	18	15	0.953	0.002	0.466	0.302	1.355
		30	25	0.583	0.019	0.324	0.137	0.862
		56	35	0.397	0.003	0.219	0.055	0.609
	3.29	18	15	0.910	−0.039	0.539	0.323	1.255
		30	25	0.538	−0.008	0.346	0.120	0.784
		56	35	0.377	−0.004	0.253	0.061	0.544

where $D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. $\hat{\Delta}^2$ is unbiased and a consistent estimator of Δ^2 under both of the approximation frameworks (see Tonda *et al.* (2017)). In the tables, the 1–2 columns indicate the rate of the dimension p and the sample size N and Δ, respectively. The 3–4 columns indicate the dimension p and the sample size N_1, respectively. In table 1, the 5–9 columns indicate 100 times the biases of the estimators for CV, I, II, III, and TNW in the case $N_1 = N_2$. In the table 2, the 5–9 columns indicate 100 times the MSEs of the estimators for CV, I, II, III, and TNW in the case $N_1 = N_2$.

In table 1, we can see that the biases of the three methods I, II, III are small than CV and TNW. On the other hand, we can see that MSE of TNW is smaller than other estimators in table 2. From figure 1 and 2, a bias of all estimators tend to 0 when N is large in both case $p/N = 1/5$ and $3/5$. From

Table 2. (MSE of estimators) ×100

p/N	Δ	p	N_1	CV	I	II	III	TNW
1/5	1.05	6	15	1.437	1.704	1.424	1.429	0.877
		10	25	0.846	0.978	0.838	0.839	0.496
		14	35	0.607	0.687	0.602	0.603	0.357
	1.68	6	15	1.111	1.294	1.092	1.094	0.625
		10	25	0.664	0.755	0.657	0.658	0.366
		14	35	0.473	0.529	0.470	0.470	0.259
	2.56	6	15	0.733	0.846	0.720	0.720	0.369
		10	25	0.433	0.487	0.429	0.429	0.212
		14	35	0.308	0.341	0.306	0.306	0.150
	3.29	6	15	0.460	0.530	0.454	0.452	0.199
		10	25	0.274	0.307	0.272	0.271	0.116
		14	35	0.194	0.215	0.193	0.193	0.081
3/5	1.05	18	15	1.679	2.187	1.671	1.709	1.033
		30	25	0.996	1.238	0.990	1.005	0.611
		56	35	0.707	0.856	0.703	0.711	0.437
	1.68	18	15	1.578	2.029	1.546	1.578	0.967
		30	25	0.920	1.132	0.908	0.921	0.565
		56	35	0.654	0.784	0.647	0.654	0.400
	2.56	18	15	1.367	1.737	1.331	1.348	0.826
		30	25	0.793	0.964	0.781	0.786	0.482
		56	35	0.564	0.669	0.558	0.560	0.341
	3.29	18	15	1.142	1.435	1.110	1.112	0.676
		30	25	0.660	0.797	0.649	0.649	0.388
		56	35	0.460	0.542	0.455	0.454	0.270

figure 3 and 4, we can see that MSEs of all estimators also tend to 0 when N is large, and MSE of the estimators in the case $p/N = 1/5$ are smaller than the case $p/N = 3/5$. Moreover, from figure 5 and 6, we can see that a variance of TNW is smaller than other estimators and a variance of the method I is larger than other estimators. The results mean that a variance of CV is large so that MSE of CV is large, and a variance of the method I is larger than CV.

5. Conclusion

In this paper, we showed that CV is an asymptotic unbiased and a consistent estimator even if the dimension is large. However, the bias of CV increases with the dimension. We knew that MSE of CV is same as MSE

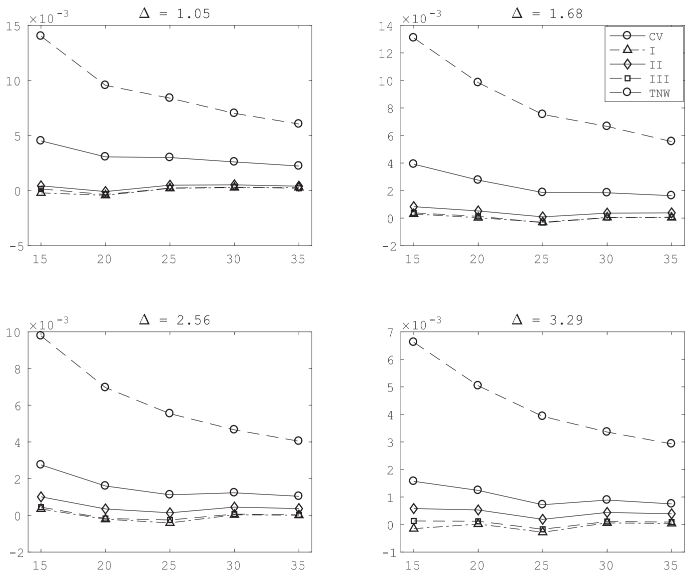

Fig. 1. The figures plot the biases of the estimators for each Δ in the case of $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

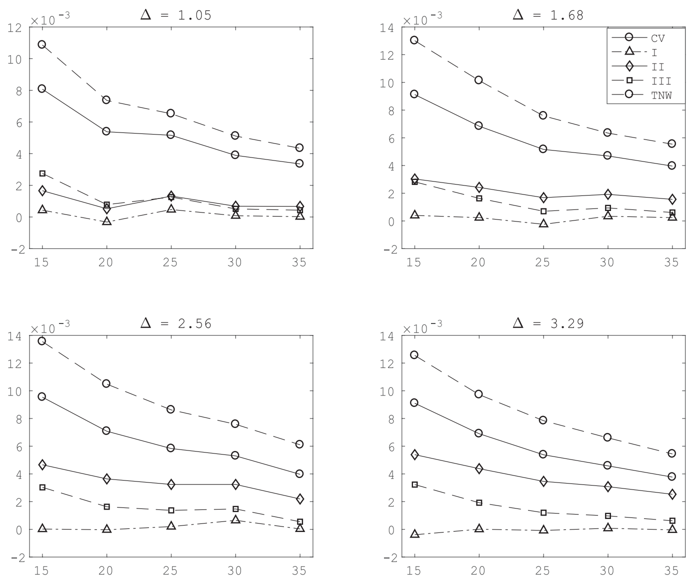

Fig. 2. The figures plot the biases of the estimators for each Δ in the case of $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

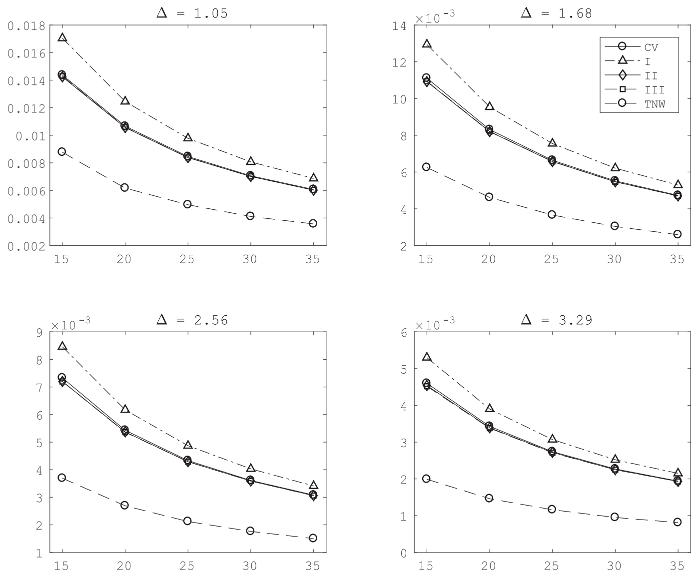

Fig. 3. The figures plot MSEs of the estimators for each Δ in the case of $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

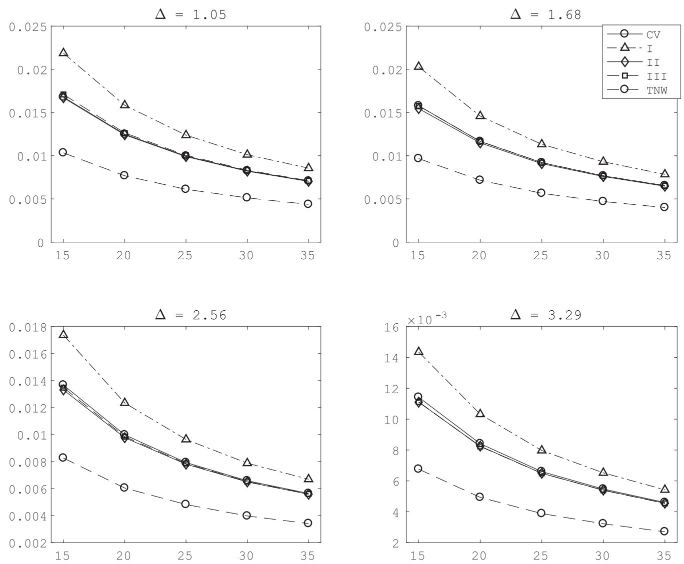

Fig. 4. The figures plot MSEs of the estimators for each Δ in the case of $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

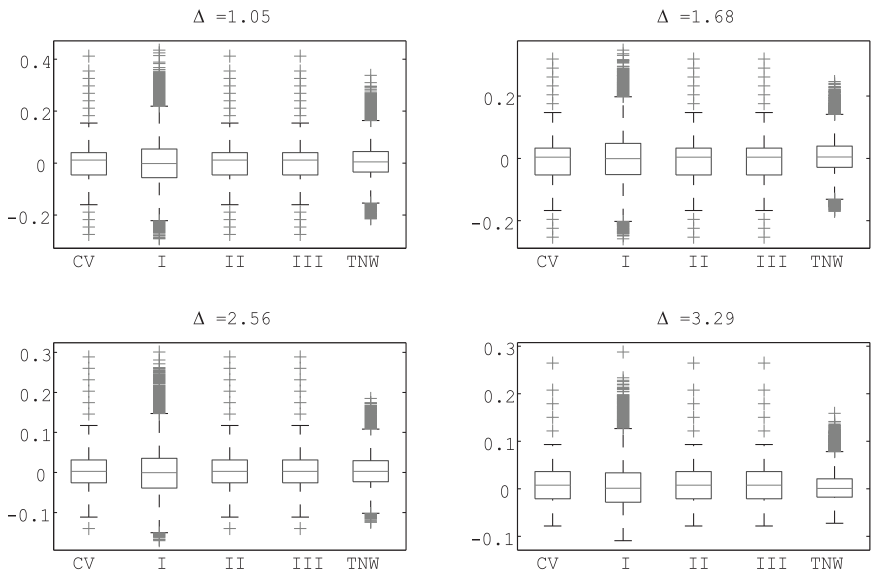

Fig. 5. The figures are the boxplots of $\hat{P}(2|1) - P(2|1)$ for each Δ in the case of $N_1 = 35$ and $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

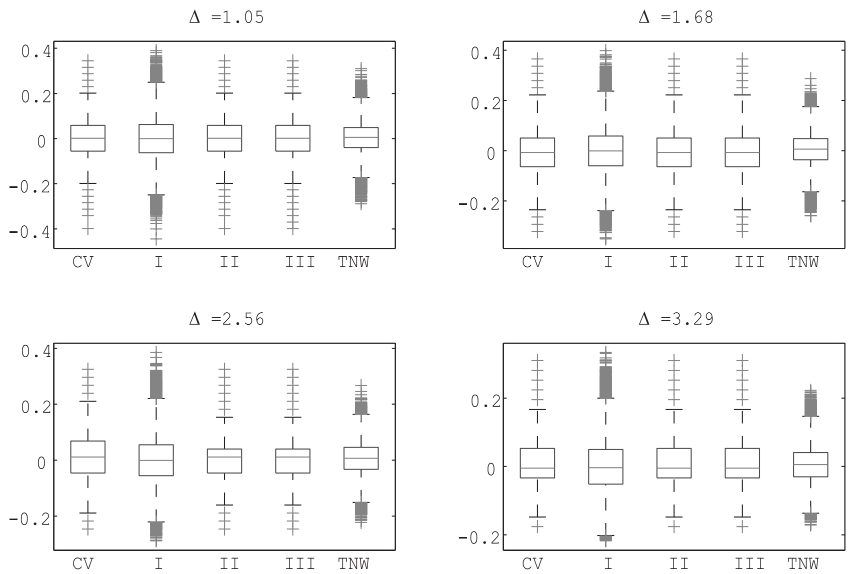

Fig. 6. The figures are the boxplot of $\hat{P}(2|1) - P(2|1)$ for each Δ in the case of $N_1 = 35$ and $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

of the estimator in Tonda *et al.* (2017). Therefore, we knew that CV is a good estimation method of probabilities of misclassification if the sample sizes are sufficiently large. Moreover, we proposed the three methods for correcting the bias of CV in the HD framework and investigated the performances of the three methods in the simulation studies. In simulation studies, we knew that the method I can be applied to many cases, while its MSE is larger than that of other methods. On the other hand, MSEs of the methods II and III are the same as that of CV, while it is necessary to derive the parameters κ and c_1. The method I is a good method if only bias correction is considered, because its method can correct the bias of CV without assumptions. On the other hand, if we can derive the optimal value of κ and d, the methods II and III are better than other methods from a viewpoint of MSE and the computational complexity. However, when the sample sizes are small, we knew that an approximation formula is better than the non-parametric methods. In the future work, we need to show asymptotic properties of CV for various cases (e.g. the non-normal case and the quadratic discriminant) and consider the non-parametric methods for decreasing MSE.

Acknowledgement

The author would like to express the deepest gratitude to Professor Hirofumi Wakaki and Professor Hirokazu Yanagihara of Hiroshima University for their support, helpful comments, and suggestions throughout this research. Finally, the author would like to thank the referees and the editor for their valuable comments and suggestions, which helped to improve this paper.

Appendix

A.1. Lemma of moments. In this section, we show key lemmas for the proof of theorems.

Lemma A.4. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $p \times p$ symmetric matrices, and let $\boldsymbol{Z}$ be an $n \times p$ random matrix and have a normal distribution with* $\mathrm{E}[\boldsymbol{Z}] = \boldsymbol{M}$ *and* $\mathrm{Cov}(\mathrm{vec}(\boldsymbol{Z}^\top)) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, *denoted by* $\boldsymbol{Z} \sim N_{n\times p}(\boldsymbol{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$. *Then, we have the following moments,*

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{Z}^\top\boldsymbol{Z})] = \mathrm{tr}\{\boldsymbol{A}(n\boldsymbol{\Sigma} + \boldsymbol{M}^\top\boldsymbol{M})\},$$

$$\begin{aligned}\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{Z}^\top\boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^\top\boldsymbol{Z})] &= n\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\ \mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + n(n+1)\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}) \\ &\quad + (n+1)\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{B}\boldsymbol{\Sigma}) + (n+1)\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{M}^\top\boldsymbol{M})\end{aligned}$$

$$+ \operatorname{tr}(\boldsymbol{A\Sigma}) \operatorname{tr}(\boldsymbol{BM}^\top \boldsymbol{M}) + \operatorname{tr}(\boldsymbol{AM}^\top \boldsymbol{M}) \operatorname{tr}(\boldsymbol{B\Sigma})$$
$$+ \operatorname{tr}(\boldsymbol{AM}^\top \boldsymbol{MBM}^\top \boldsymbol{M}),$$

$$\begin{aligned}
\mathrm{E}[\operatorname{tr}(\boldsymbol{AZ}^\top \boldsymbol{Z}) \operatorname{tr}(\boldsymbol{BZ}^\top \boldsymbol{Z})] &= n^2 \operatorname{tr}(\boldsymbol{A\Sigma}) \operatorname{tr}(\boldsymbol{B\Sigma}) + 2n \operatorname{tr}(\boldsymbol{A\Sigma B\Sigma}) \\
&\quad + n \operatorname{tr}(\boldsymbol{AM}^\top \boldsymbol{M}) \operatorname{tr}(\boldsymbol{B\Sigma}) + n \operatorname{tr}(\boldsymbol{A\Sigma}) \operatorname{tr}(\boldsymbol{BM}^\top \boldsymbol{M}) \\
&\quad + 2 \operatorname{tr}(\boldsymbol{A\Sigma BM}^\top \boldsymbol{M}) + 2 \operatorname{tr}(\boldsymbol{AM}^\top \boldsymbol{MB\Sigma}) \\
&\quad + \operatorname{tr}(\boldsymbol{AM}^\top \boldsymbol{M}) \operatorname{tr}(\boldsymbol{BM}^\top \boldsymbol{M}).
\end{aligned}$$

The proof of the lemma is given in Gupta and Nagar (2000). From Lemma A.4, we have the following lemma.

LEMMA A.5. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $p \times p$ symmetric matrices, and let $\boldsymbol{W}$ be a $p \times p$ random matrix and have a central Wishart distribution with n degrees of freedom, covariance matrix $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{W} \sim W_p(n, \boldsymbol{\Sigma})$. Then, we have the following moments,*

$$\begin{aligned}
&\mathrm{E}[\operatorname{tr}(\boldsymbol{AW})] = n \operatorname{tr}(\boldsymbol{A\Sigma}), \\
&\mathrm{E}[\operatorname{tr}(\boldsymbol{AW}) \operatorname{tr}(\boldsymbol{BW})] = 2n \operatorname{tr}(\boldsymbol{A\Sigma B\Sigma}) + n^2 \operatorname{tr}(\boldsymbol{A\Sigma}) \operatorname{tr}(\boldsymbol{B\Sigma}), \\
&\mathrm{E}[\operatorname{tr}(\boldsymbol{AWBW})] = n(n+1) \operatorname{tr}(\boldsymbol{A\Sigma B\Sigma}) + n \operatorname{tr}(\boldsymbol{A\Sigma}) \operatorname{tr}(\boldsymbol{B\Sigma}).
\end{aligned}$$

LEMMA A.6. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $p \times p$ symmetric matrices, and let $\boldsymbol{Z} \sim N_{n \times p}(\boldsymbol{M}, \boldsymbol{I}_p \otimes \boldsymbol{I}_n)$ and*

$$\boldsymbol{W} = \sqrt{n}\left(\frac{1}{n}\boldsymbol{Z}^\top \boldsymbol{Z} - \boldsymbol{\Omega}\right).$$

Then, it holds that

$$\begin{aligned}
&\mathrm{E}[\exp\{\operatorname{tr}(\boldsymbol{AW})\} g(\boldsymbol{Z}^\top \boldsymbol{Z})] \\
&\quad = \left|\boldsymbol{I}_p - \frac{2}{\sqrt{n}}\boldsymbol{A}\right|^{-n/2} \mathrm{E}[g(\tilde{\boldsymbol{Z}}^\top \tilde{\boldsymbol{Z}})] \\
&\qquad \times \exp\left[-n^{1/2} \operatorname{tr}(\boldsymbol{A\Omega}) + n^{-1/2} \operatorname{tr}\left\{\boldsymbol{M}^\top \boldsymbol{MA}\left(\boldsymbol{I}_p - \frac{2}{\sqrt{p}}\boldsymbol{A}\right)^{-1}\right\}\right],
\end{aligned}$$

where $\boldsymbol{\Omega} = \boldsymbol{I}_p + n^{-1}\boldsymbol{M}^\top \boldsymbol{M}$ and

$$\tilde{\boldsymbol{Z}} \sim N_{n \times p}\left(\boldsymbol{M}\left(\boldsymbol{I}_p - \frac{2}{\sqrt{n}}\boldsymbol{A}\right)^{-1}, \left(\boldsymbol{I}_p - \frac{2}{\sqrt{n}}\boldsymbol{A}\right)^{-1} \otimes \boldsymbol{I}_n\right).$$

A.2. Proof of Lemma 1. Suppose that

$$\boldsymbol{u} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{W} = n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2} \sim W_p(n, \boldsymbol{I}_p),$$

$$\boldsymbol{z}_1 = \sqrt{N_1}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{z}_2 = \sqrt{N_2}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1) \sim N_p(\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{I}_p),$$

where $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Omega} = \boldsymbol{M}^\top\boldsymbol{M}$ and $\boldsymbol{M} = (\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{0}, \boldsymbol{0})$. Let $\boldsymbol{Q} = (\boldsymbol{u}, \boldsymbol{z}_1, \boldsymbol{z}_2)$, then

$$\boldsymbol{V}_2 = \boldsymbol{Q}^\top\boldsymbol{Q} \sim W_3(p, \boldsymbol{I}_3, \Omega),$$

$$\boldsymbol{V}_1 = \boldsymbol{T}(\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q})^{-1}\boldsymbol{T}^\top \sim W_3(N-p, \boldsymbol{I}_3),$$

$$\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top,$$

where $\boldsymbol{T}$ is Bartlett's decomposition of $\boldsymbol{V}_2$, that is, $\boldsymbol{V}_2 = \boldsymbol{T}\boldsymbol{T}^\top$. Let $\boldsymbol{U} = (u_{ij}) = \boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q}$ then we show that $D(\boldsymbol{x})$ is expressed by u_{ij}. Therefore, we easily have (4).

A.3. Proof of Theorem 3. Put $\boldsymbol{W}_1 = \sqrt{N-p}((N-p)^{-1}\boldsymbol{V}_1 - \boldsymbol{I}_3) = O_p(1)$. From Lemma 1,

$$\begin{aligned}\boldsymbol{U} &= \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top \\ &= \frac{p}{N-p}\{\tilde{\boldsymbol{V}}_2 - (N-p)^{-1/2}\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + (N-p)^{-1}\tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top\} + O_p((N-p)^{3/2}),\end{aligned}$$

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{U}) = \frac{p}{N-p}\{\operatorname{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + a_0 + a_1\} + O_p((N-p)^{-1}),$$

where $\tilde{\boldsymbol{T}} = p^{-1/2}\boldsymbol{T}$ and $\tilde{\boldsymbol{V}}_2 = p^{-1}\boldsymbol{V}_2$,

$$a_\ell = \frac{p}{N-p}(-1)^{\ell+1}(N-p)^{-(\ell+1)/2}\operatorname{tr}(\boldsymbol{A}\tilde{\boldsymbol{T}}\boldsymbol{W}_1^{\ell+1}\tilde{\boldsymbol{T}}^\top).$$

Then it can be expanded as

$$\begin{aligned}&\mathrm{E}[\exp\{it\operatorname{tr}(\boldsymbol{A}\boldsymbol{U})\} \mid \boldsymbol{V}_2] \\ &\quad = \mathrm{E}\left[\exp\left[it\frac{p}{N-p}\{\operatorname{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + a_0 + a_1\}\right]\middle|\, \boldsymbol{V}_2\right] + O_p((N-p)^{-1}) \\ &\quad = \exp\left\{it\frac{p}{N-p}\operatorname{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)\right\}\mathrm{E}[e^{ita_0}(1+b_1) \mid \boldsymbol{V}_2] + O_p((N-p)^{-1}),\end{aligned}$$

where $i = \sqrt{-1}$,

$$b_1 = it\frac{p}{N-p}a_1.$$

From $\boldsymbol{V}_1 \sim W_3(N-p, \boldsymbol{I}_3)$, $a_0 = \mathrm{tr}(\boldsymbol{M}_0\boldsymbol{W}_1)$ and Lemma A.6,

$$\begin{aligned}
&\mathrm{E}[e^{ita_0}g(\boldsymbol{V}_1) \mid \boldsymbol{V}_2] \\
&\quad = \left|\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right|^{-(N-p)/2} \exp\{-\sqrt{N-p}\ \mathrm{tr}(\boldsymbol{M}_0)\}\mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1)] \\
&\quad = \exp\{\mathrm{tr}(\boldsymbol{M}_0^2)\}\left\{1 + \frac{4}{3\sqrt{N-p}}\ \mathrm{tr}(\boldsymbol{M}_0^3)\right\}\mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1)] + O_p((N-p)^{-1}),
\end{aligned}$$

where

$$\tilde{\boldsymbol{Z}}_1 \sim N_{(N-p)\times 3}\left(\boldsymbol{O}, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1} \otimes \boldsymbol{I}_{N-p}\right),$$

$$\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 \sim W_3\left(N-p, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1}\right)$$

are independent of $\boldsymbol{V}_2$, and $\boldsymbol{M}_0 = -itp(N-p)^{-3/2}\tilde{\boldsymbol{T}}^\top \boldsymbol{A}\tilde{\boldsymbol{T}}$ and $g(\boldsymbol{V}_1) = 1 + b_1$. The moments are given by

$$\begin{aligned}
\mathrm{E}[b_1|\boldsymbol{V}_2] &= it\frac{p}{N-p}\mathrm{E}[a_1|\boldsymbol{V}_2], \\
\mathrm{E}[a_1|\boldsymbol{V}_2] &= \mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top \boldsymbol{A}\tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_3\right)^2\right\}\right] \\
&= \frac{1}{N-p}[4\ \mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + 3\ \mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}\tilde{\boldsymbol{T}}\boldsymbol{M}_0)] + O_p((N-p)^{-1}) \\
&= \frac{1}{N-p}\left[4\ \mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) - 3it\frac{p}{(N-p)^{3/2}}\ \mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^2\}\right] + O_p((N-p)^{-1}).
\end{aligned}$$

Secondly, let $\boldsymbol{W}_2 = \sqrt{p}(p^{-1}\boldsymbol{V}_2 - \boldsymbol{\Omega}^*)$, then $\boldsymbol{W}_2 = O_p(1)$ from the central limit theorem, where $\boldsymbol{\Omega}^* = \boldsymbol{I}_3 + p^{-1}\boldsymbol{\Omega}$. We can obtain the following expansions:

$$\begin{aligned}
\boldsymbol{V}_2 &= p(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2), \\
\mathrm{tr}(\boldsymbol{M}_0) &= itp(N-p)^{-3/2}\ \mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2),
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) &= \mathrm{tr}(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2)) \\
&= \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*) + p^{-1/2}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2),
\end{aligned}$$

$$\mathrm{tr}(\boldsymbol{M}_0^2) = (it)^2 p^2 (N-p)^{-3}\,\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^2\},$$

$$\begin{aligned}
\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^2\} &= \mathrm{tr}\{(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2))^2\} \\
&= \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^2\} + 2p^{-1/2}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p(1),
\end{aligned}$$

$$\mathrm{tr}(\boldsymbol{M}_0^3) = -(it)^3 p^3 (N-p)^{-9/2}\,\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^3\},$$

$$\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^3\} = \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^3\} + O_p(p^{-1/2}) = O_{1/2}.$$

Since $\boldsymbol{V}_2 \sim W_3(p, \boldsymbol{I}_3, \boldsymbol{\Omega})$, we obtain the following expansions:

$$\begin{aligned}
&\exp\left\{it\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + \mathrm{tr}(\boldsymbol{M}_0^2)\right\} \\
&\quad = \exp\left\{it\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*) + it\frac{p^{1/2}}{N-p}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2) \right. \\
&\qquad\qquad \left. + (it)^2 p^2 (N-p)^{-3}\,\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^2\}\right\} \\
&\qquad \times \exp\{2(it)^2 p^{3/2} (N-p)^{-3}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p((N-p)^{-1})\},
\end{aligned}$$

$$\begin{aligned}
&\exp\{2(it)^2 p^{3/2} (N-p)^{-3}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p((N-p)^{-1})\} \\
&\quad = 1 + 2(it)^2 p^{3/2} (N-p)^{-3}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_1.
\end{aligned}$$

Put $\boldsymbol{M}_0^* = itp^{3/2}(N-p)^{-3}\boldsymbol{A}$. From Lemma A.6, we can have

$$\begin{aligned}
&\mathrm{E}[\exp\{\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}h(\boldsymbol{Z}_2^\top \boldsymbol{Z}_2)] \\
&\quad = \left|\boldsymbol{I}_3 - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right|^{-p/2} \mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)] \\
&\qquad \times \exp\left[-p^{1/2}\,\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{\Omega}^*) + p^{-1/2}\,\mathrm{tr}\left\{\boldsymbol{\Omega}\boldsymbol{M}_0^*\left(\boldsymbol{I}_3 - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right\}\right] \\
&\quad = \exp[\mathrm{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\}]\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)] \\
&\qquad \times \left[1 + \frac{4}{3\sqrt{p}}\,\mathrm{tr}\{(\boldsymbol{I}_3 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\}\right] + O_1.
\end{aligned}$$

Moreover, since $\mathrm{tr}\{(\boldsymbol{I}_3 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\} = O_{1/2}$,

$$
\begin{aligned}
&\mathrm{E}[\exp\{\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}h(\boldsymbol{Z}_2^\top\boldsymbol{Z}_2)] \\
&\qquad = \exp[\mathrm{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\}]\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2)] + O_1,
\end{aligned}
$$

where $h(\boldsymbol{Z}_2^\top\boldsymbol{Z}_2) = (1 + 2(it)^2 p^{-1/2}(N-p)^{-1}\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2))$, and $\boldsymbol{Z}_1$ and $\tilde{\boldsymbol{Z}}$ are the random matrices that satisfy

$$
\begin{aligned}
&\boldsymbol{V}_2 = \boldsymbol{Z}_2^\top\boldsymbol{Z}_2, \\
&\boldsymbol{Z}_2 \sim N_{p\times 3}(\boldsymbol{M}, \boldsymbol{I}_3 \otimes \boldsymbol{I}_p), \\
&\tilde{\boldsymbol{Z}}_2 \sim N_{p\times 3}(\boldsymbol{M}(\boldsymbol{I}_3 - 2p^{-1/2}\boldsymbol{M}_0^*)^{-1}, (\boldsymbol{I}_3 - 2p^{-1/2}\boldsymbol{M}_0^*)^{-1} \otimes \boldsymbol{I}_p).
\end{aligned}
$$

The moments are given by

$$
\begin{aligned}
&\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2)] = 1 + 2(it)^2 p^{-1/2}(N-p)^{-1}\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\mathrm{E}[\tilde{\boldsymbol{W}}_2]), \\
&\mathrm{E}[\boldsymbol{W}_2] = \sqrt{p}\left\{\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right. \\
&\qquad\qquad \left. + p^{-1}\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\boldsymbol{\Omega}\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right\} - \boldsymbol{\Omega}^* = O_{1/2},
\end{aligned}
$$

where

$$
\tilde{\boldsymbol{W}}_2 = \sqrt{p}\left(\frac{1}{p}\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2 - \boldsymbol{\Omega}^*\right).
$$

From above result, we have

$$
\begin{aligned}
\eta &= \frac{p}{N-p}\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*) = \frac{n}{N-p}\left(\Delta^2 + \frac{p}{N_2} - \frac{bp}{N_1} + p(1-b)\right), \\
s^2 &= 2\left[p^2(N-p)^{-3}\ \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^2\} + \frac{p}{(N-p)^2}\ \mathrm{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})\boldsymbol{A}^2\}\right] \\
&= 4\frac{n^2 N}{(N-p)^3}\left(\Delta^2 + \frac{pb^2}{N_1} + \frac{p}{N_2}\right).
\end{aligned}
$$

Therefore, we have the characteristic function $\phi(t)$ of $D_b(\boldsymbol{x})$ as

$$
\phi(t) = \exp(it\eta - t^2 s^2/2) + O_1.
$$

From this expansion, we can have the result of Theorem 3 by using the inversion formula.

A.4. Proof of Lemma 3. The proof of Lemma 3 imitates the proof of Lemma 1. Suppose that

$$\boldsymbol{u}_{1i} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_{1i} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{u}_{1j} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{W} = (n-2)\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}^{(i,j)}\boldsymbol{\Sigma}^{-1/2} \sim W_p(n-2, \boldsymbol{I}_p),$$

$$\boldsymbol{z}_1 = \sqrt{n_1 - 1}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_1^{(i,j)} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),$$

$$\boldsymbol{z}_2 = \sqrt{N_2}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1) \sim N_p(\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{I}_p),$$

where $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Omega} = \boldsymbol{M}^\top\boldsymbol{M}$ and $\boldsymbol{M} = (\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$.

Let $\boldsymbol{Q} = (\boldsymbol{u}_{1i}, \boldsymbol{u}_{1j}, \boldsymbol{z}_1, \boldsymbol{z}_2)$, then

$$\boldsymbol{V}_2 = \boldsymbol{Q}^\top\boldsymbol{Q} \sim W_4(p, \boldsymbol{I}_4, \boldsymbol{\Omega}),$$

$$\boldsymbol{V}_1 = \boldsymbol{T}^\top(\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q})^{-1}\boldsymbol{T} \sim W_4(N-p, \boldsymbol{I}_4),$$

$$\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top,$$

where $\boldsymbol{T}$ is Bartlett's decomposition of $\boldsymbol{V}_2$, that is, $\boldsymbol{V}_2 = \boldsymbol{T}\boldsymbol{T}^\top$. Let $\boldsymbol{U} = (u_{ij}) = \boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q}$ then $D^{(i)}(\boldsymbol{x}_{1i})$ and $D^{(j)}(\boldsymbol{x}_{1j})$ are expressed by u_{ij} from Lemma 2. Therefore, we easily have (6), (7).

A.5. Expansion of $\phi(\boldsymbol{t})$. Let

$$\boldsymbol{W}_1 = \sqrt{N-p}\left(\frac{1}{N-p}\boldsymbol{V}_1 - \boldsymbol{I}_4\right),$$

then $\boldsymbol{W}_1 = O_p(1)$ from the central limit theorem. From Lemma 2,

$$\boldsymbol{V}_1 = (N-p)\left(\boldsymbol{I}_4 + \frac{1}{\sqrt{N-p}}\boldsymbol{W}_1\right),$$

$$D_b^{(-i)}(\boldsymbol{x}_{1i}) = \mathrm{tr}(\boldsymbol{A}_i\boldsymbol{U}) - T_i^{-1}\boldsymbol{a}_i^\top\boldsymbol{U}\boldsymbol{A}_i\boldsymbol{U}\boldsymbol{a}_i, \qquad (i = 1, 2).$$

Then, we obtain an expansion of $\boldsymbol{U}$ as follows:

$$\boldsymbol{U} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top$$
$$= \frac{p}{N-p}\tilde{\boldsymbol{T}}\left\{\boldsymbol{I}_4 - \frac{1}{\sqrt{N-p}}\boldsymbol{W}_1 + \frac{1}{N-p}\boldsymbol{W}_1^2 - \frac{1}{(\sqrt{N-p})^3}\boldsymbol{W}_1^3\right\}\tilde{\boldsymbol{T}}^\top + O_p(N^{-2}),$$

where $\tilde{\boldsymbol{T}} = p^{-1/2}\boldsymbol{T} = O_p(1)$. From above result, it can be expanded as

$$\mathrm{tr}(\boldsymbol{A}_i\boldsymbol{U}) = \frac{p}{N-p}\{\mathrm{tr}(\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2) + a_{i,0} + a_{i,1}\} + O_p(N^{-3/2}),$$

$$T_i = \frac{N_1-1}{N_1-2} + \mathrm{tr}(\boldsymbol{B}_i\boldsymbol{U}) = b_{i,0} + b_{i,1} + b_{i,2} + O_p(N^{-3/2}),$$

$$\boldsymbol{a}_i^\top \boldsymbol{U}\boldsymbol{A}_i\boldsymbol{U}\boldsymbol{a}_i = \frac{p^2}{(N-p)^2}\{\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i + c_{i,0} + c_{i,1} + c_{i,2}\} + O_p(N^{-3/2}),$$

$$T_i^{-1} = s_{i,0} + s_{i,1} + s_{i,2} + O_p(N^{-3/2}),$$

where $\tilde{\boldsymbol{V}}_2 = p^{-1}\boldsymbol{V}_2$ and

$$a_{i,\ell} = (-1)^{\ell+1}(N-p)^{-(\ell+1)/2}\,\mathrm{tr}(\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^{\ell+1}\tilde{\boldsymbol{T}}^\top), \qquad (\ell = 0,1,2),$$

$$b_{i,0} = \frac{N_1-1}{N_1-2} + \frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2),$$

$$b_{i,\ell} = (-1)^{\ell}(N-p)^{-\ell/2}\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^{\ell}\tilde{\boldsymbol{T}}^\top), \qquad (\ell = 1,2),$$

$$c_{i,0} = -(N-p)^{-1/2}\boldsymbol{a}_i^\top(\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2)\boldsymbol{a}_i,$$

$$c_{i,1} = (N-p)^{-1}\boldsymbol{a}_i^\top(\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2)\boldsymbol{a}_i,$$

$$s_{i,0} = b_{i,0}^{-1}, \qquad s_{i,1} = b_{i,1}b_{i,0}^{-2} \qquad s_{i,2} = b_{i,0}^{-3}(b_{i,1}^2 - b_{i,0}b_{i,2}).$$

Then $D_b^{(-i)}$ is expanded as follows:

$$D_b^{(-i)}(\boldsymbol{x}_{1i}) = \frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2) - s_{i,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i + D_{i,0} + D_{i,1} + O_p(N^{-1}),$$

where

$$D_{i,0} = \frac{p}{N-p}a_{i,0} - \frac{p^2}{(N-p)^2}(s_{i,0}c_{i,0} + s_{i,1}\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i),$$

$$D_{i,1} = \frac{p}{N-p}a_{i,1} - \frac{p^2}{(N-p)^2}(s_{i,0}c_{i,1} + s_{i,1}c_{i,0} + s_{i,2}\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i).$$

We consider the characteristic function of joint distribution of $D_b^{(-1)}(\boldsymbol{x}_{11})$ and $D_b^{(-2)}(\boldsymbol{x}_{12})$, that is,

$$\phi(\boldsymbol{t}) = \mathrm{E}[\exp\{it_1D_b^{(-1)}(\boldsymbol{x}_{11}) + it_2D_b^{(-2)}(\boldsymbol{x}_{12})\}],$$

where $\boldsymbol{t} = (t_1,t_2)^\top$ and $i = \sqrt{-1}$.

Firstly, we consider the following conditional expectation given $\tilde{\boldsymbol{V}}_2$,

$$
\begin{aligned}
&\mathrm{E}[\exp\{it_1 D_b^{(-1)}(\boldsymbol{x}_{11}) + it_2 D_b^{(-2)}(\boldsymbol{x}_{12})\} | \tilde{\boldsymbol{V}}_2] \\
&\quad = \exp\left[it_1 \left\{ \frac{p}{N-p} \operatorname{tr}(\boldsymbol{A}_1 \tilde{\boldsymbol{V}}_2) - s_{1,0} \frac{p^2}{(N-p)^2} \boldsymbol{a}_1^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_1 \tilde{\boldsymbol{V}}_2 \boldsymbol{a}_1 \right\} \right. \\
&\qquad\qquad \left. + it_2 \left\{ \frac{p}{N-p} \operatorname{tr}(\boldsymbol{A}_2 \tilde{\boldsymbol{V}}_2) - s_{2,0} \frac{p^2}{(N-p)^2} \boldsymbol{a}_2^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_2 \tilde{\boldsymbol{V}}_2 \boldsymbol{a}_2 \right\} \right] \\
&\qquad\qquad \times \mathrm{E}[\exp(it_1 D_{1,0} + it_2 D_{2,0} + it_1 D_{1,1} + it_2 D_{2,1} + O_p(N^{-1})) \mid \tilde{\boldsymbol{V}}_2].
\end{aligned}
$$

We expand the following conditional expectation,

$$
\begin{aligned}
&\mathrm{E}[\exp(it_1 D_{1,0} + it_2 D_{2,0} + it_1 D_{1,1} + it_2 D_{2,1} + O_p(N^{-1})) \mid \tilde{\boldsymbol{V}}_2] \\
&\quad = \mathrm{E}[\exp(it_1 D_{1,0} + it_2 D_{2,0})(1 + it_1 D_{1,1} + it_2 D_{2,1}) \mid \tilde{\boldsymbol{V}}_2] + O_p(N^{-1}).
\end{aligned}
$$

Let

$$
\begin{aligned}
\boldsymbol{M}_0 &= it_1 \boldsymbol{M}_{1,0} + it_2 \boldsymbol{M}_{2,0}, \\
\boldsymbol{M}_{j,0} &= (N-p)^{-1/2} \tilde{\boldsymbol{T}}^\top \left[-\frac{p}{N-p} \boldsymbol{A}_j + \frac{p^2}{(N-p)^2} s_{j,0} \right. \\
&\qquad \left. \times \left\{ \boldsymbol{B}_j \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_j + \boldsymbol{A}_j \tilde{\boldsymbol{V}}_2 \boldsymbol{B}_j + s_{j,0} \frac{p}{N-p} \boldsymbol{a}_j \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_j \tilde{\boldsymbol{V}}_2 \boldsymbol{a}_j \boldsymbol{B}_j \right\} \right] \tilde{\boldsymbol{T}},
\end{aligned}
$$

then we have

$$
\exp(it_1 D_{1,0} + it_2 D_{2,0}) = \exp\{\operatorname{tr}(\boldsymbol{M}_0 \boldsymbol{W}_1)\}.
$$

From $\boldsymbol{V}_1 \sim W_4(N-p, \boldsymbol{I}_4)$ and and Lemma A.6,

$$
\begin{aligned}
&\mathrm{E}[\exp\{\operatorname{tr}(\boldsymbol{M}_0 \boldsymbol{W}_1)\} g(\boldsymbol{V}_1) \mid \boldsymbol{V}_2] \\
&\quad = \left| \boldsymbol{I}_4 - \frac{2}{\sqrt{N-p}} \boldsymbol{M}_0 \right|^{-(N-p)/2} \exp\{-\sqrt{N-p} \operatorname{tr}(\boldsymbol{M}_0)\} \mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1) \mid \boldsymbol{V}_2],
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{\boldsymbol{Z}}_1 &\sim N_{(N-p)\times 4}\left(\boldsymbol{O}, \left(\boldsymbol{I}_4 - \frac{2}{\sqrt{N-p}} \boldsymbol{M}_0 \right)^{-1} \otimes \boldsymbol{I}_{N-p} \right), \\
\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 &\sim W_4\left(N-p, \left(\boldsymbol{I}_4 - \frac{2}{\sqrt{N-p}} \boldsymbol{M}_0 \right)^{-1} \right)
\end{aligned}
$$

are independent of $\boldsymbol{V}_2$, and

$$g(\boldsymbol{V}_1) = 1 + it_1 D_{1,1} + it_2 D_{2,1}.$$

Put

$$h(\boldsymbol{V}_2) = \mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1) \mid \boldsymbol{V}_2] = 1 + it_1 \mathrm{E}[D_{1,1}|\boldsymbol{V}_2] + it_2 \mathrm{E}[D_{2,1}|\boldsymbol{V}_2].$$

For $i = 1, 2$, we have

$$\begin{aligned}\mathrm{E}[D_{i,1}|\boldsymbol{V}_2] &= \frac{p}{N-p}\mathrm{E}[a_{i,1}|\boldsymbol{V}_2] \\ &\quad - \frac{p^2}{(N-p)^2}(s_{i,0}\mathrm{E}[c_{i,1}|\boldsymbol{V}_2] + \mathrm{E}[s_{i,1}c_{i,0}|\boldsymbol{V}_2] + \boldsymbol{a}_i^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{a}_i \mathrm{E}[s_{i,2}|\boldsymbol{V}_2]).\end{aligned}$$

The moments are given by

$$\begin{aligned}\mathrm{E}[a_{i,1}|\boldsymbol{V}_2] &= \mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)^2\right\} \middle| \boldsymbol{V}_2\right] \\ &= \frac{1}{N-p}[5\,\mathrm{tr}(\boldsymbol{A}_i \tilde{\boldsymbol{V}}_2) + 4\,\mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{T}} \boldsymbol{M}_0^2)] + O_p((N-p)^{-1}), \\ \mathrm{E}[c_{i,1}|\boldsymbol{V}_2] &= \mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)\right\}\right. \\ &\qquad \left. + \mathrm{tr}\left\{(\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{T}} + \tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{B}_i \tilde{\boldsymbol{T}})\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)^2\right\} \middle| \boldsymbol{V}_2\right] \\ &= \frac{1}{N-p}\{\mathrm{tr}(\boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2) + 4\,\mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{T}} \boldsymbol{M}_0 \tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{T}} \boldsymbol{M}_0) \\ &\quad + \mathrm{tr}(\boldsymbol{B}_i \tilde{\boldsymbol{V}}_2)\,\mathrm{tr}(\boldsymbol{A}_i \tilde{\boldsymbol{V}}_2) + 5\,\mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{T}} + \tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{B}_i \tilde{\boldsymbol{T}}) \\ &\quad + 4\,\mathrm{tr}((\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{T}} + \tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{B}_i \tilde{\boldsymbol{T}})\boldsymbol{M}_0^2)\} + O_p((N-p)^{-1}), \\ \mathrm{E}[s_{i,1}c_{i,0}|\boldsymbol{V}_2] &= b_{i,0}^{-2}\mathrm{E}[b_{i,1}c_{i,0}|\boldsymbol{V}_2], \\ \mathrm{E}[s_{i,2}|\boldsymbol{V}_2] &= b_{i,0}^{-3}(\mathrm{E}[b_{i,1}^2|\boldsymbol{V}_2] - b_{i,0}\mathrm{E}[b_{i,2}|\boldsymbol{V}_2]), \\ \mathrm{E}[b_{i,1}c_{i,0}|\boldsymbol{V}_2] &= \frac{p}{N-p}\mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)\right\}\right. \\ &\qquad \left. \times \mathrm{tr}\left\{(\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_i \tilde{\boldsymbol{T}} + \tilde{\boldsymbol{T}}^\top \boldsymbol{A}_i \tilde{\boldsymbol{V}}_2 \boldsymbol{B}_i \tilde{\boldsymbol{T}})\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)\right\} \middle| \boldsymbol{V}_2\right]\end{aligned}$$

$$= \frac{2p}{(N-p)^2}[\mathrm{tr}\{\boldsymbol{B}_i(\tilde{\boldsymbol{V}}_2\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2 + \tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2)\}$$

$$+ 2\,\mathrm{tr}(\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{M}_0)\,\mathrm{tr}\{(\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{T}} + \tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{B}_i\tilde{\boldsymbol{T}})\boldsymbol{M}_0\}]$$

$$+ O_p((N-p)^{-1}),$$

$$\mathrm{E}[b_{i,1}^2|\boldsymbol{V}_2] = \frac{p^2}{(N-p)^2}\mathrm{E}\left[\left(\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top\tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)\right\}\right)^2\,\middle|\,\boldsymbol{V}_2\right]$$

$$= \frac{2p^2}{(N-p)^3}[\mathrm{tr}\{(\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2)^2\} + 2\{\mathrm{tr}(\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{M}_0)\}^2] + O_p((N-p)^{-3/2}),$$

$$\mathrm{E}[b_{i,2}|\boldsymbol{V}_2] = \frac{p}{(N-p)}\mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top\tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4\right)^2\right\}\,\middle|\,\boldsymbol{V}_2\right]$$

$$= \frac{p}{(N-p)^2}[5\,\mathrm{tr}(\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2) + 4\,\mathrm{tr}(\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{M}_0^2)] + O_p((N-p)^{-3/2}).$$

Secondly, let

$$\boldsymbol{W}_2 = \sqrt{p}\left(\frac{1}{p}\boldsymbol{V}_2 - \boldsymbol{\Omega}^*\right), \tag{A.10}$$

then $\boldsymbol{W}_2 = O_p(1)$ from the central limit theorem where $\boldsymbol{\Omega}^* = \boldsymbol{I}_4 + p^{-1}\boldsymbol{\Omega}$. We obtain the following expansion by using (A.5).

$$\boldsymbol{V}_2 = p\left(\boldsymbol{\Omega}^* + \frac{1}{\sqrt{p}}\boldsymbol{W}_2\right),$$

$$\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2) - s_{i,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i$$

$$= \frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_i\boldsymbol{\Omega}^*) - s_{i,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_i^\top\boldsymbol{\Omega}^*\boldsymbol{A}_i\boldsymbol{\Omega}^*\boldsymbol{a}_i$$

$$+ a_{i,0}^* + a_{i,1}^* + O_p(p^{-1}),$$

$$b_{i,0} = b_{i,0,0} + b_{i,0,1},$$

$$s_{i,0} = s_{i,0,0} + s_{i,0,1} + s_{i,0,2} + O_p(p^{-3/2}),$$

$$\mathrm{tr}(\boldsymbol{M}_0^2) = \mathrm{tr}\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^2\} + 2\,\mathrm{tr}\{\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*(p^{-1/2}\boldsymbol{\Xi}_0\boldsymbol{W}_2 + \boldsymbol{\Xi}_1\boldsymbol{\Omega}^*)\} + O_p(p^{-1}),$$

$$\mathrm{tr}(\boldsymbol{M}_0^3) = \mathrm{tr}\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^3\} + O_p(p^{-1/2}),$$

$$h(\boldsymbol{V}_2) = h(p\boldsymbol{\Omega}^*) + O_p(p^{-1/2}(N-p)^{-1/2})$$
$$= 1 + O_1,$$

where

$$a_{i,0}^* = -\frac{p^2}{(N-p)^2} s_{i,0,1} \boldsymbol{a}_i^\top \boldsymbol{\Omega}^* \boldsymbol{A}_i \boldsymbol{\Omega}^* \boldsymbol{a}_i$$
$$+ \frac{1}{\sqrt{p}} \left\{ \frac{p}{N-p} \operatorname{tr}(\boldsymbol{A}_i \boldsymbol{W}_2) - \frac{p^2}{(N-p)^2} s_{i,0,0} (\boldsymbol{a}_i^\top \boldsymbol{\Omega}^* \boldsymbol{A}_i \boldsymbol{W}_2 \boldsymbol{a}_i + \boldsymbol{a}_i^\top \boldsymbol{W}_2 \boldsymbol{A}_i \boldsymbol{\Omega}^* \boldsymbol{a}_i) \right\},$$

$$a_{i,1}^* = -\frac{p^2}{(N-p)^2} s_{i,0,2} \boldsymbol{a}_i^\top \boldsymbol{\Omega}^* \boldsymbol{A}_i \boldsymbol{\Omega}^* \boldsymbol{a}_i - \frac{p}{(N-p)^2} s_{i,0,0} \boldsymbol{a}_i^\top \boldsymbol{W}_2 \boldsymbol{A}_i \boldsymbol{W}_2 \boldsymbol{a}_i$$
$$- \frac{p^{3/2}}{(N-p)^2} s_{i,0,1} (\boldsymbol{a}_i^\top \boldsymbol{\Omega}^* \boldsymbol{A}_i \boldsymbol{W}_2 \boldsymbol{a}_i + \boldsymbol{a}_i^\top \boldsymbol{W}_2 \boldsymbol{A}_i \boldsymbol{\Omega}^* \boldsymbol{a}_i),$$

$$b_{i,0,0} = \frac{n_1}{n_1 - 1} + \frac{p}{N-p} \operatorname{tr}(\boldsymbol{B}_i \boldsymbol{\Omega}^*) = \frac{n_1}{n_1 - 1} + \frac{p}{N-p}\left(1 + \frac{1}{n_1}\right),$$

$$b_{i,0,1} = \frac{\sqrt{p}}{N-p} \operatorname{tr}(\boldsymbol{B}_i \boldsymbol{W}_2),$$

$$s_{i,0,0} = b_{i,0,0}^{-1} = \frac{n_1(n_1-1)(N-p)}{n_1^2(N-p) + p(n_1-1)(n_1+1)},$$

$$s_{i,0,1} = -b_{i,0,0}^{-2} b_{i,0,1}, \qquad s_{i,0,2} = b_{i,0,0}^{-3} b_{i,0,1}^2,$$

$$\boldsymbol{\Xi}_s = it_1 \boldsymbol{\Xi}_{1,s} + it_2 \boldsymbol{\Xi}_{2,s},$$

$$\boldsymbol{\Xi}_{j,0} = (N-p)^{-1/2} \left[-\frac{p}{N-p} \boldsymbol{A}_j \right.$$
$$\left. + \frac{p^2}{(N-p)^2} s_{j,0,0} \left\{ \boldsymbol{B}_j \boldsymbol{\Omega}^* \boldsymbol{A}_j + \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{B}_j + s_{j,0,0} \frac{p}{N-p} \boldsymbol{a}_j \boldsymbol{\Omega}^* \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{a}_j \boldsymbol{B}_j \right\} \right],$$

$$\boldsymbol{\Xi}_{j,1} = (N-p)^{-1/2} \frac{p^2}{(N-p)^2} s_{j,0,1} (\boldsymbol{B}_j \boldsymbol{\Omega}^* \boldsymbol{A}_j + \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{B}_j + s_{j,0,0} \boldsymbol{a}_j^\top \boldsymbol{\Omega}^* \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{a}_j \boldsymbol{B}_j)$$
$$+ \frac{1}{\sqrt{p}} s_{j,0,0}^2 (\boldsymbol{a}_j^\top \boldsymbol{\Omega}^* \boldsymbol{A}_j \boldsymbol{W}_2 \boldsymbol{a}_j \boldsymbol{B}_j + \boldsymbol{a}_j^\top \boldsymbol{W}_2 \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{a}_j \boldsymbol{B}_j) + s_{j,0,1} \boldsymbol{a}_j^\top \boldsymbol{\Omega}^* \boldsymbol{A}_j \boldsymbol{\Omega}^* \boldsymbol{a}_j \boldsymbol{B}_j.$$

Let

$$\boldsymbol{M}_0^* = it_1\boldsymbol{M}_{1,0}^* + it_2\boldsymbol{M}_{2,0}^*$$

$$\boldsymbol{M}_{j,0}^* = \frac{\sqrt{p}}{N-p}\left\{\frac{p^2}{(N-p)^2}s_{j,0,0}^2\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j + \boldsymbol{A}_j\right.$$

$$\left. - \frac{p}{N-p}s_{j,0,0}(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j + \boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{B}_j)\right\},$$

then we have

$$\exp(it_1a_{1,0}^* + it_2a_{2,0}^*) = \exp\{\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}.$$

Therefore, we can expand $\phi(\boldsymbol{t})$ as follows:

$$\phi(\boldsymbol{t}) = \mathrm{E}\left[\exp\left[it_1\left\{\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_1\tilde{\boldsymbol{V}}_2) - s_{1,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_1^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_1\tilde{\boldsymbol{V}}_2\boldsymbol{a}_1\right\}\right.\right.$$

$$\left. + it_2\left\{\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_2\tilde{\boldsymbol{V}}_2) - s_{2,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_2^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_2\tilde{\boldsymbol{V}}_2\boldsymbol{a}_2\right\}\right]$$

$$\left.\times\exp\{\mathrm{tr}(\boldsymbol{M}_0^2)\}\left\{1 + \frac{4}{3\sqrt{N-p}}\,\mathrm{tr}(\boldsymbol{M}_0^3)\right\}h(\boldsymbol{V}_2)\right] + O((N-p)^{-1})$$

$$= \exp\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^2\}[1 + \mathrm{tr}\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^3\}]h(p\boldsymbol{\Omega}^*)$$

$$\times\exp\left[it_1\left\{\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_1\boldsymbol{\Omega}^*) - s_{1,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_1^\top\boldsymbol{\Omega}^*\boldsymbol{A}_1\boldsymbol{\Omega}^*\boldsymbol{a}_1\right\}\right.$$

$$\left. it_2\left\{\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_2\boldsymbol{\Omega}^*) - s_{2,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_2^\top\boldsymbol{\Omega}^*\boldsymbol{A}_2\boldsymbol{\Omega}^*\boldsymbol{a}_2\right\}\right]$$

$$\times\mathrm{E}[\exp\{\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}[1 + it_1a_{1,1}^* + it_2a_{2,1}^*$$

$$+ 2\,\mathrm{tr}\{\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*(p^{-1/2}\boldsymbol{\Xi}_0\boldsymbol{W}_2 + \boldsymbol{\Xi}_1\boldsymbol{\Omega}^*)\}]] + O_1.$$

From above result and and Lemma A.6,

$$\mathrm{E}[\exp\{\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}h^*(\boldsymbol{V}_2)]$$

$$= \left|\boldsymbol{I}_4 - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right|^{-p/2}\mathrm{E}[h^*(\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2)]$$

$$\times\exp\left\{-p^{1/2}\,\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{\Omega}^*) + p^{-1/2}\,\mathrm{tr}\left[\boldsymbol{\Omega}\boldsymbol{M}_0^*\left(\boldsymbol{I} - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right]\right\}$$

$$= \exp[\mathrm{tr}\{(\boldsymbol{I}_4 + 2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\}]\mathrm{E}[h^*(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)]$$

$$\times \left[1 + \frac{4}{3\sqrt{p}}\,\mathrm{tr}\{(\boldsymbol{I}_4 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\}\right] + O(p^{-1}),$$

where

$$\tilde{\boldsymbol{Z}}_2 \sim N_{p\times 4}\left(\boldsymbol{M}\left(\boldsymbol{I} - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}, \left(\boldsymbol{I} - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1} \otimes \boldsymbol{I}_p\right).$$

The moments are given by

$$\mathrm{E}[\mathrm{tr}\{\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*(p^{-1/2}\boldsymbol{\Xi}_0\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2 + \boldsymbol{\Xi}_1\boldsymbol{\Omega}^*)\}] = O(p^{-1}),$$

$$\mathrm{E}[a_{j,1}^*] = -\frac{p^2}{(N-p)^2}\mathrm{E}[s_{j,0,2}]\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j - \frac{p}{(N-p)^2}s_{j,0,0}\mathrm{E}[\boldsymbol{a}_j^\top\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j]$$

$$-\frac{p^{3/2}}{(N-p)^2}\mathrm{E}[s_{j,0,1}(\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j + \boldsymbol{a}_j^\top\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j)],$$

$$\mathrm{E}[s_{j,0,2}] = b_{j,0,0}^{-3}\frac{p}{(N-p)^2}\mathrm{E}[\{\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)^2\}],$$

$$\mathrm{E}[\{\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)^2\}] = 4\{\mathrm{tr}(\boldsymbol{B}_j\boldsymbol{M}_0^*)\}^2 + O(p^{-1/2}) = O(p^{-1/2}),$$

$$\mathrm{E}[\boldsymbol{a}_j^\top\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j] = \mathrm{E}[\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2)] = 4\,\mathrm{tr}(\boldsymbol{M}_0^*\boldsymbol{B}_j\boldsymbol{M}_0^*\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*) + O(1) = O(1),$$

$$\mathrm{E}[s_{j,0,1}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j] = \frac{\sqrt{p}}{N-p}b_{j,0,0}^{-2}\mathrm{E}[\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)\ \mathrm{tr}(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2)],$$

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)\ \mathrm{tr}(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2)] = 4\,\mathrm{tr}(\boldsymbol{B}_j\boldsymbol{M}_0^*)\ \mathrm{tr}(\boldsymbol{B}_j\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{M}_0^*) + O(1) = O(1),$$

where

$$\tilde{\boldsymbol{W}}_2 = \sqrt{p}\left(\frac{1}{p}\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2 - \boldsymbol{\Omega}^*\right).$$

Since $(\boldsymbol{I}_4 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3 = O_{1/2}$, we have the following expansion:

$$\mathrm{E}[\exp(i\boldsymbol{t}^\top\boldsymbol{D}_b)] = \exp\{i\boldsymbol{t}^\top\boldsymbol{\eta} - \boldsymbol{t}^\top\boldsymbol{\Lambda}\boldsymbol{t}/2\} + O_1, \tag{A.11}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$ and

$$\eta_j = \frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_j\boldsymbol{\Omega}^*) - s_{j,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j,$$

$$= \frac{n-1}{N-p}\left\{\Delta^2 + \frac{p}{N_2} - b\frac{p(n_1-1)}{n_1^2} + p\left(1 - b\left(1 + \frac{1}{n_1^2}\right)\right)\right\}$$

$$+\frac{bp^2(n_1-1)(n-1)}{n_1^3(N-p)^2+n_1(N-p)p(n_1-1)(n_1+1)}\left(1+\frac{2(n_1-1)^{1/2}}{n_1^{1/2}}+\frac{n_1-1}{n_1}\right)$$

$$=\frac{n-1}{N-p}\left\{\Delta^2+\frac{p}{N_2}-b\frac{p(n_1-1)}{n_1^2}+p\left(1-b\left(1+\frac{1}{n_1^2}\right)\right)\right\}+O_1,$$

$$\Lambda=\begin{pmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{21} & \lambda_{22},\end{pmatrix},\qquad \lambda_{ij}=2[\mathrm{tr}(\boldsymbol{\Xi}_{i,0}\boldsymbol{\Omega}^*\boldsymbol{\Xi}_{j,0}\boldsymbol{\Omega}^*)+\mathrm{tr}\{(\boldsymbol{I}_4+2p^{-1}\boldsymbol{\Omega})\boldsymbol{M}_{i,0}^*\boldsymbol{M}_{j,0}^*\}],$$

$$\lambda_{jj}=\mathrm{tr}\{(\boldsymbol{\Xi}_{j,0}\boldsymbol{\Omega}^*)^2\}+\mathrm{tr}\{(\boldsymbol{I}_4+2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_{j,0}^*)^2\}$$

$$=4\left(\frac{(n-1)^2N}{(N-p)^3}\right)\left(\Delta^2+\frac{p}{N_2}-b^2\frac{p}{n_1^2}+b^2\frac{p}{n_1}\right)+O_1,$$

$$\lambda_{12}=\mathrm{tr}(\boldsymbol{\Xi}_{1,0}\boldsymbol{\Omega}^*\boldsymbol{\Xi}_{2,0}\boldsymbol{\Omega}^*)+\mathrm{tr}((\boldsymbol{I}_4+2p^{-1}\boldsymbol{\Omega})\boldsymbol{M}_{1,0}^*\boldsymbol{M}_{2,0}^*)$$

$$=O_1.$$

A.6. Derivation of $\kappa(\Delta)$ in the case of d_F. In this section, we show that $\kappa(\Delta)$ is decided as (9) in the case of d_F and $c=0$. $d_F^{(-\lambda)}$ is estimator of d_F for the method II and is derived as

$$d_F^{(-\lambda)}(\boldsymbol{x}_k)=(\bar{\boldsymbol{x}}_1^{(-k,\lambda)}-\bar{\boldsymbol{x}}_2)^\top\{\boldsymbol{S}^{(-k,\lambda)}\}^{-1}\left\{\boldsymbol{x}-\frac{1}{2}(\bar{\boldsymbol{x}}_1^{(-k,\lambda)}+\bar{\boldsymbol{x}}_2)\right\}$$

$$=\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{U})-(1-\lambda)T_\lambda^{-1}\boldsymbol{a}^\top\boldsymbol{U}\boldsymbol{A}_\lambda\boldsymbol{U}\boldsymbol{a},$$

$$T_\lambda=\frac{N_1^{(\lambda)}}{N_1^{(1)}}+(1-\lambda)\,\mathrm{tr}(\boldsymbol{B}\boldsymbol{U}),$$

where $\bar{\boldsymbol{x}}_1^{(-k,\lambda)}$ and $\boldsymbol{S}^{(-k,\lambda)}$ are given by (8), and $\boldsymbol{U}$ is the same as $\boldsymbol{U}$ in Lemma 1, and $\boldsymbol{a}=(0,n_1^{-1/2},1)^\top$, $\boldsymbol{B}=\boldsymbol{a}\boldsymbol{a}^\top$ and

$$\boldsymbol{A}_\lambda=\frac{N^{(-\lambda)}}{2}\begin{pmatrix} N_2^{-1} & 0 & -N_2^{-1/2}\\ 0 & -n_1\{N_1^{(-\lambda)}\}^{-2} & n_1^{3/2}\{N_1^{(-\lambda)}\}^{-2}\\ -N_2^{-1/2} & n_1^{3/2}\{N_1^{(\lambda)}\}^{-2} & \frac{1-\lambda}{N_1^{(-\lambda)}}\left(2-\frac{1-\lambda}{N_1^{(-\lambda)}}\right)\end{pmatrix}.$$

Put $\boldsymbol{W}_1=\sqrt{N-p}\{(N-p)^{-1}\boldsymbol{V}_1-\boldsymbol{I}_3\}$. From $1-\kappa/N$, we have an expansion of $d_F^{(-\lambda)}(\boldsymbol{x})$ as

$$T_\lambda=\frac{N_1^{(-\lambda)}}{N_1^{(-1)}}+O_p(N^{-1}),$$

$$\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{U})-(1-\lambda)T_\lambda^{-1}\boldsymbol{a}^\top\boldsymbol{U}\boldsymbol{A}_\lambda\boldsymbol{U}\boldsymbol{a}=\frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_\lambda\tilde{\boldsymbol{V}}_2)+a_0+a_1+O_p((N-p)^{-1}),$$

$$a_0 = -\frac{p}{(N-p)^{3/2}} \operatorname{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}} \boldsymbol{W}_1),$$

$$a_1 = \frac{p}{(N-p)^2} \operatorname{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}} \boldsymbol{W}_1^2) + \frac{\kappa}{N} \frac{p^2}{(N-p)^2} \boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a},$$

where $\tilde{\boldsymbol{T}} = p^{-1/2}\boldsymbol{T}$ and $\tilde{\boldsymbol{V}}_2 = \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^\top$. Then, the characteristic function of $d_F^{(\lambda)}(\boldsymbol{x})$ is expressed as

$$\mathrm{E}[e^{itd_F^{(\lambda)}(\boldsymbol{x})}] = \mathrm{E}\left[\mathrm{E}\left[\exp\left\{it\left(\frac{p}{N-p}\operatorname{tr}(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2) + a_0 + a_1\right) + O_p((N-p)^{-1})\right\} \,\middle|\, \boldsymbol{V}_2\right]\right],$$

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{it\frac{p}{N-p}\operatorname{tr}(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2) + a_0 + a_1 + O_p((N-p)^{-1})\right\} \,\middle|\, \boldsymbol{V}_2\right] \\
&\quad = \exp\left\{it\frac{p}{N-p}\operatorname{tr}(\boldsymbol{A}_\lambda \boldsymbol{V}_2)\right\}\mathrm{E}[\{e^{ita_0}(1+ita_1)\} \mid \boldsymbol{V}_2] + O_p((N-p)^{-1}).
\end{aligned}$$

Put $\boldsymbol{M}_0 = itp(N-p)^{-3/2}\boldsymbol{T}^\top \boldsymbol{A}_\lambda \boldsymbol{T}$. From Lemma A.6,

$$\begin{aligned}
&\mathrm{E}[\{e^{ita_0}(1+ita_1)\} \mid \boldsymbol{V}_2] \\
&\quad = \left|\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right|^{-(N-p)/2} \exp\{-\sqrt{N-p}\operatorname{tr}(\boldsymbol{M}_0)\}\mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1)] \\
&\quad = \exp\{\operatorname{tr}(\boldsymbol{M}_0^2)\}\left\{1 + \frac{4}{3\sqrt{N-p}}\operatorname{tr}(\boldsymbol{M}_0^3)\right\}\mathrm{E}[g(\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1)] + O_p((N-p)^{-1}),
\end{aligned}$$

where

$$\tilde{\boldsymbol{Z}}_1 \sim N_{(N-p)\times 3}\left(\boldsymbol{O}, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1} \otimes \boldsymbol{I}_{N-p}\right),$$

$$\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 \sim W_3\left(N-p, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1}\right)$$

are independent of $\boldsymbol{V}_2$, and $\boldsymbol{M}_0 = -itp(N-p)^{-3/2}\tilde{\boldsymbol{T}}^\top \boldsymbol{A}\tilde{\boldsymbol{T}}$ and $g(\boldsymbol{V}_1) = 1 + ita_1$. The moments are given by

$$\begin{aligned}
\mathrm{E}[a_1 | \boldsymbol{V}_2] &= \frac{p}{N-p}\mathrm{E}\left[\operatorname{tr}\left\{\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_3\right)^2\right\}\right] \\
&\quad + \frac{\kappa}{N}\frac{p^2}{(N-p)^2}\boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a}
\end{aligned}$$

$$= \frac{1}{N-p}[4\ \mathrm{tr}(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2) + 3\ \mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}} \boldsymbol{M}_0)]$$

$$+ \frac{\kappa}{N} \frac{p^2}{(N-p)^2} \boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a} + O_p((N-p)^{-1})$$

$$= \frac{1}{N-p}\left[4\ \mathrm{tr}(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2) - 3it\frac{p}{(N-p)^{3/2}}\ \mathrm{tr}\{(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2)^2\}\right]$$

$$+ \frac{\kappa}{N} \frac{p^2}{(N-p)^2} \boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a} + O_p((N-p)^{-1}).$$

Secondly, let $\boldsymbol{W}_2 = \sqrt{p}(p^{-1}\boldsymbol{V}_2 - \boldsymbol{\Omega}^*)$, then $\boldsymbol{W}_2 = O_p(1)$ from the central limit theorem where $\boldsymbol{\Omega}^* = \boldsymbol{I}_3 + p^{-1}\boldsymbol{\Omega}$. We can obtain the following expansions:

$$\boldsymbol{V}_2 = p(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2),$$

$$\mathrm{tr}(\boldsymbol{M}_0) = itp(N-p)^{-3/2}\ \mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2),$$

$$\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) = \mathrm{tr}(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2))$$

$$= \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*) + p^{-1/2}\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2),$$

$$\mathrm{tr}(\boldsymbol{M}_0^2) = (it)^2 p^2 (N-p)^{-3}\ \mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^2\},$$

$$\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^2\} = \mathrm{tr}\{(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2))^2\}$$

$$= \mathrm{tr}\{(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*)^2\} + 2p^{-1/2}\ \mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*\boldsymbol{A}_\lambda\boldsymbol{W}_2) + O_p(1),$$

$$\boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a} = \boldsymbol{a}^\top \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{a} + O_p(1) = O_p(1),$$

$$\mathrm{tr}(\boldsymbol{M}_0^3) = -(it)^3 p^3 (N-p)^{-9/2}\ \mathrm{tr}\{(\boldsymbol{A}_\lambda\tilde{\boldsymbol{V}}_2)^3\},$$

$$\mathrm{tr}\{(\boldsymbol{A}_\lambda\tilde{\boldsymbol{V}}_2)^3\} = \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^3\} + O_p(p^{-1/2}) = O_{1/2}.$$

Since $\boldsymbol{V}_2 \sim W_3(p, \boldsymbol{I}_3, \boldsymbol{\Omega})$, we obtain the following expansions:

$$\exp\left\{it\frac{p}{N-p}\ \mathrm{tr}(\boldsymbol{A}_\lambda\tilde{\boldsymbol{V}}_2) + \mathrm{tr}(\boldsymbol{M}_0^2)\right\}$$

$$= \exp\left\{it\frac{p}{N-p}\ \mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*) + it\frac{p^{1/2}}{N-p}\ \mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{W}_2)\right.$$

$$\left. + (it)^2 p^2 (N-p)^{-3}\ \mathrm{tr}\{(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*)^2\}\right\}$$

$$\times \exp\{2(it)^2 p^{3/2}(N-p)^{-3}\ \mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*\boldsymbol{A}_\lambda\boldsymbol{W}_2) + O_p((N-p)^{-1})\},$$

$$\exp\{2(it)^2 p^{3/2}(N-p)^{-3} \operatorname{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{W}_2) + O_p((N-p)^{-1})\}$$
$$= 1 + 2(it)^2 p^{3/2}(N-p)^{-3} \operatorname{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{W}_2) + O_1.$$

Put $\boldsymbol{M}_0^* = itp^{3/2}(N-p)^{-3}\boldsymbol{A}_\lambda$. From Lemma A.6, we can have

$$\mathrm{E}[\exp\{\operatorname{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}h(\boldsymbol{Z}_2^\top \boldsymbol{Z}_2)]$$
$$= \left|\boldsymbol{I}_3 - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right|^{-p/2} \mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)]$$
$$\times \exp\left[-p^{1/2}\operatorname{tr}(\boldsymbol{M}_0^*\boldsymbol{\Omega}^*) + p^{-1/2}\operatorname{tr}\left\{\boldsymbol{\Omega}\boldsymbol{M}_0^*\left(\boldsymbol{I}_3 - \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right\}\right]$$
$$= \exp[\operatorname{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\}]\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)]$$
$$\times \left[1 + \frac{4}{3\sqrt{p}}\operatorname{tr}\{(\boldsymbol{I}_3 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\}\right] + O_1.$$

Moreover, since $\operatorname{tr}\{(\boldsymbol{I}_3 + 3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\} = O_{1/2}$,

$$\mathrm{E}[\exp\{\operatorname{tr}(\boldsymbol{M}_0^*\boldsymbol{W}_2)\}h(\boldsymbol{Z}_2^\top \boldsymbol{Z}_2)]$$
$$= \exp[\operatorname{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\}]\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)] + O_1,$$

where $h(\boldsymbol{Z}_2^\top \boldsymbol{Z}_2) = (1 + 2(it)^2 p^{-1/2}(N-p)^{-1}\operatorname{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{W}_2))$, and $\boldsymbol{Z}_1$ and $\tilde{\boldsymbol{Z}}$ are the random matrices that satisfy

$$\boldsymbol{V}_2 = \boldsymbol{Z}_2^\top \boldsymbol{Z}_2,$$
$$\boldsymbol{Z}_2 \sim N_{p\times 3}(\boldsymbol{M}, \boldsymbol{I}_3 \otimes \boldsymbol{I}_p),$$
$$\tilde{\boldsymbol{Z}}_2 \sim N_{p\times 3}(\boldsymbol{M}(\boldsymbol{I}_3 - 2p^{-1/2}\boldsymbol{M}_0^*)^{-1}, (\boldsymbol{I}_3 - 2p^{-1/2}\boldsymbol{M}_0^*)^{-1} \otimes \boldsymbol{I}_p).$$

The moments are given by

$$\mathrm{E}[h(\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2)] = 1 + 2(it)^2 p^{-1/2}(N-p)^{-1}\operatorname{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \mathrm{E}[\tilde{\boldsymbol{W}}_2]),$$
$$\mathrm{E}[\boldsymbol{W}_2] = \sqrt{p}\left\{\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right.$$
$$\left. + p^{-1}\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\boldsymbol{\Omega}\left(\boldsymbol{I}_3 + \frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right\} - \boldsymbol{\Omega}^* = O_{1/2},$$

where

$$\tilde{\boldsymbol{W}}_2 = \sqrt{p}\left(\frac{1}{p}\tilde{\boldsymbol{Z}}_2^\top \tilde{\boldsymbol{Z}}_2 - \boldsymbol{\Omega}^*\right).$$

From the above result, we have

$$\begin{aligned}\eta_\lambda &= \frac{p}{N-p}\,\mathrm{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^*)\\ &= \frac{N^{(-\lambda)}}{2(N-p)}\left(\Delta^2 + \frac{p}{N_2} - \frac{n_1 p}{\{N_1^{(-\lambda)}\}^2} - \frac{(1-\lambda)^2 p}{\{N_1^{(-\lambda)}\}^2} + \frac{2(1-\lambda)p}{N_1^{(-\lambda)}}\right),\\ s_\lambda^2 &= 2\left[p^2(N-p)^{-3}\,\mathrm{tr}\{(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*)^2\} + \frac{p}{(N-p)^2}\,\mathrm{tr}\{(\boldsymbol{I}_3 + 2p^{-1}\boldsymbol{\Omega})\boldsymbol{A}_\lambda^2\}\right]\\ &= \frac{\{N^{(-\lambda)}\}^2 N}{(N-p)^3}\left(\Delta^2 + \frac{p n_1^3}{\{N_1^{(-\lambda)}\}^4} + \frac{p}{N_2}\right).\end{aligned}$$

Therefore, we have the characteristic function $\phi(t)$ of $d_F^{(-\lambda)}(\boldsymbol{x})$ as

$$\phi(t) = \exp(it\eta_\lambda - t^2 s_\lambda^2/2) + O_1.$$

By using the inversion formula, we have

$$\Pr(d_F^{(-1;\lambda)}(\boldsymbol{x}_{i1}) \le 0) = \Phi(-s_\lambda^{-1}\eta_\lambda) + O_1.$$

From Theorem 3, the probability of misclassification $P(2|1)$ of d_F is given as

$$\begin{aligned}&\Pr(d_F^{(-1)}(\boldsymbol{x}) \le 0 \mid x \in \Pi_1)\\ &\quad = \Phi\left(-\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2}\right) + O_1.\end{aligned}$$

Since $\lambda = 1 - \kappa/N$,

$$\begin{aligned}s_\lambda^{-1}\eta_\lambda = &-\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2}\\ &+ \frac{1}{4}\frac{p}{n_1 N_1}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2}\\ &\times\left\{2 - \left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\right\}\end{aligned}$$

$$- \frac{\kappa p}{n_1 N}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{n_1} + \frac{p}{N_2}\right)^{-1/2} + O_2$$

$$= -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2} + O_2.$$

Therefore, κ is given as (9).

References

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

[2] B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation, J. Amer. Statist. Assoc., **78** (1983), 316–331.

[3] B. Efron and R. Tibshirani, Improvement on cross-validation: The .632+ Bootstrap method, J. Amer. Statist. Assoc., **92** (1997), 548–560.

[4] R. A. Fisher, The use of multiple measurements in taxonomic problems, The Annals of Human Genetics, **7** (1936), 111–132.

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

[6] Y. Fujikoshi, V. V. Ulyanv and R. Shimizu, Multivariate statistics: High-dimensional and large-sample approximations, Wiley, Hoboken, NJ, 2010.

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

[8] A. K. Gupta and D. K. Nagar, Matrix Variate Distributions, Chapman & Hall, 2000.

[9] P. A. Lachenbruch and M. R. Mickey, Estimation of error rates in discriminant analysis, Technometrics, **10** (1968), 1–11.

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

[11] M. Okamoto, An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Statist., **34** (1963), 1286–301.

[12] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. R. Statist. Soc., **B36** (1974), 111–147.

[13] T. Tonda, T. Nakagawa and H. Wakaki, EPMC estimation in discriminant analysis when the dimension and sample sizes are large, Hiroshima Math. J., **47** (2017), No. 1, 43–62.

[14] H. Yanagihara and H. Fujisawa, Iterative bias correction of the cross-validation criterion, Scand. J. Stat., **39** (2012), 116–130.

[15] H. Yanagihara, T. Tonda and C. Matsumoto, Bias correction of cross-validation criterion based on Kullback-Laeibler information under a general condition, J. Multivariate Anal., **97** (2006), 1965–1975.

[16] H. Yanagihara, K.-H. Yuan, H. Fujisawa and K. Hayashi, A class of cross-validatory model selection criteria, Hiroshima Math. J., **43** (2013), 149–177.

Tomoyuki Nakagawa
Depertment of Information Sciences
Faculty of Science & Technology
Tokyo University of Science
Noda-shi Chiba 278-8510 *Japan*
E-mail: nakagawa.stat@gmail.com