広島大学学位請求論文

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

（高次元大標本の場合での CV を用いた誤判別確率の推定）

２０１８年

広島大学大学院理学研究科
数学専攻

中川　智之

目　次

１．主論文

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

(高次元大標本の場合での CV を用いた誤判別確率の推定)

中川　智之

Hiroshima Mathematical Journal, 出版予定.

２．参考論文

(1) EPMC estimation in discriminant analysis when the dimension and sample sizes are large.

Tetsuji Tonda, Tomoyuki Nakagawa and Hirofumi Wakaki

Hiroshima Mathematical Journal, **47**(1)(2017)43-62.

(2) Selection of the linear and the quadratic discriminant functions when the difference between two covariance matrices is small.

Tomoyuki Nakagawa and Hirofumi Wakaki

Journal of the Japan Statistical Society, **47**(2)(2017)145-165.

主論文

Hiroshima Math. J.
00 (0000), 1–39

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

Tomoyuki Nakagawa

(Received Xxx 00, 0000)

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 unknown 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 for a classification rule constructed from a training data. The probabilities of misclassification 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 $k = 1, 2$, the training data $\boldsymbol{X}_k = (\boldsymbol{x}_{k1}, \dots, \boldsymbol{x}_{kN_k})^\top$ 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. For observation $\boldsymbol{x}$, the statistician wishes to estimate the probabilities of misclassification for a classification rule. In this paper, we consider the following classification rule using the discriminant function $d_{\boldsymbol{X}}(\boldsymbol{x}) = d(\boldsymbol{x})$. It is to classify $\boldsymbol{x}$ as coming from Π_1 if $d(\boldsymbol{x}) > c$ and from Π_2 if $d(\boldsymbol{x}) \le c$, where c is a cut-off point. 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). As another sample, we consider the following

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

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

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 is introduce in Fujikoshi and Seo (1998) and includes various discriminant functions, for example D_b is the same as d_F when $b = 1$. Then, by using the discriminant function d, the probabilities of misclassification are given by

$$P(2|1) = \Pr\{d(\boldsymbol{x}) \le c \mid \boldsymbol{x} \in \Pi_1\},$$
$$P(1|2) = \Pr\{d(\boldsymbol{x}) > c \mid \boldsymbol{x} \in \Pi_2\}.$$

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 make accurate discrimination. So, we want to obtain the probabilities of misclassification. 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 observation. Estimation methods of the probabilities of misclassification are separated to the parametric and the non-parametric methods. In the parametric methods, we assume a distribution and a classification rule and derives an approximation formula of the probabilities of misclassification. It is 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. Hence, an approximation formula needs to be derived for each assumption. On the other hand, estimators of the probabilities of misclassification are used for the Cross-validation (CV) 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). In 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 known well in the high-dimensional case. Hence, we investigate asymptotic properties of CV when the dimension and the sample sizes tend to be 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 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:

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

where $N = N_1 + N_2$. Regarding the estimation of the probabilities of misclassification, it is also 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 have been observed in recently. Therefore we consider an asymptotic theory based on the high-dimensional (HD) framework:

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

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

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

Firstly, we consider the bias of the estimator by CV. The estimator $\hat{P}_{CV}$ of the probability of misclassification using CV expresses as

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

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

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

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

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

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 order of its bias is larger than the LS framework. For the classification with d_F, the following theorem is given in Tonda *et al.* (2017).

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

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

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

$$\begin{aligned}\nu &= \nu\left(\Delta^2\right) \\ &= -\frac{1}{2}\left(\frac{N-p}{N-1}\right)^{1/2}\left\{\Delta^2 + \frac{(N_1-N_2)(p-1)}{N_1N_2}\right\}\left\{\Delta^2 + \frac{N(p-1)}{N_1N_2}\right\}^{-1/2}.\end{aligned}$$

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

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

where $m = N_1N_2/N$ *and*

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

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

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

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

Therefore, we obtain the following corollary.

COROLLARY 1. *In the case of classification with* 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

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

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

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

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 the 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:*

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

where

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

η and s are given in the appendix.

The proof is given in Fujikoshi and Seo (1998) and Fujikoshi (2000). In this paper, we can show the different way of the proof in the appendix. This theorem means that the estimator of the probabilities of misclassification by using CV is an asymptotic unbiased estimator in the case of classification with 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:*

$$\bar{\boldsymbol{x}}_k = \frac{n_1}{N_1}\bar{\boldsymbol{x}}_k^{(-i)} + \frac{1}{N_1}\boldsymbol{x}_{ki},$$
$$n_1\boldsymbol{S}_1 = (n_1 - 1)\boldsymbol{S}_1^{(-i)} + \frac{n_1 - 1}{n_1}\left(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)}\right)\left(\boldsymbol{x} - \bar{\boldsymbol{x}}_1^{(-i)}\right)^\top,$$

for $k=1,2$. Moreover,

$$
\begin{aligned}
n\boldsymbol{S} &= (n-1)\boldsymbol{S}^{(-i)} + \frac{n_1-1}{n_1}\left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(-i)}\right)\left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(i)}\right)^\top, \\
\boldsymbol{S}^{-1} &= \frac{n}{n-1}\left[\left\{\boldsymbol{S}^{(-i)}\right\}^{-1} - T^{-1}\left\{\boldsymbol{S}^{(-i)}\right\}^{-1}\left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(-i)}\right)\left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(-i)}\right)^\top\left\{\boldsymbol{S}^{(-i)}\right\}^{-1}\right], \\
T &= \frac{N_1(n-1)}{n_1} + \left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(-i)}\right)^\top\left\{\boldsymbol{S}^{(-i)}\right\}^{-1}\left(\boldsymbol{x}-\bar{\boldsymbol{x}}_1^{(-i)}\right),
\end{aligned}
$$

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 $k=1$ express as

$$
\begin{aligned}
\bar{\boldsymbol{x}}_1^{(-i)} &= n_1^{-1}\sum_{j\neq i}^{N_1} x_{1j}, \\
\boldsymbol{S}_1^{(-i)} &= (N_1-2)^{-1}\sum_{j\neq i}^{N_1}\left(\boldsymbol{x}_{1j}-\bar{\boldsymbol{x}}_1^{(-j)}\right)\left(\boldsymbol{x}_{1j}-\bar{\boldsymbol{x}}_1^{(-j)}\right)^\top, \\
\boldsymbol{S}^{(-i)} &= (N_1-2)\boldsymbol{S}_1^{(-i)} + n_2\boldsymbol{S}_2.
\end{aligned}
$$

It is easy to proof of Lemma 2 so that we leave out its 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*

$$
\begin{aligned}
D_b^{(-i)}(\boldsymbol{x}_{1i}) &= \mathrm{tr}(\boldsymbol{A}_1\boldsymbol{U}) - T_1^{-1}\boldsymbol{a}_1^\top\boldsymbol{U}\boldsymbol{A}_1\boldsymbol{U}\boldsymbol{a}_1, \qquad (6)\\
T_1 &= \frac{N_1-1}{N_1-2} + \mathrm{tr}(\boldsymbol{B}_1\boldsymbol{U}), \\
D_b^{(-j)}(\boldsymbol{x}_{1j}) &= \mathrm{tr}(\boldsymbol{A}_2\boldsymbol{U}) - T_2^{-1}\boldsymbol{a}_2^\top\boldsymbol{U}\boldsymbol{A}_2\boldsymbol{U}\boldsymbol{a}_2, \qquad (7)\\
T_2 &= \frac{N_1-1}{N_1-2} + \mathrm{tr}(\boldsymbol{B}_2\boldsymbol{U}),
\end{aligned}
$$

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 by the appendix. Using this lemma, we obtain the following theorem.

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

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

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

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

Therefore, it is used the inversion formula

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

From this formula,

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

Since Theorem 3,

$$\Pr\left(D_b^{(-1)}(\boldsymbol{x}_{11}) \le c\right)^2 = \Phi\left(\lambda^{-1}(c-\eta)\right)^2 + O_1$$

Therefore, we complete the proof of this theorem.

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

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

Therefore, Theorems 4 and 5 mean that MSE of CV is the same order as MSE 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. The previous section, we showed that if the sample sizes are sufficiently large, the estimator of CV is good estimator even for the high-dimension. However, the bias of CV estimator is large for the small sample sizes and increases with the dimension. Therefore, it is necessary to correct the bias of CV estimator in the HD framework.

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

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 consider using this idea to the estimation of the probabilities of misclassification. The leave-two-out CV is expressed by

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

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

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

Therefore, a new estimator is given by

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

Then it holds that

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

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 estimating the probabilities of misclassification. Suppose that $F_{N-1}^{(-i)}$ and F_i are the empirical distributions of $\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1i-1}, \boldsymbol{x}_{1i+1} \ldots, \boldsymbol{x}_{1N_1}$ 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, MLE 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

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

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

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

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

This method is called leave-λ-out CV in this paper. Let $\lambda = 1 - \kappa/N$, and if we obtain an expansion by

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

where $Q_0^*(x_1,x_2,x_3)$ and $Q_1^*(x_1,x_2,x_3)$ are C^1 class functions around $(p/N_1, p/N_2, 1)$. Then, it is hold that

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

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

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

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:*

$$
\begin{aligned}
&\lambda = 1 - \kappa(\Delta)/N, \\
&\kappa(\Delta) = \frac{N}{4N_1}\left\{2 - \left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\right\}
\end{aligned}
\tag{9}
$$

A derivation of this κ is given in the appendix.

This method has the same calculation load as CV and can correct the bias of CV. On the other hand, we must derive λ for correcting the bias.

3.3. Method III : Modified a cutoff point. We propose a method for correcting the bias by modifying a cut-off point c.

$$
\begin{aligned}
P(2|1) &= \Pr(d(\boldsymbol{x}) \le c + c_1/N | x \in \Pi_1) \\
&= Q_0^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c + \frac{c_1}{N}\right) + Q_1^\dagger\left(\frac{p}{N_1}, \frac{p}{N_2}, c + \frac{c_1}{N}\right) + O_2,
\end{aligned}
$$

where $Q_0^\dagger(x_1,x_2,x_3)$ and $Q_1^\dagger(x_1,x_2,x_3)$ are C^1 class functions around $(p/N_1, p/N_2, c)$.

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

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

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

From Theorem 3, we have the following.

EXAMPLE 3. *In the case of D_b, we can have c_1 as follows;*

$$\begin{aligned}\eta^{(1)} &= \frac{n-1}{N-p-1}\left(\Delta^2+\frac{p}{N_2}-\frac{bp}{n_1}+p(1-b)\right) = \eta+\eta_1+O_2\\(\lambda^{(1)})^2 &= 4\frac{(n-1)^2(N-1)}{(N-p-1)^3}\left(\Delta^2+\frac{pb^2}{n_1}+\frac{p}{N_2}\right)\\&= \lambda^2+\lambda_1+O_2\\c_1(\Delta) &= \frac{N}{\lambda}\left\{\frac{\lambda_1}{2}(c-\eta)-\lambda\eta_1\right\}.\end{aligned}$$

where η and λ^2 are given by Theorem 3

$$\begin{aligned}\eta_1 &= \left(\frac{1}{N-p}+\frac{n}{(N-p)^2}\right)\left(\Delta^2+\frac{p}{N_2}-\frac{bp}{N_1}+p(1-b)\right)-\frac{bnp}{(N-p)N_1^2},\\\lambda_1 &= 4\frac{Nn^2}{(N-p)^3}\left(\frac{3}{N-p}-\frac{2}{n}-\frac{1}{N}\right)\left(\Delta^2+\frac{pb^2}{N_1}+\frac{p}{N_2}\right)+4\frac{pb^2Nn^2}{N_1^2(N-p)^3}.\end{aligned}$$

This method have the same the calculation load as CV and can correct the bias of CV. On the other hand, we must derive c_1.

4. Numerical study

In this section, we investigate performances of CV and the three methods for the classification rule with d_F by the Monte Carlo method. Without loss of generality, we can assume that $\boldsymbol{\mu}_1 = \Delta(1,...,1)'/2\sqrt{p}$, $\boldsymbol{\mu}_2 = -\Delta(1,...,1)'/2\sqrt{p}$ 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 Δ were $N_1, N_2 = 15, 20, 25, 30, 35$, $p/N = 1/5, 3/5$ and $\Delta = 1.05, 1.68, 2.56, 3.29$ satisfying $N-p-2>0$. The values of Δ correspond to the values 0.30, 0.20, 0.10, 0.05 of $\Phi(-\Delta/2)$, respectively. an estimator of Δ is necessary to use the methods II and III, so that $\hat{\Delta}^2$ was given by

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

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 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. Furthermore, 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. While the method I can be applied to many cases, its MSE is larger than MSE of other methods. On the other hand, while MSEs of the methods II and III are the same as CV, it is necessary to derive the parameters κ and c_1. We consider that CV is better than other methods if the sample sizes are sufficiently large. The method I makes the bias smaller than CV without assumptions, it is a good method if only bias correction is considered. On the other hand, the methods II, III are better than other methods if we can derive the optimal value of κ and d. Moreover, when the sample sizes are small, we consider 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.

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

Table 2. (MSE of estimators) $\times 100$

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	0460	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

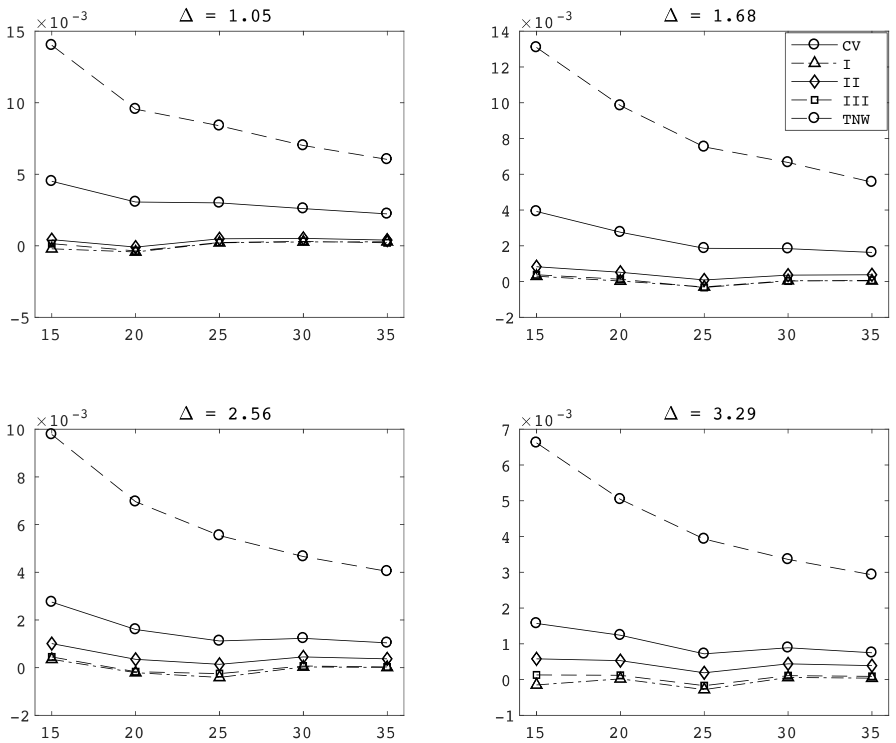

Figure 1. The figures plot the biases of the estimators for each Δ in the case of $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

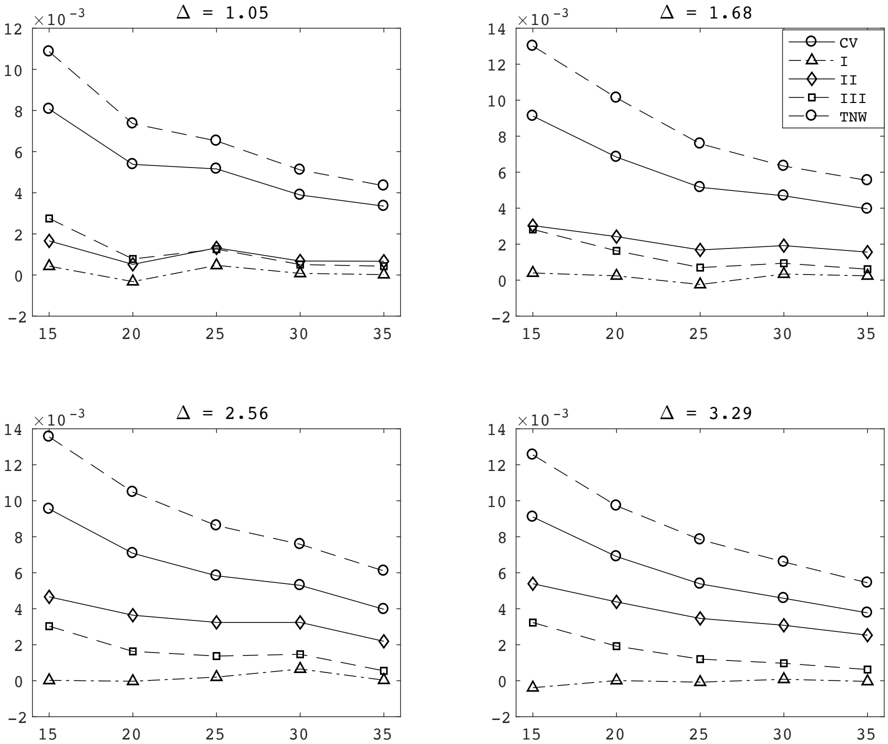

FIGURE 2. The figures plot the biases of the estimators for each Δ in the case of $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively.

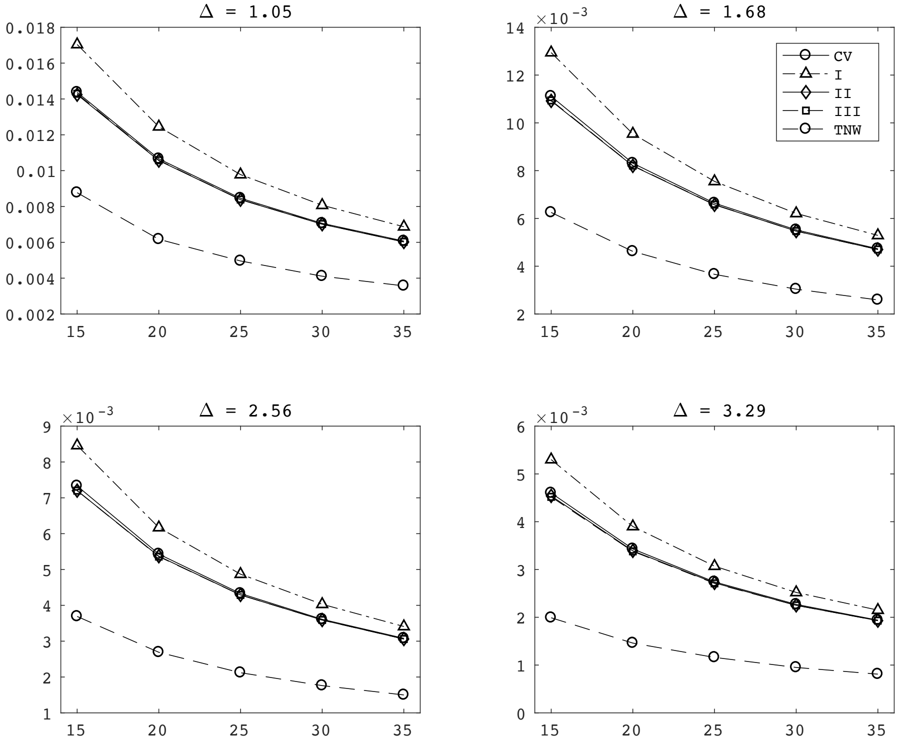

FIGURE 3. The figures plot MSEs of the estimators for each Δ in the case of $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively

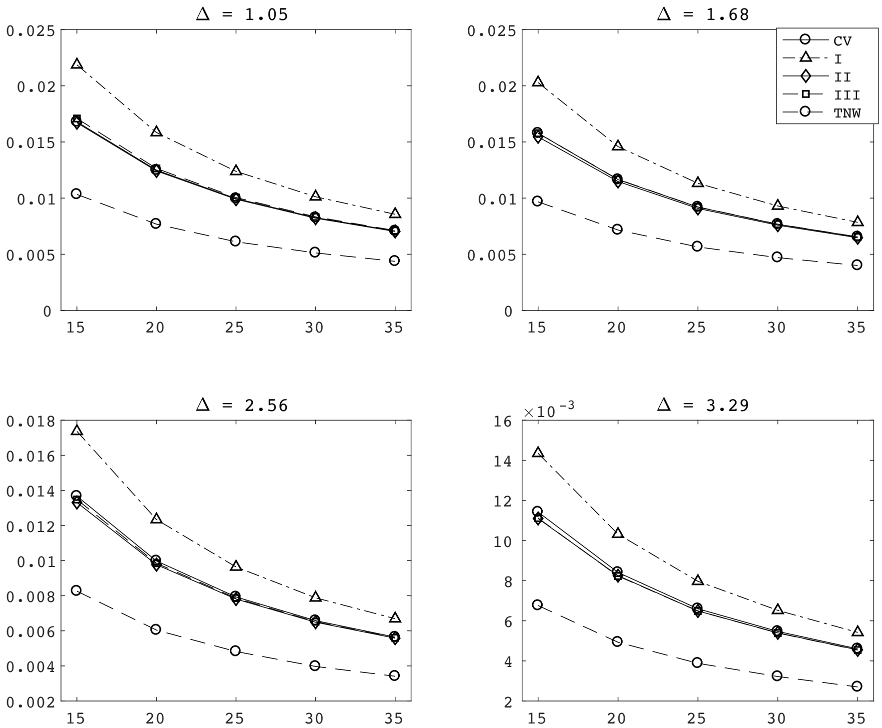

FIGURE 4. The figures plot MSEs of the estimators for each Δ in the case of $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively

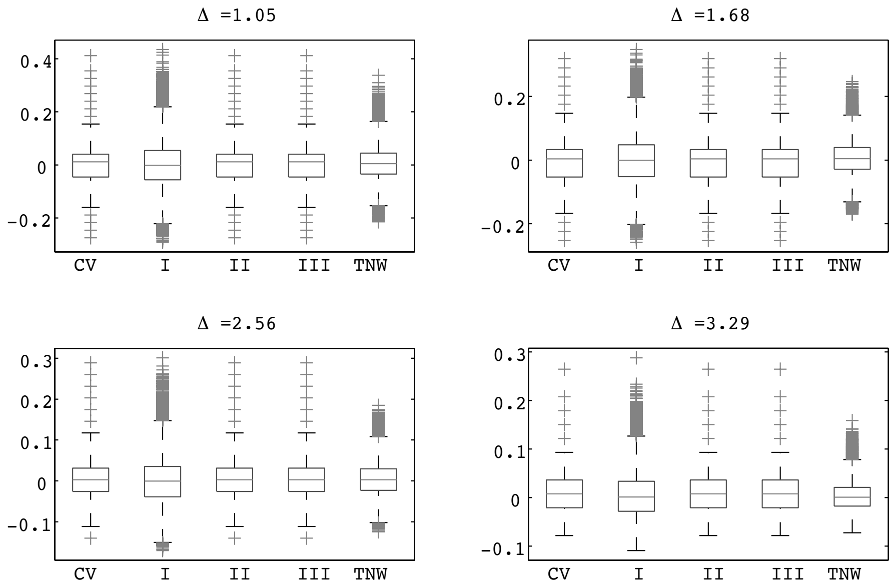

FIGURE 5. The figures are the boxplots of $\hat{P}(2|1) - P(2|1)$ for each Δ in the case of $N_1 = 35$ and $p/N = 1/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively

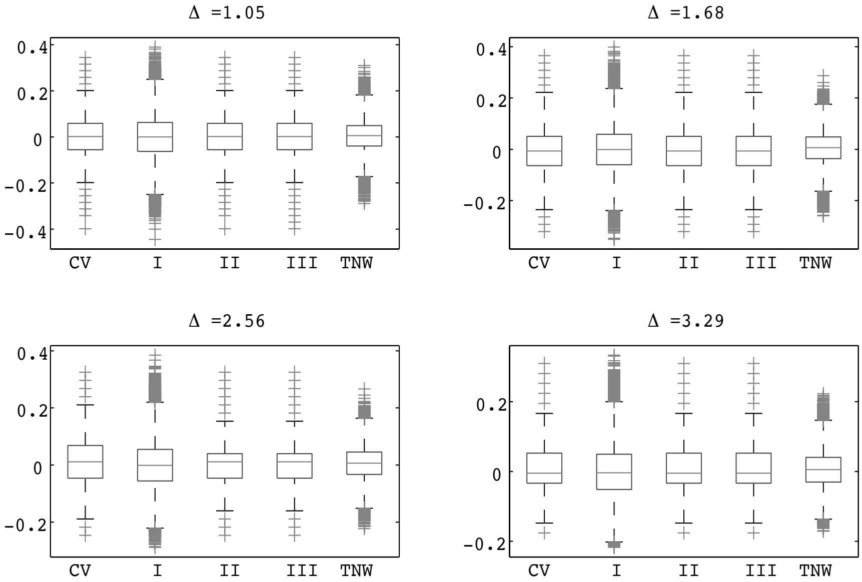

FIGURE 6. The figures are the boxplot of $\hat{P}(2|1) - P(2|1)$ for each Δ in the case of $N_1 = 35$ and $p/N = 3/5$. CV, I, II, III, and TNW indicate the cross-validation, the methods I, II, III in section 3, and the estimator in Tonda *et al.* (2017), respectively

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,*

$$\begin{aligned}
&\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{Z}^\top \boldsymbol{Z})] = \mathrm{tr}\left\{\boldsymbol{A}\left(n\boldsymbol{\Sigma} + \boldsymbol{M}^\top \boldsymbol{M}\right)\right\}, \\
&\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{B}\boldsymbol{Z}^\top \boldsymbol{Z})] = n\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + n(n+1)\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}) \\
&\qquad + (n+1)\mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{B}\boldsymbol{\Sigma}) + (n+1)\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}) \\
&\qquad + \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}) + \mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + \mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}), \\
&\mathrm{E}[\mathrm{tr}(\boldsymbol{A}\boldsymbol{Z}^\top \boldsymbol{Z})\mathrm{tr}(\boldsymbol{B}\boldsymbol{Z}^\top \boldsymbol{Z})] = n^2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + 2n\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}) \\
&\qquad + n\mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M})\mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + n\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}) \\
&\qquad + 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}) + 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{B}\boldsymbol{\Sigma}) + \mathrm{tr}(\boldsymbol{A}\boldsymbol{M}^\top \boldsymbol{M})\mathrm{tr}(\boldsymbol{B}\boldsymbol{M}^\top \boldsymbol{M}).
\end{aligned}$$

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,*

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

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

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

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

$$\begin{aligned}
&\boldsymbol{u} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),\\
&\boldsymbol{W} = n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2} \sim W_p(n, \boldsymbol{I}_p),\\
&\boldsymbol{z}_1 = \sqrt{N_1}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),\\
&\boldsymbol{z}_2 = \sqrt{N_2}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1) \sim N_p(\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{I}_p),
\end{aligned}$$

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

$$\begin{aligned}
&\boldsymbol{V}_2 = \boldsymbol{Q}^\top\boldsymbol{Q} \sim W_3(p, \boldsymbol{I}_3, \Omega),\\
&\boldsymbol{V}_1 = \boldsymbol{T}(\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q})^{-1}\boldsymbol{T}^\top \sim W_3(N-p, \boldsymbol{I}_3),\\
&\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top.
\end{aligned}$$

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. Let

$$\boldsymbol{W}_1 = \sqrt{N-p}\left(\frac{1}{N-p}\boldsymbol{V}_1 - \boldsymbol{I}_3\right) = O_p(1).$$

From Lemma 1,

$$\begin{aligned}
\boldsymbol{U} &= \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top\\
&= \frac{p}{N-p}\left\{\tilde{\boldsymbol{V}}_2 - (N-p)^{-1/2}\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + (N-p)^{-1}\tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top\right\} + O_p((N-p)^{3/2}),\\
\mathrm{tr}(\boldsymbol{A}\boldsymbol{U}) &= \frac{p}{N-p}\left\{\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + a_0 + a_1\right\} + O_p((N-p)^{-1}),
\end{aligned}$$

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

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

Then it can be expanded as

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

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,

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

where

$$\begin{aligned}
\tilde{\boldsymbol{Z}}_1 &\sim N_{(N-p)\times 3}\left(\boldsymbol{O}, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1} \otimes \boldsymbol{I}_{N-p}\right), \\
\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 &\sim W_3\left(N-p, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1}\right),
\end{aligned}$$

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

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

Secondly, let

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

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:

$$
\begin{aligned}
&\boldsymbol{V}_2 = p\left(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2\right),\\
&\mathrm{tr}(\boldsymbol{M}_0) = itp(N-p)^{-3/2}\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2),\\
&\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) = \mathrm{tr}(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2))\\
&\qquad = \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Omega}^*\right) + p^{-1/2}\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2),\\
&\mathrm{tr}(\boldsymbol{M}_0^2) = (it)^2p^2(N-p)^{-3}\mathrm{tr}\left\{\left(\boldsymbol{A}\tilde{\boldsymbol{V}}_2\right)^2\right\},\\
&\mathrm{tr}\left\{\left(\boldsymbol{A}\tilde{\boldsymbol{V}}_2\right)^2\right\} = \mathrm{tr}\left\{\left(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2)\right)^2\right\}\\
&\qquad = \mathrm{tr}\left\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^2\right\} + 2p^{-1/2}\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p(1),\\
&\mathrm{tr}(\boldsymbol{M}_0^3) = -(it)^3p^3(N-p)^{-9/2}\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^3\},\\
&\mathrm{tr}\{(\boldsymbol{A}\tilde{\boldsymbol{V}}_2)^3\} = \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^{\boldsymbol{*}})^3\} + O_p(p^{-1/2}) = O_{1/2}.
\end{aligned}
$$

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

$$
\begin{aligned}
&\exp\left\{it\frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) + \mathrm{tr}(\boldsymbol{M}_0^2)\right\}\\
&\quad = \exp\left\{it\frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*) + it\frac{p^{1/2}}{N-p}\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2) + (it)^2p^2(N-p)^{-3}\mathrm{tr}\left\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^2\right\}\right\}\\
&\quad \times \exp\left\{2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p((N-p)^{-1})\right\},\\
&\exp\left\{2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_p((N-p)^{-1})\right\}\\
&\quad = 1 + 2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Omega}^*\boldsymbol{A}\boldsymbol{W}_2) + O_1.
\end{aligned}
$$

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

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

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

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

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

$$\begin{aligned}&\boldsymbol{V}_2=\boldsymbol{Z}_2^\top\boldsymbol{Z}_2,\\&\boldsymbol{Z}_2\sim N_{p\times3}(\boldsymbol{M},\boldsymbol{I}_3\otimes\boldsymbol{I}_p),\\&\tilde{\boldsymbol{Z}}_2\sim N_{p\times3}\left(\boldsymbol{M}\left(\boldsymbol{I}_3-2p^{-1/2}\boldsymbol{M}_0^*\right)^{-1},\left(\boldsymbol{I}_3-2p^{-1/2}\boldsymbol{M}_0^*\right)^{-1}\otimes\boldsymbol{I}_p\right).\end{aligned}$$

The moments are given by

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

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

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

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

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

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

$$\begin{aligned}
\boldsymbol{u}_{1i} &= \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_{1i} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),\\
\boldsymbol{u}_{1j} &= \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}_{1j} - \boldsymbol{\mu}_1) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),\\
\boldsymbol{W} &= (n-2)\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}^{(i,j)}\boldsymbol{\Sigma}^{-1/2} \sim W_p(n-2, \boldsymbol{I}_p),\\
\boldsymbol{z}_1 &= \sqrt{n_1 - 1}\boldsymbol{\Sigma}^{-1/2}\left(\bar{\boldsymbol{x}}_1^{(i,j)} - \boldsymbol{\mu}_1\right) \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p),\\
\boldsymbol{z}_2 &= \sqrt{N_2}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1) \sim N_p(\sqrt{N_2}\boldsymbol{\delta}, \boldsymbol{I}_p),
\end{aligned}$$

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

$$\begin{aligned}
&\boldsymbol{V}_2 = \boldsymbol{Q}^\top\boldsymbol{Q} \sim W_4(p, \boldsymbol{I}_4, \boldsymbol{\Omega}),\\
&\boldsymbol{V}_1 = \boldsymbol{T}^\top(\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q})^{-1}\boldsymbol{T} \sim W_4(N-p, \boldsymbol{I}_4),\\
&\boldsymbol{Q}^\top\boldsymbol{W}^{-1}\boldsymbol{Q} = \boldsymbol{T}\boldsymbol{V}_1^{-1}\boldsymbol{T}^\top,
\end{aligned}$$

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,

$$\begin{aligned}
&\boldsymbol{V}_1 = (N-p)\left(\boldsymbol{I}_4 + \frac{1}{\sqrt{N-p}}\boldsymbol{W}_1\right),\\
&D_b^{(-i)}(\boldsymbol{x}_{1i}) = \mathrm{tr}(\boldsymbol{A}_i\boldsymbol{U}) - T_i^{-1}\boldsymbol{a}_i^\top\boldsymbol{U}\boldsymbol{A}_i\boldsymbol{U}\boldsymbol{a}_i, \quad (i = 1, 2).
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\mathrm{tr}(\boldsymbol{A}_i\boldsymbol{U}) = \frac{p}{N-p}\left\{\mathrm{tr}\left(\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\right) + a_{i,0} + a_{i,1}\right\} + O_p(N^{-3/2}),\\
&T_i = \frac{N_1-1}{N_1-2} + \mathrm{tr}(\boldsymbol{B}_i\boldsymbol{U}) = b_{i,0} + b_{i,1} + b_{i,2} + O_p(N^{-3/2}),\\
&\boldsymbol{a}_i^\top\boldsymbol{U}\boldsymbol{A}_i\boldsymbol{U}\boldsymbol{a}_i = \frac{p^2}{(N-p)^2}\left\{\boldsymbol{a}_i^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\boldsymbol{a}_i + c_{i,0} + c_{i,1} + c_{i,2}\right\} + O_p(N^{-3/2}),\\
&T_i^{-1} = s_{i,0} + s_{i,1} + s_{i,2} + O_p(N^{-3/2}),
\end{aligned}$$

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

$$\begin{aligned}
&a_{i,\ell} = (-1)^{\ell+1}(N-p)^{-(\ell+1)/2}\mathrm{tr}\left(\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^{\ell+1}\tilde{\boldsymbol{T}}^\top\right), \quad (\ell = 0,1,2),\\
&b_{i,0} = \frac{N_1-1}{N_1-2} + \frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2\right),\\
&b_{i,\ell} = (-1)^{\ell}(N-p)^{-\ell/2}\frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^{\ell}\tilde{\boldsymbol{T}}^\top\right), \quad (\ell = 1,2),\\
&c_{i,0} = -(N-p)^{-1/2}\boldsymbol{a}_i^\top\left(\tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\right)\boldsymbol{a}_i,\\
&c_{i,1} = (N-p)^{-1}\boldsymbol{a}_i^\top\left(\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{V}}_2\boldsymbol{A}_i\tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top + \tilde{\boldsymbol{T}}\boldsymbol{W}_1^2\tilde{\boldsymbol{T}}^\top\boldsymbol{A}_i\tilde{\boldsymbol{V}}_2\right)\boldsymbol{a}_i,\\
&s_{i,0} = b_{i,0}^{-1}, \quad s_{i,1} = b_{i,1}b_{i,0}^{-2} \quad s_{i,2} = b_{i,0}^{-3}\left(b_{i,1}^2 - b_{i,0}b_{i,2}\right).
\end{aligned}$$

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

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

where

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

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

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

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

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

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

We expand the following conditional expectation,

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

Let

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

Then

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

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

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

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

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

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

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

The moments are given by

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

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

$$\begin{aligned}\mathrm{E}\left[s_{i,1} c_{i,0} | \boldsymbol{V}_2 \right] &= b_{i,0}^{-2} \mathrm{E}\left[b_{i,1} c_{i,0} | \boldsymbol{V}_2 \right], \\ \mathrm{E}\left[s_{i,2} | \boldsymbol{V}_2 \right] &= b_{i,0}^{-3} \left(\mathrm{E}\left[b_{i,1}^2 | \boldsymbol{V}_2 \right] - b_{i,0} \mathrm{E}\left[b_{i,2} | \boldsymbol{V}_2 \right] \right),\end{aligned}$$

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

$$\begin{aligned}\mathrm{E}\left[b_{i,1}^2 | \boldsymbol{V}_2 \right] &= \frac{p^2}{(N-p)^2} \mathrm{E}\left[\left(\mathrm{tr}\left\{ \tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{T}} \left(\frac{1}{N-p} \tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 - \boldsymbol{I}_4 \right) \right\} \right)^2 \middle| \boldsymbol{V}_2 \right] \\ &= \frac{2p^2}{(N-p)^3} \left[\mathrm{tr}\left\{ \left(\boldsymbol{B}_i \tilde{\boldsymbol{V}}_2 \right)^2 \right\} + 2\left\{ \mathrm{tr}\left(\tilde{\boldsymbol{T}}^\top \boldsymbol{B}_i \tilde{\boldsymbol{T}} \boldsymbol{M}_0 \right) \right\}^2 \right] + O_p\left((N-p)^{-3/2} \right),\end{aligned}$$

$$\mathrm{E}\left[b_{i,2}\,|\boldsymbol{V}_2\right] = \frac{p}{(N-p)}\mathrm{E}\left[\mathrm{tr}\left\{\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\left(\frac{1}{N-p}\tilde{\boldsymbol{Z}}_1^\top\tilde{\boldsymbol{Z}}_1-\boldsymbol{I}_4\right)^2\right\}\middle|\,\boldsymbol{V}_2\right]$$
$$= \frac{p}{(N-p)^2}\left[5\mathrm{tr}\left(\boldsymbol{B}_i\tilde{\boldsymbol{V}}_2\right)+4\mathrm{tr}\left(\tilde{\boldsymbol{T}}^\top\boldsymbol{B}_i\tilde{\boldsymbol{T}}\boldsymbol{M}_0^2\right)\right]+O_p\left((N-p)^{-3/2}\right).$$

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.10).

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

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

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

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

$$\mathrm{tr}\left(\boldsymbol{M}_0^3\right) = \mathrm{tr}\left\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^3\right\}+O_p(p^{-1/2}),$$
$$h(\boldsymbol{V}_2) = h(p\boldsymbol{\Omega}^*)+O_p(p^{-1/2}(N-p)^{-1/2})$$
$$= 1+O_1,$$

where

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

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

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

$$
\begin{aligned}
&b_{i,0,1} = \frac{\sqrt{p}}{N-p}\mathrm{tr}\left(\boldsymbol{B}_i \boldsymbol{W}_2\right),\\
&s_{i,0,0} = b_{i,0,0}^{-1} = \frac{n_1(n_1-1)(N-p)}{n_1^2(N-p)+p(n_1-1)(n_1+1)},\\
&s_{i,0,1} = -b_{i,0,0}^{-2}b_{i,0,1}, \quad s_{i,0,2} = b_{i,0,0}^{-3}b_{i,0,1}^2,\\
&\boldsymbol{\Xi}_s = it_1\boldsymbol{\Xi}_{1,s} + it_2\boldsymbol{\Xi}_{2,s},\\
&\boldsymbol{\Xi}_{j,0} = (N-p)^{-1/2}\left[-\frac{p}{N-p}\boldsymbol{A}_j\right.\\
&\qquad \left.+\frac{p^2}{(N-p)^2}s_{j,0,0}\left\{\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j + \boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{B}_j + s_{j,0,0}\frac{p}{N-p}\boldsymbol{a}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j\right\}\right],\\
&\boldsymbol{\Xi}_{j,1} = (N-p)^{-1/2}\frac{p^2}{(N-p)^2}s_{j,0,1}\left(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j + \boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{B}_j + s_{j,0,0}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j\right)\\
&\qquad + \frac{1}{\sqrt{p}}s_{j,0,0}^2\left(\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{W}_2\boldsymbol{a}_j\boldsymbol{B}_j + \boldsymbol{a}_j^\top\boldsymbol{W}_2\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j\right) + s_{j,0,1}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j.
\end{aligned}
$$

Let

$$
\begin{aligned}
\boldsymbol{M}_0^* &= it_1\boldsymbol{M}_{1,0}^* + it_2\boldsymbol{M}_{2,0}^*\\
\boldsymbol{M}_{j,0}^* &= \frac{\sqrt{p}}{N-p}\left\{\frac{p^2}{(N-p)^2}s_{j,0,0}^2\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j\boldsymbol{B}_j + \boldsymbol{A}_j\right.\\
&\qquad \left. - \frac{p}{N-p}s_{j,0,0}\left(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j + \boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{B}_j\right)\right\}.
\end{aligned}
$$

Then

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

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

$$
\begin{aligned}
\phi(\boldsymbol{t}) = \mathrm{E}&\left[\exp\left[it_1\left\{\frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{A}_1\tilde{\boldsymbol{V}}_2\right) - s_{1,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_1^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_1\tilde{\boldsymbol{V}}_2\boldsymbol{a}_1\right\}\right.\right.\\
&\qquad \left. + it_2\left\{\frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{A}_2\tilde{\boldsymbol{V}}_2\right) - s_{2,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_2^\top\tilde{\boldsymbol{V}}_2\boldsymbol{A}_2\tilde{\boldsymbol{V}}_2\boldsymbol{a}_2\right\}\right]\\
&\quad \left.\times\exp\left\{\mathrm{tr}\left(\boldsymbol{M}_0^2\right)\right\}\left\{1 + \frac{4}{3\sqrt{N-p}}\mathrm{tr}\left(\boldsymbol{M}_0^3\right)\right\}h(\boldsymbol{V}_2)\right] + O((N-p)^{-1})\\
&= \exp\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^2\}\left[1 + \mathrm{tr}\left\{(\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*)^3\right\}\right]h(p\boldsymbol{\Omega}^*)\\
&\quad \times\exp\left[it_1\left\{\frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{A}_1\boldsymbol{\Omega}^*\right) - s_{1,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_1^\top\boldsymbol{\Omega}^*\boldsymbol{A}_1\boldsymbol{\Omega}^*\boldsymbol{a}_1\right\}\right.\\
&\qquad \left. it_2\left\{\frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{A}_2\boldsymbol{\Omega}^*\right) - s_{2,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_2^\top\boldsymbol{\Omega}^*\boldsymbol{A}_2\boldsymbol{\Omega}^*\boldsymbol{a}_2\right\}\right]\\
&\quad \times\mathrm{E}\left[\exp\left\{\mathrm{tr}\left(\boldsymbol{M}_0^*\boldsymbol{W}_2\right)\right\}\left[1 + it_1a_{1,1}^* + it_2a_{2,1}^*\right.\right.
\end{aligned}
$$

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

From above result and and Lemma A.6,

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\mathrm{tr}\left(\boldsymbol{M}_0^*\boldsymbol{W}_2\right)\right\}h^*\left(\boldsymbol{V}_2\right)\right]\\
&=\left|\boldsymbol{I}_4-\frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right|^{-p/2}\mathrm{E}\left[h^*\left(\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2\right)\right]\\
&\qquad\times\exp\left\{-p^{1/2}\mathrm{tr}\left(\boldsymbol{M}_0^*\boldsymbol{\Omega}^*\right)+p^{-1/2}\mathrm{tr}\left[\boldsymbol{\Omega}\boldsymbol{M}_0^*\left(\boldsymbol{I}-\frac{2}{\sqrt{p}}\boldsymbol{M}_0^*\right)^{-1}\right\}\right]\\
&=\exp\left[\mathrm{tr}\left\{(\boldsymbol{I}_4+2p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^2\right\}\right]\mathrm{E}\left[h^*\left(\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2\right)\right]\\
&\qquad\times\left[1+\frac{4}{3\sqrt{p}}\mathrm{tr}\left\{(\boldsymbol{I}_4+3p^{-1}\boldsymbol{\Omega})(\boldsymbol{M}_0^*)^3\right\}\right]+O(p^{-1}),
\end{aligned}$$

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

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{tr}\left\{\boldsymbol{\Xi}_0\boldsymbol{\Omega}^*\left(p^{-1/2}\boldsymbol{\Xi}_0\tilde{\boldsymbol{Z}}_2^\top\tilde{\boldsymbol{Z}}_2+\boldsymbol{\Xi}_1\boldsymbol{\Omega}^*\right)\right\}\right]=O(p^{-1}),\\
&\mathrm{E}\left[a_{j,1}^*\right]=-\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}\left[\boldsymbol{a}_j^\top\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j\right]\\
&\qquad-\frac{p^{3/2}}{(N-p)^2}\mathrm{E}\left[s_{j,0,1}\left(\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\right)\right],\\
&\mathrm{E}[s_{j,0,2}]=b_{j,0,0}^{-3}\frac{p}{(N-p)^2}\mathrm{E}\left[\left\{\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)^2\right\}\right],\\
&\mathrm{E}\left[\left\{\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)^2\right\}\right]=4\left\{\mathrm{tr}(\boldsymbol{B}_j\boldsymbol{M}_0^*)\right\}^2+O(p^{-1/2})=O(p^{-1/2}),\\
&\mathrm{E}\left[\boldsymbol{a}_j^\top\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j\right]=\mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\right)\right]=4\mathrm{tr}\left(\boldsymbol{M}_0^*\boldsymbol{B}_j\boldsymbol{M}_0^*\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\right)+O(1)=O(1),\\
&\mathrm{E}\left[s_{j,0,1}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2\boldsymbol{a}_j\right]=\frac{\sqrt{p}}{N-p}b_{j,0,0}^{-2}\mathrm{E}\left[\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)\mathrm{tr}(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2)\right],\\
&\mathrm{E}\left[\mathrm{tr}(\boldsymbol{B}_j\tilde{\boldsymbol{W}}_2)\mathrm{tr}(\boldsymbol{B}_j\boldsymbol{\Omega}^*\boldsymbol{A}_j\tilde{\boldsymbol{W}}_2)\right]=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),
\end{aligned}$$

where

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

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

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

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

$$\begin{aligned}
\eta_j &= \frac{p}{N-p}\mathrm{tr}\left(\boldsymbol{A}_j\boldsymbol{\Omega}^*\right) - s_{j,0,0}\frac{p^2}{(N-p)^2}\boldsymbol{a}_j^\top\boldsymbol{\Omega}^*\boldsymbol{A}_j\boldsymbol{\Omega}^*\boldsymbol{a}_j, \\
&= \frac{n-1}{N-p}\left\{\Delta^2 + \frac{p}{N_2} - b\frac{p(n_1-1)}{n_1^2} + p\left(1 - b\left(1 + \frac{1}{n_1^2}\right)\right)\right\} \\
&\quad + \frac{bp^2(n_1-1)(n-1)}{n_1^3(N-p)^2 + n_1(N-p)p(n_1-1)(n_1+1)}\left(1 + \frac{2(n_1-1)^{1/2}}{n_1^{1/2}} + \frac{n_1-1}{n_1}\right) \\
&= \frac{n-1}{N-p}\left\{\Delta^2 + \frac{p}{N_2} - b\frac{p(n_1-1)}{n_1^2} + p\left(1 - b\left(1 + \frac{1}{n_1^2}\right)\right)\right\} + O_1,
\end{aligned}$$

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

$$\begin{aligned}
\lambda_{jj} &= \mathrm{tr}\left\{\left(\boldsymbol{\Xi}_{j,0}\boldsymbol{\Omega}^*\right)^2\right\} + \mathrm{tr}\left\{\left(\boldsymbol{I}_4 + 2p^{-1}\boldsymbol{\Omega}\right)\left(\boldsymbol{M}_{j,0}^*\right)^2\right\} \\
&= 4\left(\frac{(n-1)^2N}{(N-p)^3}\right)\left(\Delta^2 + \frac{p}{N_2} - b^2\frac{p}{n_1^2} + b^2\frac{p}{n_1}\right) + O_1, \\
\lambda_{12} &= \mathrm{tr}\left(\boldsymbol{\Xi}_{1,0}\boldsymbol{\Omega}^*\boldsymbol{\Xi}_{2,0}\boldsymbol{\Omega}^*\right) + \mathrm{tr}\left(\left(\boldsymbol{I}_4 + 2p^{-1}\boldsymbol{\Omega}\right)\boldsymbol{M}_{1,0}^*\boldsymbol{M}_{2,0}^*\right) \\
&= O_1.
\end{aligned}$$

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

$$\begin{aligned}
d_F^{(-\lambda)}(\boldsymbol{x}_k) &= (\bar{\boldsymbol{x}}_1^{(-k,\lambda)} - \bar{\boldsymbol{x}}_2)^\top\left\{\boldsymbol{S}^{(-k,\lambda)}\right\}^{-1}\left\{\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1^{(-k,\lambda)} + \bar{\boldsymbol{x}}_2)\right\} \\
&= \mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{U}) - (1-\lambda)T_\lambda^{-1}\boldsymbol{a}^\top\boldsymbol{U}\boldsymbol{A}_\lambda\boldsymbol{U}\boldsymbol{a}, \\
T_\lambda &= \frac{N_1^{(\lambda)}}{N_1^{(1)}} + (1-\lambda)\mathrm{tr}(\boldsymbol{B}\boldsymbol{U}),
\end{aligned}$$

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

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

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

$$
\begin{aligned}
T_\lambda &= \frac{N_1^{(-\lambda)}}{N_1^{(-1)}} + O_p\left(N^{-1}\right), \\
\mathrm{tr}(\boldsymbol{A}_\lambda \boldsymbol{U}) - (1-\lambda)T_\lambda^{-1}\boldsymbol{a}^\top \boldsymbol{U}\boldsymbol{A}_\lambda \boldsymbol{U}\boldsymbol{a} &= \frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2) + a_0 + a_1 + O_p\left((N-p)^{-1}\right), \\
a_0 &= -\frac{p}{(N-p)^{3/2}}\mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}}\boldsymbol{W}_1), \\
a_1 &= \frac{p}{(N-p)^2}\mathrm{tr}(\tilde{\boldsymbol{T}}^\top \boldsymbol{A}_\lambda \tilde{\boldsymbol{T}}\boldsymbol{W}_1^2) + \frac{\kappa}{N}\frac{p^2}{(N-p)^2}\boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a},
\end{aligned}
$$

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

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

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

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

where

$$
\begin{aligned}
\tilde{\boldsymbol{Z}}_1 &\sim N_{(N-p)\times 3}\left(\boldsymbol{O}, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1} \otimes \boldsymbol{I}_{N-p}\right), \\
\tilde{\boldsymbol{Z}}_1^\top \tilde{\boldsymbol{Z}}_1 &\sim W_3\left(N-p, \left(\boldsymbol{I}_3 - \frac{2}{\sqrt{N-p}}\boldsymbol{M}_0\right)^{-1}\right)
\end{aligned}
$$

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

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

Secondly, let

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

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:

$$\begin{aligned}
&\boldsymbol{V}_2 = p\left(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2\right), \\
&\mathrm{tr}(\boldsymbol{M}_0) = itp(N-p)^{-3/2}\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2), \\
&\mathrm{tr}(\boldsymbol{A}\tilde{\boldsymbol{V}}_2) = \mathrm{tr}(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2)) \\
&\qquad = \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{\Omega}^*\right) + p^{-1/2}\mathrm{tr}(\boldsymbol{A}\boldsymbol{W}_2), \\
&\mathrm{tr}(\boldsymbol{M}_0^2) = (it)^2 p^2 (N-p)^{-3}\mathrm{tr}\left\{\left(\boldsymbol{A}\tilde{\boldsymbol{V}}_2\right)^2\right\}, \\
&\mathrm{tr}\left\{\left(\boldsymbol{A}\tilde{\boldsymbol{V}}_2\right)^2\right\} = \mathrm{tr}\left\{\left(\boldsymbol{A}(\boldsymbol{\Omega}^* + p^{-1/2}\boldsymbol{W}_2)\right)^2\right\} \\
&\qquad = \mathrm{tr}\left\{(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^*)^2\right\} + 2p^{-1/2}\mathrm{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{W}_2) + O_p(1), \\
&\boldsymbol{a}^\top \tilde{\boldsymbol{V}}_2 \boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2 \boldsymbol{a} = \boldsymbol{a}^\top \boldsymbol{\Omega}^* \boldsymbol{A}_\lambda \boldsymbol{\Omega}^* \boldsymbol{a} + O_p(1) = O_p(1), \\
&\mathrm{tr}(\boldsymbol{M}_0^3) = -(it)^3 p^3 (N-p)^{-9/2}\mathrm{tr}\{(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2)^3\}, \\
&\mathrm{tr}\{(\boldsymbol{A}_\lambda \tilde{\boldsymbol{V}}_2)^3\} = \mathrm{tr}\{(\boldsymbol{A}\boldsymbol{\Omega}^*)^3\} + O_p(p^{-1/2}) = O_{1/2}.
\end{aligned}$$

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

$$\begin{aligned}
&\exp\left\{it\frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}_\lambda\tilde{\boldsymbol{V}}_2) + \mathrm{tr}(\boldsymbol{M}_0^2)\right\} \\
&\quad= \exp\left\{it\frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*) + it\frac{p^{1/2}}{N-p}\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{W}_2) + (it)^2p^2(N-p)^{-3}\mathrm{tr}\left\{(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*)^2\right\}\right\} \\
&\quad\times\exp\left\{2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*\boldsymbol{A}_\lambda\boldsymbol{W}_2) + O_p((N-p)^{-1})\right\}, \\
&\exp\left\{2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*\boldsymbol{A}_\lambda\boldsymbol{W}_2) + O_p((N-p)^{-1})\right\} \\
&\quad= 1 + 2(it)^2p^{3/2}(N-p)^{-3}\mathrm{tr}(\boldsymbol{A}_\lambda\boldsymbol{\Omega}^*\boldsymbol{A}_\lambda\boldsymbol{W}_2) + O_1.
\end{aligned}$$

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

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

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

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

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

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

The moments are given by

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

where

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

From the above result, we have

$$\eta_\lambda = \frac{p}{N-p}\mathrm{tr}(\boldsymbol{A}_\lambda \boldsymbol{\Omega}^*) = \frac{N^{(-\lambda)}}{2(N-p)}\left(\Delta^2 + \frac{p}{N_2} - \frac{n_1 p}{\left\{N_1^{(-\lambda)}\right\}^2} - \frac{(1-\lambda)^2 p}{\left\{N_1^{(-\lambda)}\right\}^2} + \frac{2(1-\lambda)p}{N_1^{(-\lambda)}}\right),$$

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

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 the following formula

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

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

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

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

$$\begin{aligned} s_\lambda^{-1}\eta_\lambda &= -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2} \\ &\quad + \frac{1}{4}\frac{p}{n_1 N_1}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2} \\ &\quad \times \left\{2 - \left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\right\} \\ &\quad - \frac{\kappa p}{n_1 N}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{n_1} + \frac{p}{N_2}\right)^{-1/2} + O_2 \\ &= -\frac{1}{2}\left(\frac{N-p}{N}\right)^{1/2}\left(\Delta^2 + \frac{p}{N_2} - \frac{p}{N_1}\right)\left(\Delta^2 + \frac{p}{N_1} + \frac{p}{N_2}\right)^{-1/2} + O_2. \end{aligned}$$

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. Am. 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.*, Vol.**47** (2017), No.1, 43–62.

[14] H. Yanagihara and H. Fujisawa, Iterative bias correction of the cross-validation criterion. *Scand. J. Stat.*, Vol. **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 Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526 JAPAN
E-mail: nakagawa.stat@gmail.com