Hiroshima Math. J.
47 (2017), 43–62

EPMC estimation in discriminant analysis when the dimension and sample sizes are large

Tetsuji Tonda, Tomoyuki Nakagawa and Hirofumi Wakaki

(Received April 15, 2016)
(Revised October 25, 2016)

Abstract. In this paper we obtain a higher order asymptotic unbiased estimator for the expected probability of misclassification (EPMC) of the linear discriminant function when both the dimension and the sample size are large. Moreover, we evaluate the mean squared error of our estimator. We also present a numerical comparison between the performance of our estimator and that of the other estimators based on Okamoto (1963, 1968) and Fujikoshi and Seo (1998). It is shown that the bias and the mean squared error of our estimator are less than those of the other estimators.

1. Introduction

For $k = 1, 2$, let Π_k be a p-variate normal population with the mean vector $\boldsymbol{\mu}_k$ and a common covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}$ is positive definite and these parameters are unknown. Thus,

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

For $k = 1, 2$, let $\overline{\boldsymbol{X}}_k$ and $\boldsymbol{S}$ be the sample mean vector and the pooled sample covariance matrix, based on a sample of N_k independent observations from Π_k, respectively.

The observation $\boldsymbol{X}$ may be classified by the linear discriminant function $W : \mathbf{R}^p \to \mathbf{R}$ defined by

$$W(\boldsymbol{X}) = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)' \boldsymbol{S}^{-1} \left\{ \boldsymbol{X} - \frac{1}{2}(\overline{\boldsymbol{X}}_1 + \overline{\boldsymbol{X}}_2) \right\},$$

where $\boldsymbol{a}'$ is the transpose of $\boldsymbol{a}$. The classification rule with $W(\boldsymbol{X})$ is as follows: a new observation $\boldsymbol{X}$ is classified as coming from Π_1 if $W(\boldsymbol{X}) > 0$ and from Π_2 otherwise, that is,

$$W(\boldsymbol{X}) > 0 \Rightarrow \boldsymbol{X} \in \Pi_1, \qquad W(\boldsymbol{X}) \leq 0 \Rightarrow \boldsymbol{X} \in \Pi_2.$$

2010 *Mathematics Subject Classification*. Primary 62H30; Secondary 62F12.
Key words and phrases. Discriminant analysis, Classification, High dimensional, Asymptotic expansion, Expected probability of misclassification.

The performance of the classification rule is evaluated by the following probabilities of misclassification:

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

For the optimal linear discriminant rule using the true values of the parameters, we have $P(2|1) = P(1|2) = \Phi(-\Delta/2)$, where Φ is the distribution function of $N(0,1)$ and Δ is the Mahalanobis distance defined by $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ (see [1, 5] for example). In the case that the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are unknown, we use the expected probabilities of misclassification (EPMC), i.e.,

$$e(2|1) = \Pr(W(\boldsymbol{X}) \le 0 | \boldsymbol{X} \in \Pi_1), \qquad e(1|2) = \Pr(W(\boldsymbol{X}) > 0 | \boldsymbol{X} \in \Pi_2),$$

as a risk of the rule with $W(\boldsymbol{X})$. In general, it is hard to obtain the exact evaluation of the EPMC's. There is considerable work for their asymptotic approximations. It may be noted that there are typically two types (type-I, type-II) of their approximations. The type-I approximations are the ones under a framework such that

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

and the type-II approximations are the ones under a framework such that

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

where $N = N_1 + N_2$ and $N - p - 2 > 0$. Okamoto [7] gave an asymptotic expansion for the EPMC of $W(\boldsymbol{X})$ under the type-I approximation framework. Moreover, McLachlan [6] gave an asymptotic unbiased estimator of the EPMC up to terms of $O(N^{-2})$. Deev [2] gave an asymptotic expansion for the EPMC of $W(\boldsymbol{X})$ in the case $N_1 = N_2$ under the type-II approximation framework. Wyman et al. [10] compared the accuracy of several approximations for $W(\boldsymbol{X})$ in the case $N_1 = N_2$, and pointed out that the approximation due to Raudys [8] had overall the best accuracy for the combinations of the parameters considered in their study. Fujikoshi and Seo [4] gave an asymptotic approximation as an extension of Raudys [8]. Fujikoshi [3] gave an asymptotic expansion and its error bound. However, as their approximations are the function of unknown parameter Δ, it must be estimated in practice. The purpose of this paper is to construct an asymptotic unbiased estimator of EPMC and to evaluate the performance of several estimating methods in simulation study.

The present paper is organized in the following way. In Section 2 an asymptotic expansion of EPMC, as the type-II approximation, is derived. In Section 3 we construct a higher order asymptotic unbiased estimator, and evaluate the mean squared error (MSE) of the estimator in the type-II approximation framework. In Section 4 we compare the performances of our estimator with that of the other methods based on [4, 7].

2. Asymptotic expansion

In this section we derive an asymptotic expansion of EPMC under the type-II approximation framework. We denote the distribution function of $W(\boldsymbol{X})$ for $\boldsymbol{X}$ coming from Π_1 by

$$\Pr(W(\boldsymbol{X}) \le w | \boldsymbol{X} \in \Pi_1) = g(w; N_1, N_2, \Delta^2).$$

Then, it is easily seen that

$$\Pr(W(\boldsymbol{X}) \le w | \boldsymbol{X} \in \Pi_2) = 1 - g(-w; N_2, N_1, \Delta^2).$$

The EPMC's of the classification rule are given by

$$e(2|1) = g(0; N_1, N_2, \Delta^2), \qquad e(1|2) = g(0; N_2, N_1, \Delta^2).$$

Hence, it is sufficient to study the distribution of $W(\boldsymbol{X})$ for $\boldsymbol{X}$ coming from Π_1. In the following we assume that $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Assuming that the initial sample $K = (\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{S})$ is fixed, $W(\boldsymbol{X})$ is conditionally distributed as $N(\mu_1(K), \sigma^2(K))$, where $\mu_1(K)$ and $\sigma^2(K)$ depend on the initial sample. Then the conditional probability of misclassification, $P_K(2|1)$, can be expressed as

$$P_K(2|1) = \Phi(T), \qquad T = -\frac{\mu_1(K)}{\sigma(K)}, \tag{1}$$

where

$$\mu_1(K) = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)' \boldsymbol{S}^{-1} \left\{ \boldsymbol{\mu}_1 - \frac{1}{2}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) \right\},$$

$$\sigma^2(K) = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)' \boldsymbol{S}^{-1} \boldsymbol{\Sigma} \boldsymbol{S}^{-1} (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2).$$

The EPMC can be obtained by evaluating $\mathrm{E}_K[P_K(2|1)]$, where $\mathrm{E}_K[\cdot]$ is the expectation with respect to K. Let $\boldsymbol{Z} = \sqrt{m}\boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$, $\boldsymbol{A} = n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}$ and

$$z_1 = \sqrt{\frac{N}{\sigma^2(K)}} (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)' \boldsymbol{S}^{-1} \left(\overline{\boldsymbol{X}}_2 + \frac{N_1}{N}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - \boldsymbol{\mu}_2 \right),$$

where $m = N_1N_2/N$ and $n = N - 2$. Then

$$\boldsymbol{Z} \sim N_p(\boldsymbol{\delta}, \boldsymbol{I}_p), \qquad \boldsymbol{A} \sim W_p(n, \boldsymbol{I}_p), \qquad z_1 \sim N(0,1),$$

and they are independent, where $\boldsymbol{\delta} = \sqrt{m}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\delta}'\boldsymbol{\delta} = m\Delta^2$. By using these variables, we can express T as

$$T = -\frac{1}{\sqrt{mN}} T_3^{-1/2}\left\{N_2T_1 + \frac{1}{2}(N_1 - N_2)T_2\right\} + \frac{1}{\sqrt{N}}z_1, \tag{2}$$

where

$$T_1 = \boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{Z}, \qquad T_2 = \boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}, \qquad T_3 = \boldsymbol{Z}'\boldsymbol{A}^{-2}\boldsymbol{Z}. \tag{3}$$

By using the similar distribution reduction in [4], we have the following lemma.

Lemma 1. *Suppose that* $n - p + 1 > 0$. *Then the statistic* (T_1, T_2, T_3) *in* (3) *can be expressed in terms of independent standard normal variables* z_i $(i = 2, 3)$ *and chi-squared variables* y_i $(i = 1, \dots, 5)$ *with* f_i *degrees of freedom as follows:*

$$T_1 = \frac{\sqrt{m}\Delta}{y_2}\left\{z_2 + \sqrt{m}\Delta + z_3\left(\frac{y_1y_3}{y_4(y_5 + z_3^2)}\right)^{1/2}\right\},$$

$$T_2 = \frac{1}{y_2}\{y_1 + (z_2 + \sqrt{m}\Delta)^2\}, \qquad T_3 = \frac{1}{y_2^2}\{y_1 + (z_2 + \sqrt{m}\Delta)^2\}\left(1 + \frac{y_3}{y_4}\right),$$

where $f_1 = f_3 = p - 1$, $f_2 = n - p + 1$, $f_4 = n - p + 2$ *and* $f_5 = p - 2$.

The proof of this lemma is given in Appendix. From this lemma, T can be written as the function of y_j's and z_j's, i.e., $T = T(y_1, \dots, y_5, z_1, z_2, z_3)$. Note that f_j's tend to infinity as N_1, N_2 and p become large. Let

$$u_j = \sqrt{f_j}\left(\frac{y_j}{f_j} - 1\right).$$

It is well known that u_j is asymptotically distributed as $N(0,2)$ when f_j tends to infinity. Using this property, the expansion of T up to a term of $O_{3/2}$ can be obtained as

$$T = T_{(0)} + T_{(1)} + T_{(2)} + T_{(3)} + O_2, \tag{4}$$

where $T_{(j)}$'s are given in Appendix and O_j means a term of jth order with respect to $(N_1^{-1}, N_2^{-1}, p^{-1})$. Evaluating the expectation $\mathrm{E}[\Phi(T)]$, up to a term of O_1, leads to the following theorem.

Theorem 1. *Suppose that $\boldsymbol{X}$ comes from $\Pi_1 : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Then, under the type-II approximation framework, $e(2|1)$ can be expanded as*

$$e(2|1) = e_{AE}(2|1) + O_2, \qquad e_{AE}(2|1) = \Phi(v) + \phi(v)F_1(\Delta),$$

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

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

and $F_1(\Delta)$ is the term of O_1 given in Appendix.

Remark 1. Δ may tend to infinity depending on p. However, $P(2|1) \to 0$ when $\Delta \to \infty$. Hence, in this paper, we assume that $\Delta = O(1)$ even when $p \to \infty$.

Corollary 1. *Under the Type-I approximation framework, $e_{AE}(2|1)$ can be expanded as*

$$e_{AE}(2|1) = \Phi\left(-\frac{\Delta}{2}\right) + \phi\left(-\frac{\Delta}{2}\right)\left[\frac{1}{16\Delta N_1}\{\Delta^2 + 12(p-1)\} + \frac{1}{16\Delta N_2}\{\Delta^2 - 4(p-1)\} + \frac{\Delta(p-1)}{4(N-1)}\right] + O(N^{-2}).$$

We can see from Corollary 1 that $e_{AE}(2|1)$ is the same as the type-I approximation of $P(2|1)$ in [7] except for $O(N^{-2})$ terms.

3. Derivation of the estimator Q_{TNW}

Under the type-I framework, several estimating techniques of EPMC are reviewed in Siotani et al. [9]. McLachlan [6] derived a higher order asymptotic unbiased estimator by using asymptotic expansions. In this section, by using a technique similar to that in [6], we derive a higher order asymptotic unbiased estimator under the type-II framework.

We consider the following estimator for EPMC:

$$Q_{TNW} = \Phi(\hat{v}) + \hat{Q}_1, \qquad \hat{v} = v(D_s^2),$$

where $\hat{Q}_1$ is a term of O_1, $D_s^2 = f_2D^2/n - f_1/m$ and $D^2 = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1} \cdot (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$. To construct an asymptotic unbiased estimator up to terms of O_1, we define Q_1 such that the bias of Q_{TNW} is O_2. The bias of Q_{TNW} can be expressed as

$$\mathrm{Bias}(Q_{TNW}) = E_K[P_K(2|1) - Q_{TNW}] = e(2|1) - \mathrm{E}[\Phi(\hat{v})] - Q_1,$$

where $\mathrm{E}[\hat{Q}_1] = Q_1$. From Theorem 1, $e(2|1)$ can be expanded as $\Phi(v) + \phi(v)F_1(\Delta) + O_2$, and the expansion of $\mathrm{E}[\Phi(\hat{v})]$ up to a term of O_1 is given in the following lemma.

LEMMA 2. *Suppose that* $\boldsymbol{X}$ *comes from* $\Pi_1 : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. *Then, under the type-II framework,* $\mathrm{E}[\Phi(\hat{v})]$ *can be expanded as*

$$E[\Phi(\hat{v})] = \Phi(v) + \phi(v)G_1(\Delta) + O_2,$$

where $G_1(\Delta)$ *is the term of* O_1 *given in Appendix.*

From Theorem 1 and Lemma 2, it follows that

$$\mathrm{Bias}(Q_{TNW}) = \phi(v)\{F_1(\Delta) - G_1(\Delta)\} - Q_1 + O_2.$$

Therefore, for $Q_1 = \phi(v)\{F_1(\Delta) - G_1(\Delta)\}$, the bias of Q_{TNW} becomes O_2. From this, the estimator of EPMC defined by

$$Q_{TNW} = \Phi(\hat{v}) + \hat{Q}_1, \qquad \hat{Q}_1 = \phi(\hat{v})\{F_1(D_s) - G_1(D_s)\} \tag{5}$$

is asymptotically unbiased up to a term of O_1. We call this estimating technique TNW method.

Moreover, $\hat{v}$ can be expanded as

$$\hat{v} = v(1 + v_{(1)} + v_{(2)} + v_{(3)}) + O_2,$$

where $v_{(i)}$'s are given in Appendix. Then the variance of our estimator is given by

$$\begin{aligned}
\mathrm{Var}(Q_{TNW}) &= \mathrm{E}[Q_{TNW}^2] - \mathrm{E}[Q_{TNW}]^2 \\
&= v^2\phi(v)^2\mathrm{E}[v_{(1)}^2] + O_{3/2} \\
&= \frac{\phi(v)^2}{4}\left(\frac{N-p}{N-1}\right)\left(\frac{f_1 + 2m\Delta^2}{(f_1 + m\Delta^2)^2} + \frac{1}{f_2}\right) \\
&\quad \times \frac{(\Delta^2 + f_1(N + 2N_2)/N_1N_2)^2}{\Delta^2 + f_1N/N_1N_2} + O_{3/2}.
\end{aligned}$$

Thus, we have the mean squared error (MSE) of our estimator as follows:

$$\begin{aligned}
\mathrm{MSE}(Q_{TNW}) &= \mathrm{E}[\{Q_{TNW} - P(2|1)\}^2] \\
&= \mathrm{Var}(Q_{TNW}) + \{\mathrm{E}[Q_{TNW}] - P(2|1)\}^2
\end{aligned}$$

$$= \frac{\phi(v)^2}{4}\left(\frac{N-p}{N-1}\right)\left(\frac{f_1+2m\Delta^2}{(f_1+m\Delta^2)^2}+\frac{1}{f_2}\right)$$
$$\times\frac{(\Delta^2+f_1(N+2N_2)/N_1N_2)^2}{\Delta^2+f_1N/N_1N_2}+O_{3/2}.$$

Therefore, the MSE of our estimator is O_1 under the type-II asymptotic framework.

4. Simulation study

We study the accuracy of asymptotic approximations and the performance of the estimator of EPMC. Without loss of generality, we assume that $\boldsymbol{\mu}_1 = (-\Delta/2, 0, \dots, 0)'$, $\boldsymbol{\mu}_2 = (\Delta/2, 0, \dots, 0)'$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. Let $e_O(2|1;\Delta)$ denote the asymptotic expansion up to the second order with respect to $(N_1^{-1}, N_2^{-1}, n^{-1})$ due to Okamoto [7]. For the type-II approximation, Fujikoshi and Seo [4] gave the asymptotic approximation defined by $e_{FS}(2|1;\Delta) = \Phi(\gamma)$, where

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

4.1. Comparison of accuracy. First we compare the accuracy of $e_{AE}(2|1;\Delta)$ with those of $e_{FS}(2|1;\Delta)$ and $e_O(2|1;\Delta)$. The configuration of the values of N_1, N_2, p and Δ are $N_1, N_2 = 10, 20, 30, 40$, $p = 5, 10, 20, 30, 40$ 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. For each of the configurations, $e(2|1)$ is obtained by Monte Calro simulation as $e(2|1) = B^{-1}\sum_{i=1}^{B} c_i(2|1)$, where $c_i(2|1)$ is the conditional probability of misclassification, defined by (1), for the ith iteration.

The overall performance of the several asymptotic approximations across all configurations of parameters is described graphically in Figure 1, which is a scatter plot of $e(2|1)$ [x-axis] versus each asymptotic approximation [y-axis]. In each graph, the circular (○), plus (+) and triangle (△) marks denote the approximations $e_{AE}(2|1;\Delta)$, $e_{FS}(2|1;\Delta)$ and $e_O(2|1;\Delta)$, respectively. Table 1 gives the approximated values of $e(2|1)$ by each methods in the case $p = 10$. From Figure 1 and Table 1, we see that $e_{AE}(2|1;\Delta)$ is better than the other ones.

4.2. A comparison of performance of EPMC estimators. Next, we compare our estimator in (5) with the other estimating methods. Under the type-I approximation framework, McLachlan [6] suggested an estimating method

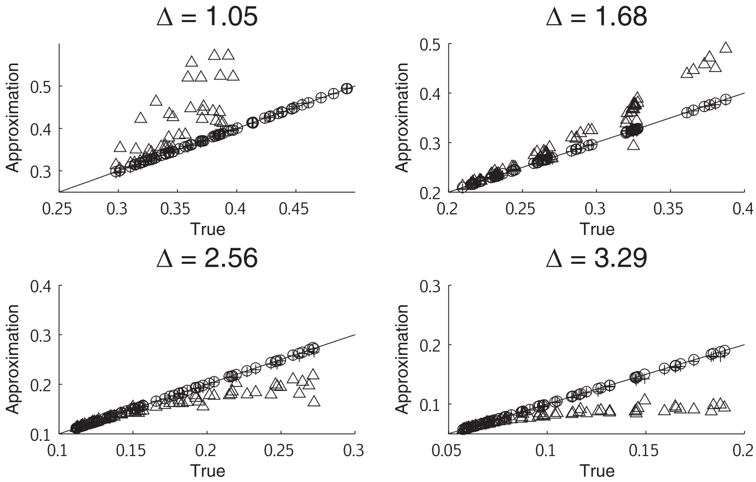

Fig. 1. True EPMC values [x-axis] versus asymptotic approximations values [y-axis].

called M method. The bias of its estimator is O_3 under the type-I approximation framework. Under the type-II approximation framework, we can consider two estimating methods, which are based on $e_{AE}(2|1;\Delta)$ and $e_{FS}(2|1;\Delta)$ with Δ^2 replaced by $\hat{\Delta}^2$, respectively. We call them AE and FS methods, respectively. $\hat{\Delta}^2$ is given by

$$\hat{\Delta}^2 = \begin{cases} \dfrac{n-p-3}{n}D^2 - \dfrac{pN}{N_1N_2} & \text{if } \hat{\Delta} \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

$\hat{\Delta}^2$ is a consistent estimator of Δ^2 under both of the approximation frameworks. The values of N_1, N_2, p and Δ are chosen as follows:

$$N_1, N_2 = 10, 20, 30, \qquad N = N_1 + N_2, \qquad p/N = 0.2, 0.3, \dots, 0.8,$$
$$\Delta = 1.05, 1.68, 2.56, 3.29, \qquad \text{satisfying } N - p - 2 > 0.$$

The performance of each estimator is evaluated by the MSE

$$B^{-1}\sum_{i=1}^{B}\{\hat{e}_i(2|1) - e(2|1)\}^2,$$

where B is the number of iterations in Monte Calro simulation and $\hat{e}_i(2|1)$ denotes the estimation of $e(2|1)$ in the ith iteration.

Figure 2 shows the box plot of bias $\mathrm{E}[\hat{e}(2|1)] - e(2|1)$ for several configurations of N_1, N_2 and p. Figure 3 shows the box plots of the difference of MSE for AE, FS and M versus TNW for several configurations of N_1,

Table 1. Values of approximations and simulation in the case $p = 10$.

(N_1, N_2)	Δ	$e(2\|1)$	$e_O(2\|1;\Delta)$	$e_{FS}(2\|1;\Delta)$	$e_{AE}(2\|1;\Delta)$
(10, 10)	1.05	0.41378	0.67276	0.41243	0.41423
	1.68	0.32707	0.38883	0.32477	0.32757
	2.56	0.21887	0.20270	0.21411	0.21956
	3.29	0.14977	0.10625	0.14261	0.15021
(10, 20)	1.05	0.43789	0.63900	0.43941	0.43794
	1.68	0.32245	0.35939	0.32418	0.32306
	2.56	0.19271	0.17791	0.19192	0.19315
	3.29	0.11767	0.09130	0.11495	0.11778
(20, 10)	1.05	0.34516	0.42634	0.34254	0.34527
	1.68	0.25910	0.28092	0.25707	0.25948
	2.56	0.15752	0.15509	0.15512	0.15785
	3.29	0.09714	0.08458	0.09394	0.09694
(20, 20)	1.05	0.37076	0.44067	0.37099	0.37080
	1.68	0.26532	0.28023	0.26595	0.26552
	2.56	0.15127	0.14725	0.15091	0.15136
	3.29	0.08742	0.07808	0.08644	0.08748
(10, 30)	1.05	0.44907	0.62125	0.45188	0.44894
	1.68	0.32061	0.34608	0.32351	0.32086
	2.56	0.18131	0.16766	0.18202	0.18192
	3.29	0.10473	0.08506	0.10358	0.10505
(30, 10)	1.05	0.31524	0.35036	0.31177	0.31508
	1.68	0.23233	0.24230	0.22931	0.23203
	2.56	0.13513	0.13503	0.13280	0.13519
	3.29	0.07897	0.07394	0.07679	0.07896
(20, 30)	1.05	0.38022	0.44019	0.38178	0.38023
	1.68	0.26554	0.27638	0.26724	0.26570
	2.56	0.14649	0.14212	0.14664	0.14634
	3.29	0.08176	0.07431	0.08137	0.08179
(30, 20)	1.05	0.34213	0.37860	0.34166	0.34215
	1.68	0.24208	0.25070	0.24223	0.24232
	2.56	0.13496	0.13336	0.13441	0.13485
	3.29	0.07556	0.07102	0.07499	0.07569
(30, 30)	1.05	0.35168	0.38389	0.35259	0.35177
	1.68	0.24412	0.25068	0.24520	0.24429
	2.56	0.13253	0.13078	0.13281	0.13261
	3.29	0.07261	0.06874	0.07249	0.07272

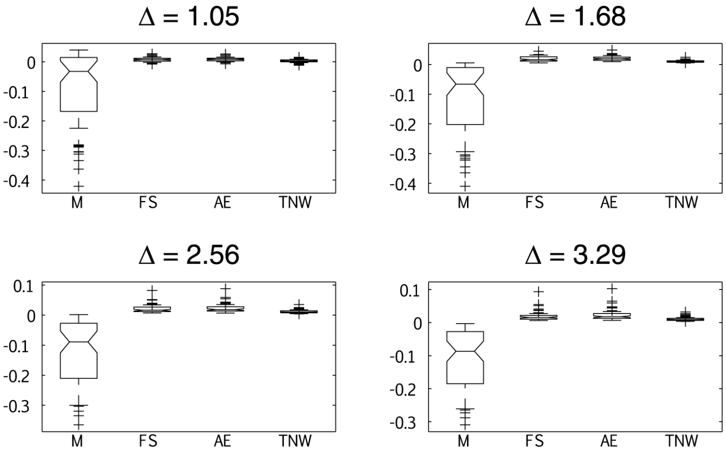

Fig. 2. Box plots of the bias $\mathrm{E}[\hat{e}(2|1)] - e(2|1)$.

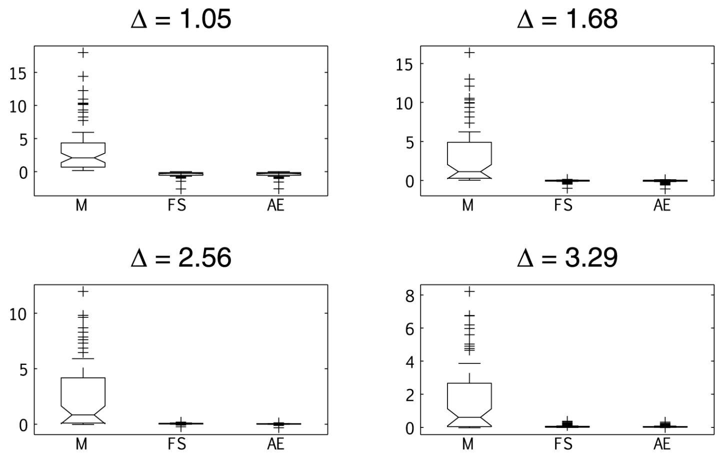

Fig. 3. Box plots of the MSE of other estimators – MSE(TNW).

N_2 and p. From Figures 2 and 3, we can see that M is worse than TNW, AE and FS. The MSE of TNW is not less than AE and FS, but the bias of TNW is better than AE and FS. Tables 2 and 3 give the values of estimators by M, FS, AE and TNW in the case that $p/N = 1/5$ and $4/5$, respectively. From Tables 2 and 3, we can see that TNW has the smaller bias than the other methods. Tables 4 and 5 give the values of $100 \times$ (the MSE of other estimators – MSE(TNW)) in the case that $p/N = 1/5$ and $4/5$, respectively.

From the above results, our estimator is better than the other estimators.

Table 2. Bias of M, FS, AE and TNW in the case $p/N = 1/5$.

(N_1, N_2)	Δ	M	FS	AE	TNW
(20, 20)	1.05	0.01998	0.01677	0.01636	0.01074
	1.68	0.00132	0.01504	0.01425	0.00930
	2.56	−0.00556	0.01195	0.01182	0.00718
	3.29	−0.00750	0.00952	0.00996	0.00485
(10, 30)	1.05	0.03717	0.02412	0.02074	0.01024
	1.68	−0.00112	0.02697	0.02375	0.01148
	2.56	−0.01427	0.01843	0.01727	0.00867
	3.29	−0.01417	0.01390	0.01426	0.00650
(30, 10)	1.05	0.00962	0.00794	0.01118	0.01193
	1.68	0.00264	0.01192	0.01462	0.01142
	2.56	−0.00226	0.01020	0.01231	0.00803
	3.29	−0.00453	0.00865	0.01045	0.00544

Table 3. Bias of M, FS, AE and TNW in the case $p/N = 4/5$.

(N_1, N_2)	Δ	M	FS	AE	TNW
(20, 20)	1.05	−0.30832	−0.00095	0.00057	0.00984
	1.68	−0.32088	0.01801	0.02068	0.01379
	2.56	−0.30083	0.02651	0.03166	0.01667
	3.29	−0.26022	0.02617	0.03363	0.01765
(10, 30)	1.05	−0.41857	−0.00330	−0.00446	0.01011
	1.68	−0.41307	0.02108	0.02124	0.01162
	2.56	−0.36602	0.03978	0.04294	0.01686
	3.29	−0.30990	0.04062	0.04697	0.01851
(30, 10)	1.05	−0.28121	−0.00801	−0.00365	0.01390
	1.68	−0.28706	0.00651	0.01170	0.01508
	2.56	−0.26659	0.01817	0.02515	0.01715
	3.29	−0.23129	0.02109	0.02982	0.01822

Appendix

A.1. Proof of the consistency of $\hat{\Delta}^2$ and D_s^2.

Let

$$\tilde{\Delta}^2 = \frac{n-p-1}{n} D^2 - \frac{pN}{N_1 N_2}.$$

Then

$$\mathrm{E}[\tilde{\Delta}^2] = \Delta^2, \qquad \tilde{\Delta}^2 \xrightarrow{p} \Delta^2.$$

Table 4. Values of the MSE of other estimators – MSE(TNW) in the case $p/N = 1/5$.

(N_1, N_2)	Δ	M	FS	AE
(20, 20)	1.05	0.472	−0.036	−0.030
	1.68	0.064	0.030	0.026
	2.56	0.004	0.010	0.005
	3.29	−0.005	0.008	0.006
(10, 30)	1.05	2.082	−0.191	−0.200
	1.68	0.295	0.133	0.104
	2.56	0.019	0.058	0.035
	3.29	−0.014	0.029	0.022
(30, 10)	1.05	0.164	−0.114	−0.104
	1.68	0.016	−0.012	0.002
	2.56	−0.004	−0.006	0.001
	3.29	−0.006	0.001	0.006

Table 5. Values of the MSE of other estimators – MSE(TNW) in the case $p/N = 4/5$.

(N_1, N_2)	Δ	M	FS	AE
(20, 20)	1.05	10.553	−0.383	−0.416
	1.68	10.394	−0.137	−0.177
	2.56	8.363	0.078	0.037
	3.29	5.913	0.077	0.047
(10, 30)	1.05	17.924	−0.897	−0.956
	1.68	16.388	−0.434	−0.525
	2.56	12.055	0.152	0.042
	3.29	8.200	0.274	0.179
(30, 10)	1.05	8.437	−0.331	−0.360
	1.68	8.162	−0.220	−0.245
	2.56	6.506	−0.052	−0.073
	3.29	4.624	−0.007	−0.017

where $N = N_1 + N_2$, $n = N - 2$. In fact, D^2 can be expressed as

$$D^2 = \frac{n}{m} \frac{(z_2 + \sqrt{m}\Delta)^2 + y_1}{y_2}, \qquad (m = N_1 N_2 / N).$$

Then

$$\mathrm{E}[D^2] = \frac{n}{n-p-1}\left(\frac{p}{m} + \Delta^2\right),$$

$$\mathrm{Var}(D^2) = \frac{n^2}{m^2}\frac{(n-p-1)(2p+4m\Delta^2) + 2(p+m\Delta)^2}{(n-p-1)^2(n-p-3)}.$$

Thus, $\mathrm{E}[\tilde{\Delta}^2] = \Delta^2$ and $\mathrm{Var}(\tilde{\Delta}^2) \to 0$. Therefore $\tilde{\Delta}^2$ is a consistent estimator of Δ^2. From this, we can easily show that $\hat{\Delta}^2$ and D_s^2 are consistent estimators of Δ^2, as desired.

A.2. Proof of Lemma 1. Suppose that the $p \times p$ orthogonal matrices $\boldsymbol{H}$ and $\boldsymbol{Q}$ are given by

$$\boldsymbol{H} = ((\boldsymbol{z}'\boldsymbol{z})^{-1/2}\boldsymbol{z}, \{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_z)\boldsymbol{\delta}\}^{-1/2}(\boldsymbol{I}_p - \boldsymbol{\Pi}_z)\boldsymbol{\delta}, \boldsymbol{H}_1),$$

$$\boldsymbol{Q} = ((\boldsymbol{\delta}'\boldsymbol{\delta})^{-1/2}\boldsymbol{\delta}, \boldsymbol{Q}_1).$$

Let $\tilde{\boldsymbol{A}} = \boldsymbol{H}'\boldsymbol{A}^{-1}\boldsymbol{H}$ and $\tilde{\boldsymbol{A}}$ be partitioned as

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} \tilde{A}_{11} & \tilde{\boldsymbol{A}}_{12} \\ \tilde{\boldsymbol{A}}_{21} & \tilde{\boldsymbol{A}}_{22} \end{pmatrix}, \qquad \tilde{A}_{11} : 1 \times 1.$$

Then, $\tilde{\boldsymbol{A}}$ is distributed as $W_p(n, \boldsymbol{I}_p)$, and

$$\tilde{\boldsymbol{A}}^{-1} = \begin{pmatrix} 0 & \boldsymbol{0}' \\ \boldsymbol{0} & \tilde{\boldsymbol{A}}_{22}^{-1} \end{pmatrix} + \begin{pmatrix} 1 \\ -\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21} \end{pmatrix} \tilde{A}_{11.2}^{-1}(1 \quad -\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1}),$$

$$\tilde{\boldsymbol{A}}^{-2} = \begin{pmatrix} 0 & -\tilde{A}_{11.2}^{-1}\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2} \\ -\tilde{A}_{11.2}^{-1}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21} & \tilde{\boldsymbol{A}}_{22}^{-2} \end{pmatrix} + \begin{pmatrix} 1 \\ -\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21} \end{pmatrix} \tilde{A}_{11.2}^{-2}(1 + \tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21})(1 \quad -\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1}),$$

where $\tilde{A}_{11.2} = \tilde{A}_{11} - \tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}$. Moreover, $\tilde{A}_{11.2}$, $\tilde{\boldsymbol{A}}_{22}$ and $\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21}$ are mutually independent, and $\tilde{A}_{11.2}$, $\tilde{\boldsymbol{A}}_{22}$ and $\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21}$ are distributed as χ^2_{n-p+1}, $W_{p-1}(n, \boldsymbol{I}_{p-1})$ and $N_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$, respectively (see [1, 5] for example). Therefore, we have

$$\begin{aligned} T_1 &= \boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{z} = \boldsymbol{\delta}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{A}^{-1}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{z} \\ &= ((\boldsymbol{z}'\boldsymbol{z})^{-1/2}(\boldsymbol{\delta}'\boldsymbol{z}), \{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_z)\boldsymbol{\delta}\}^{1/2}, \boldsymbol{0}')\tilde{\boldsymbol{A}}^{-1}\begin{pmatrix} (\boldsymbol{z}'\boldsymbol{z})^{1/2} \\ \boldsymbol{0} \end{pmatrix} \\ &= \tilde{A}_{11.2}^{-1}\{\boldsymbol{\delta}'\boldsymbol{z} - (\boldsymbol{z}'\boldsymbol{z})^{1/2}\{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_z)\boldsymbol{\delta}\}^{1/2}\boldsymbol{e}_1'\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\}, \\ &\quad (\boldsymbol{e}_1 = (1, 0, \ldots, 0)' : (p-1) \times 1) \end{aligned}$$

$$= \tilde{A}_{11.2}^{-1}\left[\boldsymbol{\delta}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z} - \{m\Delta^2(\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z}) - (\boldsymbol{\delta}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z})^2\}^{1/2}\right.$$
$$\left.\times \frac{\boldsymbol{e}_1'\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}}{(\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21})^{1/2}}\{(\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1/2})\tilde{\boldsymbol{A}}_{22}^{-1}(\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21})\}^{1/2}\right]$$
$$= \frac{1}{y_2}\left[\sqrt{m}\Delta z_2 + m\Delta^2 - \{m\Delta^2(y_1 + (z_2 + \sqrt{m}\Delta)^2) - (\sqrt{m}\Delta z_2 + m\Delta^2)^2\}^{1/2}\right.$$
$$\left.\times \frac{\tilde{z}_3}{(y_5 + \tilde{z}_3^2)^{1/2}}\left(\frac{y_3}{y_4}\right)^{1/2}\right]$$
$$= \frac{\sqrt{m}\Delta}{y_2}\left\{z_2 + \sqrt{m}\Delta - \tilde{z}_3\left(\frac{y_1 y_3}{y_4(y_5 + \tilde{z}_3^2)}\right)^{1/2}\right\}$$
$$= \frac{\sqrt{m}\Delta}{y_2}\left\{z_2 + \sqrt{m}\Delta + z_3\left(\frac{y_1 y_3}{y_4(y_5 + z_3^2)}\right)^{1/2}\right\}, \qquad (z_3 = -\tilde{z}_3, \tilde{z}_3 \sim N(0,1)),$$
$$T_2 = \boldsymbol{z}'\boldsymbol{A}^{-1}\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{A}^{-1}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{z}$$
$$= \tilde{A}_{11.2}^{-1}(\boldsymbol{z}'\boldsymbol{z}) = \tilde{A}_{11.2}^{-1}(\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z}) = \frac{y_1 + (z_2 + \sqrt{m}\Delta)^2}{y_2}$$

and

$$T_3 = \boldsymbol{z}'\boldsymbol{A}^{-2}\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{A}^{-2}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{z}$$
$$= (\boldsymbol{z}'\boldsymbol{z})\tilde{A}_{11.2}^{-2}(1 + \tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21})$$
$$= (\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z})\tilde{A}_{11.2}^{-2}\{1 + (\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1/2})\tilde{\boldsymbol{A}}_{22}^{-1}(\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21})\}$$
$$= \frac{y_1 + (z_2 + \sqrt{m}\Delta)^2}{y_2^2}\left(1 + \frac{y_3}{y_4}\right).$$

A.3. Calculation of $F_1(\Delta)$. The expansions of T_j's, up to O_1 terms, can be given by

$$T_i = t_{i,0}(1 + t_{i,1} + t_{i,2} + t_{i,3}) + O_2, \qquad i = 1, 2, 3,$$

where $t_{i,j}$ $(j = 0, 1, 2, 3)$ are given by

$$t_{1,0}=\frac{m\Delta^2}{f_2},\qquad t_{1,1}=\frac{1}{\sqrt{m}\Delta}\left(z_2+z_3\sqrt{\frac{f_1f_3}{f_4f_5}}\right)-\frac{u_2}{\sqrt{f_2}},$$

$$t_{1,2}=\frac{z_3}{2\sqrt{m}\Delta}\sqrt{\frac{f_1f_3}{f_4f_5}}\left(\frac{u_1}{\sqrt{f_1}}+\frac{u_3}{\sqrt{f_3}}-\frac{u_4}{\sqrt{f_4}}-\frac{u_5}{\sqrt{f_5}}\right)$$
$$+\frac{u_2^2}{f_2}-\frac{u_2}{\sqrt{mf_2}\Delta}\left(z_2+z_3\sqrt{\frac{f_1f_3}{f_4f_5}}\right),$$

$$t_{1,3}=\frac{z_3}{4\sqrt{m}\Delta}\sqrt{\frac{f_1f_3}{f_4f_5}}\left(-\frac{u_1^2}{f_1}-\frac{u_3^2}{f_3}+\frac{u_4^2}{f_4}+\frac{u_5^2}{f_5}-\frac{z_3^2}{f_5}\right.$$
$$\left.+\frac{2u_1u_3}{\sqrt{f_1f_3}}-\frac{2u_1u_4}{\sqrt{f_1f_4}}-\frac{2u_1u_5}{\sqrt{f_1f_5}}-\frac{2u_3u_4}{\sqrt{f_3f_4}}-\frac{2u_3u_5}{\sqrt{f_3f_5}}+\frac{2u_4u_5}{\sqrt{f_4f_5}}\right)$$
$$-\frac{u_2^3}{f_2\sqrt{f_2}}+\frac{u_2^2}{f_4\sqrt{m}\Delta}\left(z_2+z_3\sqrt{\frac{f_1f_3}{f_4f_5}}\right)$$
$$-\frac{z_3u_2}{2\sqrt{mf_4}\Delta}\sqrt{\frac{f_1f_3}{f_4f_5}}\left(\frac{u_1}{\sqrt{f_1}}+\frac{u_3}{\sqrt{f_3}}-\frac{u_4}{\sqrt{f_4}}-\frac{u_5}{\sqrt{f_5}}\right),$$

$$t_{2,0}=\frac{f_1+m\Delta^2}{f_2},\qquad t_{2,1}=\frac{2\sqrt{m}\Delta z_2+\sqrt{f_1}u_1}{f_1+m\Delta^2}-\frac{u_2}{\sqrt{f_2}},$$

$$t_{2,2}=\frac{z_2^2}{f_1+m\Delta^2}+\frac{u_2^2}{f_2}-\frac{u_2(2\sqrt{m}\Delta z_2+\sqrt{f_1}u_1)}{\sqrt{f_2}(f_1+m\Delta^2)},$$

$$t_{2,3}=\frac{u_2^2(2\sqrt{m}\Delta z_2+\sqrt{f_1}u_1)}{f_2(f_1+m\Delta^2)}-\frac{z_2^2u_2}{\sqrt{f_2}(f_1+m\Delta^2)}-\frac{u_2^3}{f_2\sqrt{f_2}},$$

$$t_{3,0}=\frac{(f_1+m\Delta^2)(f_3+f_4)}{f_2^2f_4},$$

$$t_{3,1}=\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}-\frac{2u_2}{\sqrt{f_2}}+\frac{\sqrt{f_3}u_3+\sqrt{f_4}u_4}{f_3+f_4}-\frac{u_4}{\sqrt{f_4}},$$

$$t_{3,2}=\frac{z_2^2}{f_1+m\Delta^2}+\frac{3u_2^2}{f_2}-\frac{2\sqrt{f_1}u_1+4\sqrt{m}\Delta z_2}{\sqrt{f_2}(f_1+m\Delta^2)}+\frac{f_3u_4^2}{f_4(f_3+f_4)}$$
$$-\frac{\sqrt{f_3}u_3u_4}{(f_3+f_4)\sqrt{f_4}}+\frac{\sqrt{f_3}u_3}{f_3+f_4}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta}\right)$$

$$- \frac{2\sqrt{f_3}u_2u_3}{(f_3+f_4)\sqrt{f_2}} - \frac{f_3u_4}{(f_3+f_4)}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2}{f_1 + m\Delta}\right)$$

$$+ \frac{f_3u_2u_4}{(f_3+f_4)\sqrt{f_2f_4}},$$

$$t_{3,3} = \frac{3}{f_2}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2}{f_1 + m\Delta^2}\right) - \frac{2u_2z_2}{(f_1+m\Delta^2)\sqrt{f_2}} - \frac{4u_2^3}{f_2\sqrt{f_2}}$$

$$+ \frac{\sqrt{f_3}u_3u_4^2}{f_4(f_3+f_4)} - \frac{f_3u_4^3}{(f_3+f_4)f_4\sqrt{f_4}}$$

$$+ \frac{f_3u_4^2}{f_4(f_3+f_4)}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta^2 z_2}{f_1 + m\Delta^2}\right) - \frac{2f_3u_2u_4^2}{f_4(f_3+f_4)\sqrt{f_2}}$$

$$- \frac{\sqrt{f_3}u_3u_4}{(f_3+f_4)\sqrt{f_4}}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2}{f_1 + m\Delta^2}\right) + \frac{2\sqrt{f_3}u_2u_3u_4}{(f_3+f_4)\sqrt{f_2f_4}}$$

$$+ \frac{\sqrt{f_3}u_3z_2^2}{(f_3+f_4)(f_1+m\Delta^2)} + \frac{3\sqrt{f_3}u_2^2u_3}{(f_3+f_4)f_2}$$

$$- \frac{2\sqrt{f_3}u_2u_3}{(f_3+f_4)\sqrt{f_2}}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2}{f_1 + m\Delta^2}\right)$$

$$- \frac{f_3u_4z_2^2}{(f_3+f_4)(f_1+m\Delta^2)\sqrt{f_4}} - \frac{3f_3u_2^2u_4}{(f_3+f_4)f_2\sqrt{f_4}}$$

$$+ \frac{2f_3u_2u_4}{(f_3+f_4)\sqrt{f_2f_4}}\left(\frac{\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2}{f_1 + m\Delta^2}\right).$$

Using these expressions, $T_{(j)}$'s in (4) can be written as

$$T_{(0)} = a_1 + a_2,$$

$$T_{(1)} = -\frac{T_{(0)}}{2}t_{3,1} + a_1t_{1,1} + a_2t_{2,1} + \frac{1}{\sqrt{N}}z_1,$$

$$T_{(2)} = T_{(0)}\left(\frac{3}{8}t_{3,1}^2 - \frac{1}{2}t_{3,2}\right) + a_1\left(t_{1,2} - \frac{1}{2}t_{3,1}t_{1,1}\right) + a_2\left(t_{2,2} - \frac{1}{2}t_{3,1}t_{2,1}\right),$$

$$T_{(3)} = T_{(0)}\left(-\frac{5}{16}t_{3,1}^3 + \frac{3}{4}t_{3,1}t_{3,2} - \frac{1}{2}t_{3,3}\right)$$

$$+ a_1\left(t_{1,3} - \frac{1}{2}t_{3,1}t_{1,2} + \frac{3}{8}t_{1,1}t_{3,1}^2 - \frac{1}{2}t_{1,1}t_{3,2}\right)$$

$$+ a_2\left(t_{2,3} - \frac{1}{2}t_{3,1}t_{2,2} + \frac{3}{8}t_{2,1}t_{3,1}^2 - \frac{1}{2}t_{2,1}t_{3,2}\right),$$

where $a_1 = aN_2t_{1,0}$, $a_2 = a(N_1 - N_2)t_{2,0}/2$ and $a = N^{-1}(mt_{3,0})$. Then $F_1(\Delta)$ can be obtained by calculating

$$F_1(\Delta) = \mathrm{E}[T_{(2)}] - \frac{T_{(0)}}{2}\mathrm{E}[T_{(1)}^2],$$

where the moments of $t_{i,j}$'s are given by

$$\mathrm{E}[t_{1,1}] = \mathrm{E}[t_{2,1}] = \mathrm{E}[t_{3,1}] = 0, \qquad \mathrm{E}[t_{1,2}] = \frac{2}{f_2}, \qquad \mathrm{E}[t_{2,2}] = \frac{1}{f_1 + m\Delta^2} + \frac{2}{f_2},$$

$$\mathrm{E}[t_{3,2}] = \frac{1}{f_1 + m\Delta^2} + \frac{6}{f_2} + \frac{2f_3}{f_4(f_3 + f_4)},$$

$$\mathrm{E}[t_{1,3}] = 0, \qquad \mathrm{E}[t_{2,3}] = \mathrm{E}[t_{3,3}] = O_2,$$

$$\mathrm{E}[t_{1,1}^2] = \frac{1}{m\Delta^2}\left(1 + \frac{f_1f_3}{f_4f_5}\right) + \frac{2}{f_2}, \qquad \mathrm{E}[t_{2,1}^2] = \frac{2f_1 + 4m\Delta^2}{(f_1 + m\Delta^2)^2} + \frac{2}{f_2},$$

$$\mathrm{E}[t_{3,1}^2] = \frac{2f_1 + 4m\Delta^2}{(f_1 + m\Delta^2)^2} + \frac{8}{f_2} + \frac{2f_3}{f_4(f_3 + f_4)},$$

$$\mathrm{E}[t_{1,1}t_{2,1}] = \frac{2}{f_1 + m\Delta^2} + \frac{2}{f_2}, \qquad \mathrm{E}[t_{1,1}t_{3,1}] = \frac{2}{f_1 + m\Delta^2} + \frac{4}{f_2},$$

$$\mathrm{E}[t_{2,1}t_{3,1}] = \frac{2f_1 + 4m\Delta^2}{(f_1 + m\Delta^2)^2} + \frac{4}{f_2},$$

and the remainder of the moments of t_{ij}'s are O_2.

A.4. Proof of Corollary 1. From Theorem 1, we have the following result under the type-I approximation:

$$T_{(0)} = -\frac{\Delta}{2} + \frac{3(p-1)}{4\Delta N_1} - \frac{p-1}{4\Delta N_2} + \frac{\Delta(p-1)}{4(N-1)} + O(N^{-2}),$$

$$a_1 = -\frac{N_2}{N}\Delta + O(N^{-1}), \qquad a_2 = -\frac{N_1 - N_2}{N}\Delta + O(N^{-1}),$$

$$\mathrm{E}[t_{1,1}^2] = \frac{1}{m\Delta^2} + \frac{2}{N} + O(N^{-2}), \qquad \mathrm{E}[t_{1,2}] = \frac{2}{N} + O(N^{-2}),$$

$$\mathrm{E}[t_{2,1}^2] = \frac{4}{m\Delta^2} + \frac{2}{N} + O(N^{-2}),$$

$$\mathrm{E}[t_{2,2}] = \frac{1}{m\Delta^2} + \frac{2}{N} + O(N^{-2}), \qquad \mathrm{E}[t_{1,1}t_{1,2}] = \frac{2}{m\Delta^2} + \frac{2}{N} + O(N^{-2}),$$

$$\mathrm{E}[t_{3,1}^2] = \frac{4}{m\Delta^2} + \frac{8}{N} + O(N^{-2}), \qquad \mathrm{E}[t_{3,2}] = \frac{1}{m\Delta^2} + \frac{6}{N} + O(N^{-2}),$$

$$\mathrm{E}[t_{1,1}t_{3,1}] = \frac{2}{m\Delta^2} + \frac{4}{N} + O(N^{-2}), \qquad \mathrm{E}[t_{3,1}t_{2,1}] = \frac{4}{m\Delta^2} + \frac{4}{N} + O(N^{-2}).$$

Then,

$$F_1(\Delta) = \frac{\Delta}{16}\left\{\frac{1}{N_1} + \frac{1}{N_2}\right\} + O(N^{-2}).$$

Therefore, Corollary 1 can be proved immediately from the above results.

A.5. Calculation of $G_1(\Delta)$. D_s^2 can be expanded as

$$D_s^2 = \Delta^2 + v_1 + O_1,$$

where v_1 is given by

$$v_1 = \frac{1}{m}(\sqrt{f_1}u_1 + 2\sqrt{m}\Delta z_2) - \frac{1}{\sqrt{f_2}}u_2\left(\Delta^2 + \frac{f_1}{m}\right).$$

Then the moment of $F_1(D_s)$ is given by

$$\mathrm{E}[F_1(D_s)] = F_1(\Delta) + F_1'(\Delta)\mathrm{E}[v_1] + O_2 = F_1(\Delta) + O_2.$$

$\hat{v}$ can be expanded as

$$\hat{v} = v(1 + v_{(1)} + v_{(2)} + v_{(3)}) + O_2,$$

where $v_{(j)}$'s are given by

$$v_{(1)} = \left(\xi - \frac{1}{2}\right)t_{2,1}, \qquad v_{(2)} = \left(\frac{3}{8}\xi - \frac{1}{2}\right)t_{2,1}^2 + \left(\xi - \frac{1}{2}\right)t_{2,2},$$

$$v_{(3)} = \left(\xi - \frac{1}{2}\right)t_{2,3} + \left(\frac{3}{4} - \frac{1}{2}\xi\right)t_{2,1}t_{2,2} + \left(\frac{3}{8}\xi - \frac{5}{16}\right)t_{2,1}^3$$

with

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

Then $G_1(\Delta)$ can be obtained by calculating

$$G_1(\Delta) = \nu\left(\mathrm{E}[\nu_{(2)}] - \frac{\nu^2}{2}\mathrm{E}[\nu_{(1)}^2]\right).$$

The necessary moments of $t_{2,j}$'s are given above. Moreover, the moment of $G_1(D_s)$ is given by $\mathrm{E}[G_1(D_s)] = G_1(\Delta) + O_2$.

Acknowledgement

The authors would like to express their gratitude to the referees and the editor for their valuable comments and suggestions.

References

[1] T. W. Anderson, An introduction to multivariate statistical analysis, Third Edision, John Wiley & Sons, Hoboken, New Jersey, 2003.

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

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

[4] 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.

[5] Y. Fujikoshi, V. V. Ulyanv and R. Shimizu, Multivariate statistics: high-dimensional and large-sample approximations, John Wiley & Sons, Hoboken, New Jersey, 2010.

[6] G. J. McLachlan, An asymptotic unbiased technique for estimating the error rates in discriminant analysis, Biometrics, **30** (1974), 239–349.

[7] M. Okamoto, An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Statist., **34** (1963), 1286–1301. (1968, Collected)

[8] S. Raudy, On the amount of a priori information in construction of a classification algorithm, Engrg. Cybernetics, **10** (1972), 710–718.

[9] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern multivariate statistical analysis: a graduate course and handbook, American Sciences Press, Ohio, 1985.

[10] F. J. Wyman, D. M. Young and D. W. Turner, A comparison of asymptotic error rate expansions for the sample linear discriminant function, Pattern Recognition, **23** (1990), 775–783.

Tetsuji Tonda
Faculty of Management and Information Systems
Prefectural University of Hiroshima
Hiroshima 734-8558, *Japan*
E-mail: ttetsuji@pu-hiroshima.ac.jp

Tomoyuki Nakagawa
Depertment of Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526, *Japan*
E-mail: nakagawa.stat@gmail.com

Hirofumi Wakaki
Depertment of Mathematics
Graduate School of Science
Hiroshima University
Higashi-Hiroshima 739-8526, *Japan*
E-mail: wakaki@math.sci.hiroshima-u.ac.jp