EPMC Estimation in Discriminant Analysis when the Dimension and Sample Sizes are Large

Tetsuji Tonda[1], Tomoyuki Nakagawa[2,*] and Hirofumi Wakaki[2]

(Last modified: March 30, 2016)

[1] *Faculty of Management and Information Systems, Prefectural University of Hiroshima 1-1-71, Ujina-Higashi, Minami-ku, Hiroshima 734-8553, Japan*
[2] *Department of Mathematics, Graduate School of Science, Hiroshima University 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan*

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 study a numerical comparison for the performance of our estimator with other estimator base on Okamoto (1963), Fujikoshi and Seo (1998). It is shown that the bias and the mean squared error of our estimator is less than the others.

1 Introduction

Let Π_k $(k = 1, 2)$ be two p-variate normal populations with the mean vector $\boldsymbol{\mu}_k$ $(k = 1, 2)$ and the covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}$ is positive definite and these parameters are unknown, that is,

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

Let $\bar{\boldsymbol{X}}_k$ and $\boldsymbol{S}$ be the sample mean vectors and the pooled sample covariance matrix, based on a sample of N_k observations from Π_k $(k = 1, 2)$, respectively.

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

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

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

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

*E-mail: nakagawa.stat@gmail.com

The performance of the classification rule is evaluated by its 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 linear discriminant rule with using the true values of the parameters, $P(2|1) = P(1|2) = \Phi(-\Delta/2)$, where Φ is the distribution function of $N(0,1)$, Δ is the Mahalanobis distance defined by $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. In the case that the parameter $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are unknown, we use the expected probabilities of misclassification (EPMC), i.e.,

$$e(2|1) = \Pr\left(W(\boldsymbol{X}) \leq 0|\boldsymbol{X} \in \Pi_1\right), \quad e(1|2) = \Pr\left(W(\boldsymbol{X}) > 0|\boldsymbol{X} \in \Pi_2\right)$$

In general, it is hard to obtain the exact evaluation of the EPMC's. There are considerable works 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 N_1 and N_2 tend to large and p is fixed, and the type-II approximations are the ones under a framework such that N_1, N_2 and p tend to large. For the type-I approximations, Okamoto (1963) gave an asymptotic expansion for the EPMC of $W(\boldsymbol{X})$. Moreover, McLachlan (1974) gave asymptotic unbiased estimator of the EPMC up to terms of $O(N^{-2})$ where $N = N_1 + N_2$. For the type-II approximations, Deev (1970) gave an asymptotic expansion for the EPMC of $W(\boldsymbol{X})$ in the case $N_1 = N_2$. Wyman et al. (1990) is compared the accuracy of several approximations for $W(\boldsymbol{X})$ in the case $N_1 = N_2$, and pointed that the approximation due to Raudys (1972) has overall the best accuracy for the combinations of the parameters considered in their study. Fujikoshi and Seo (1998) gave an asymptotic approximation as an extension of Raudys (1972). Fujikoshi (2000) gave an asymptotic expansion and 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 its estimator for the type-II approximation. In section 4 we compared the performances of our estimator with other methods based on Fujikoshi and Seo (1998) and Okamoto (1963). In section 5 we present a discussion and our conclusions.

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

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

Then, it is easily seen that

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

The EPMC's of the classification rule are given by

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

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 $K = (\bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S})$ 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

$$P_K(2|1) = \Phi(T), \qquad T = -\frac{\mu_1(K)}{\sigma(K)}, \tag{2.1}$$

where

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

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 $\boldsymbol{Z} = \sqrt{m}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$, $\boldsymbol{A} = n\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}$ and

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

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

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

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

$$T = -\frac{1}{\sqrt{mN}}T_3^{-1/2}\left\{N_2T_1 + \frac{1}{2}(N_1 - N_2)T_2\right\} + \frac{1}{\sqrt{N}}z_1 \tag{2.2}$$

where

$$T_1 = \boldsymbol{\delta}'\boldsymbol{A}^{-1}\boldsymbol{Z}, \quad T_2 = \boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}, \quad T_3 = \boldsymbol{Z}'\boldsymbol{A}^{-2}\boldsymbol{Z}. \tag{2.3}$$

By using the similar distribution reduction in Fujikoshi and Seo (1998), we have the following lemma.

Lemma 2.1. *Suppose that $n - p + 1 > 0$. Then the statistic (T_1, T_2, T_3) in (2.3) can be expressed in the term of independent standard normal variable z_i $(i = 2, 3)$ and chi-squared variables y_i $(i = 1, ..., 5)$ with f_i degrees of freedom as follow,*

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

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

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 variables y_j's and z_j's, i.e., $T = T(y_1, ..., y_5, z_1, z_2, z_3)$. 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 the term of $O_{3/2}$ can be obtained as follow

$$T = T_{(0)} + T_{(1)} + T_{(2)} + T_{(3)} + O_2 \tag{2.4}$$

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

Theorem 2.1. *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*

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

where

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

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

3 Derivation of the estimator Q_{TW}

Under the type-I framework, several estimating techniques of EPMC are reviewed in Siotani et al. (1985). McLachlan (1974) derived a higher order asymptotic unbiased estimator by using asymptotic expansions. In this section, by using the similar technique in McLachlan (1974), we derive a higher order asymptotic unbiased estimator under the type-II framework.

We consider a following estimator for EPMC,

$$Q_{TW} = \Phi(\hat{\nu}) + Q_1, \quad \hat{\nu} = \nu\left(D_s^2\right),$$

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

$$\mathrm{Bias}(Q_{TW}) = E_K[P_K(2|1) - Q_{TW}] = e(2|1) - \mathrm{E}[\Phi(\hat{\nu})] - Q_1.$$

From Theorem 2.1, $e(2|1)$ can be expanded as $\Phi(\nu) + \phi(\nu)F_1(\Delta) + O_2$, and the expansion of $\mathrm{E}[\Phi(\hat{\nu})]$ up to the term of O_1 is given in the following lemma.

Lemma 3.1. *Suppose that* $\boldsymbol{X}$ *comes from* $\Pi_1 : N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$*. Then, under the type-II framework,* $\mathrm{E}[\Phi(\hat{\nu})]$ *can be expanded as*

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

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

From Theorem 2.1 and Lemma 3.1, it follow that

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

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

$$Q_{TW} = \Phi(\hat{\nu}) + \hat{Q}_1, \quad \hat{Q}_1 = \phi(\hat{\nu})\{F_1(D_s) - G_1(D_s)\} \tag{3.1}$$

is asymptotically unbiased up to the term of O_1. We call this estimating technique TW method.

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

$$\hat{\nu} = \nu(1 + \nu_{(1)} + \nu_{(2)} + \nu_{(3)}) + O_2,$$

where $\nu_{(i)}$'s are given in appendix. Then the variance of our estimator is given by the following.

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

Thus, we have the MSE of our estimator as follows.

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

Therefore, the mean squared error of our estimator converges to 0 in 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 $\boldsymbol{\mu}_1 = (-\Delta/2, 0, ..., 0)'$, $\boldsymbol{\mu}_2 = (\Delta/2, 0, ..., 0)'$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_p$. Let $e_{OK}(2|1;\Delta)$ denotes the asymptotic expansion up to the second order with respect to $(N_1^{-1}, N_2^{-1}, n^{-1})$ due to Okamoto (1963, 1968). For the type-II approximation, Fujikoshi and Seo (1998) gave the asymptotic approximation defined by $e_{FS}(2|1;\Delta) = \Phi(\gamma)$ where

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

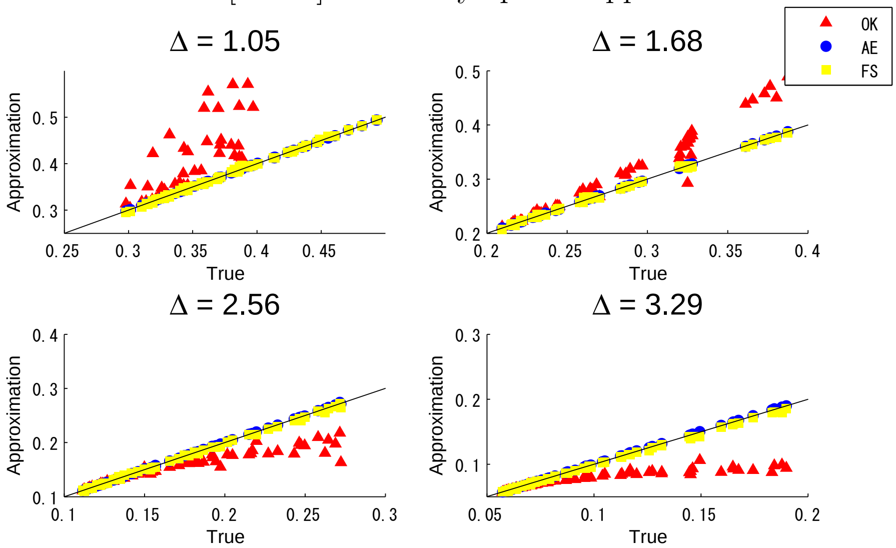
4.1 Comparison of accuracy

The first is a comparison of the accuracy of $e_{AE}(2|1;\Delta)$ with $e_{FS}(2|1;\Delta)$ and $e_{OK}(2|1;\Delta)$. 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 value of Δ correspond to $0.30, 0.20, 0.10, 0.05$ defined by $\Phi(-\Delta/2)$. For each of the configurations, a corresponding EPMC, $e(2|1)$, is obtained by Monte Calro simulation: $e(2|1) = B^{-1}\sum_{i=1}^{B} c_i(2|1)$, where $c_i(2|1)$ is the conditional probability of misclassification, defined by (2.1), for 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 (●), square (■), triangle (▲) mark denote the approximation $e_{AE}(2|1;\Delta)$, $e_{FS}(2|1;\Delta)$ and $e_{OK}(2|1;\Delta)$, 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, the approximation $e_{AE}(2|1;\Delta)$ are better than the other ones.

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



4.2 A comparison of performance of EPMC estimators

Next, we compare our estimator in (3.1) with the other estimating methods. Under the type-I approximation framework, McLachlan(1974) suggested an estimating method 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 replace by $\hat{\Delta}^2$, respectively. We call them AE and FS method, respectively. $\hat{\Delta}^2$ is given by

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

$\hat{\Delta}^2$ have consistency for Δ^2 under the both approximation frameworks. Moreover, we call our new estimating method TW method. Because $\Delta^2 \geq 0$, TW should be modified by changing D_s into 0 if $D_s < 0$. The values of N_1, N_2, p and Δ were chosen as follows,

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

The performance of each estimator is evaluated by MSE, $B^{-1}\sum_{i=1}^{B}\{\hat{e}_i(2|1) - e(2|1)\}^2$, where B is the number of iteration in Monte Calro simulation and $\hat{e}_i(2|1)$ denotes a estimation of $e(2|1)$ in i th 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 TW for several configurations of N_1, N_2 and p. From Figure 2 and 3, we can see that M

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

(N_1, N_2)	Δ	$e(2\|1)$	$e_{OK}(2\|1;\Delta)$	$e_{FS}(2\|1;\Delta)$	$e_{AE}(2\|1;\Delta)$
(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

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

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

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

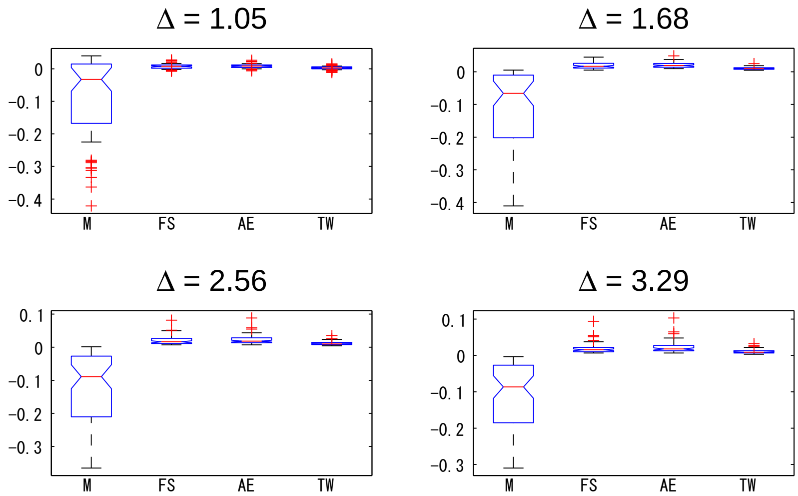


Figure 3: Box plots of the MSE of other estimators $-$ MSE(TW)

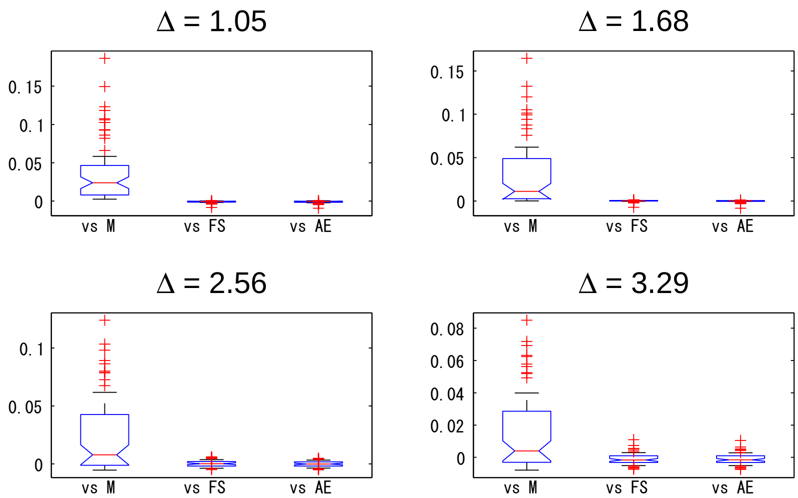


Table 2: Bias of M, FS, AE and TW in the case $p/N = 1/5$.

$(N_1,\ N_2)$	Δ	M	FS	AE	TW
(20, 20)	1.05	0.01973	0.01659	0.01618	0.00892
	1.68	0.00218	0.01587	0.01508	0.01012
	2.56	-0.00560	0.01191	0.01179	0.00714
	3.29	-0.00752	0.00951	0.00995	0.00484
(10, 30)	1.05	0.03825	0.02484	0.02147	0.00578
	1.68	-0.00158	0.02660	0.02338	0.01100
	2.56	-0.01472	0.01798	0.01682	0.00823
	3.29	-0.01473	0.01329	0.01367	0.00591
(30, 10)	1.05	0.00995	0.00813	0.01137	0.00958
	1.68	0.00230	0.01164	0.01435	0.01106
	2.56	-0.00253	0.00995	0.01206	0.00777
	3.29	-0.00460	0.00858	0.01037	0.00537

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

$(N_1,\ N_2)$	Δ	M	FS	AE	TW
(20, 20)	1.05	-0.30721	-0.00037	0.00114	-0.00025
	1.68	-0.32105	0.01798	0.02065	0.01088
	2.56	-0.30114	0.02599	0.03115	0.01601
	3.29	-0.26048	0.02585	0.03332	0.01733
(10, 30)	1.05	-0.41768	-0.00309	-0.00423	-0.00969
	1.68	-0.41237	0.02160	0.02177	0.00487
	2.56	-0.36578	0.03950	0.04268	0.01591
	3.29	-0.30974	0.04136	0.04768	0.01914
(30, 10)	1.05	-0.28066	-0.00787	-0.00352	0.00353
	1.68	-0.28670	0.00698	0.01215	0.01162
	2.56	-0.26635	0.01862	0.02559	0.01714
	3.29	-0.23155	0.02131	0.03003	0.01838

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

(N_1, N_2)	Δ	M	FS	AE
(20, 20)	1.05	0.527	0.028	0.034
	1.68	0.064	0.031	0.027
	2.56	0.005	0.010	0.005
	3.29	-0.005	0.008	0.006
(10, 30)	1.05	2.324	0.059	0.049
	1.68	0.296	0.139	0.111
	2.56	0.020	0.056	0.034
	3.29	-0.012	0.028	0.021
(30, 10)	1.05	0.246	-0.038	-0.027
	1.68	0.018	-0.008	0.005
	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(TW) in the case $p/N = 4/5$.

(N_1, N_2)	Δ	M	FS	AE
(20, 20)	1.05	10.750	-0.142	-0.173
	1.68	10.498	-0.047	-0.086
	2.56	8.390	0.084	0.043
	3.29	5.933	0.078	0.047
(10, 30)	1.05	18.451	-0.311	-0.371
	1.68	16.614	-0.150	-0.242
	2.56	12.081	0.189	0.078
	3.29	8.199	0.282	0.189
(30, 10)	1.05	8.611	-0.131	-0.159
	1.68	8.246	-0.122	-0.147
	2.56	6.512	-0.036	-0.055
	3.29	4.633	-0.006	-0.015

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

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

where $N = N_1 + N_2$, $n = N - 2$.

Proof. D^2 can be expressed in the following.

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

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 $\mathrm{Var}(\tilde{\Delta}^2) \to 0$. From the above $\tilde{\Delta}^2$ has consistency for Δ^2. □

Hence, we can easily show that $\hat{\Delta}^2$ and D_s^2 have consistency for Δ^2.

A.2 Proof of Lemma 2.1

Suppose that the $p \times p$ orthogonal matrices $\boldsymbol{H}$, $\boldsymbol{Q}$ are given by

$$\boldsymbol{H} = \left((\boldsymbol{z}'\boldsymbol{z})^{-1/2}\boldsymbol{z}, \{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_{\boldsymbol{z}})\boldsymbol{\delta}\}^{-1/2}(\boldsymbol{I}_p - \boldsymbol{\Pi}_{\boldsymbol{z}})\boldsymbol{\delta}, \boldsymbol{H}_1\right), \quad \boldsymbol{Q} = \left((\boldsymbol{\delta}'\boldsymbol{\delta})^{-1/2}\boldsymbol{\delta}, \boldsymbol{Q}_1\right).$$

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

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

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

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

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 $\tilde{A}_{11.2}$, $\tilde{\boldsymbol{A}}_{22}$, $\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21}$ are distributed as χ^2_{n-p+1}, $W_{p-1}(n, \boldsymbol{I}_{p-1})$, $N_{p-1}(\boldsymbol{0}, \boldsymbol{I}_{p-1})$, respectively. From above results,

$$
\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} \\
&= \left((\boldsymbol{z}'\boldsymbol{z})^{-1/2}(\boldsymbol{\delta}'\boldsymbol{z}), \{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_{\boldsymbol{z}})\boldsymbol{\delta}\}^{1/2}, \boldsymbol{0}'\right) \tilde{\boldsymbol{A}}^{-1} \begin{pmatrix} (\boldsymbol{z}'\boldsymbol{z})^{1/2} \\ \boldsymbol{0} \end{pmatrix} \\
&= \tilde{A}_{11.2}^{-1}\left\{\boldsymbol{\delta}'\boldsymbol{z} - (\boldsymbol{z}'\boldsymbol{z})^{1/2}\{\boldsymbol{\delta}'(\boldsymbol{I}_p - \boldsymbol{\Pi}_{\boldsymbol{z}})\boldsymbol{\delta}\}^{1/2}\boldsymbol{e}_1'\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\right\}, \quad (\boldsymbol{e}_1 = (1, 0, ..., 0)' : (p-1)\times 1) \\
&= \tilde{A}_{11.2}^{-1}\Bigg[\boldsymbol{\delta}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z} - \left\{m\Delta^2(\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z}) - (\boldsymbol{\delta}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z})^2\right\}^{1/2} \\
&\qquad\qquad \times \frac{\boldsymbol{e}_1'\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}}{\left(\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21}\right)^{1/2}}\left\{\left(\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1/2}\right)\tilde{\boldsymbol{A}}_{22}^{-1}\left(\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21}\right)\right\}^{1/2}\Bigg] \\
&= \frac{1}{y_2}\Bigg[\sqrt{m}\Delta z_2 + m\Delta^2 \\
&\qquad - \left\{m\Delta^2\left(y_1 + (z_2 + \sqrt{m}\Delta)^2\right) - \left(\sqrt{m}\Delta z_2 + m\Delta^2\right)^2\right\}^{1/2} \frac{\tilde{z}_3}{(y_5 + \tilde{z}_3^2)^{1/2}}\left(\frac{y_3}{y_4}\right)^{1/2}\Bigg] \\
&= \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\}, (z_3 = -\tilde{z}_3, \tilde{z}_3 \sim N(0,1))
\end{aligned}
$$

$$
\begin{aligned}
T_2 &= \boldsymbol{z}'\boldsymbol{A}^{-1}\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{A}^{-1}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{z} \\
&= \tilde{A}_{11.2}^{-1}(\boldsymbol{z}'\boldsymbol{z}) = \tilde{A}_{11.2}^{-1}(\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z}) = \frac{y_1 + (z_2 + \sqrt{m}\Delta)^2}{y_2}, \\
T_3 &= \boldsymbol{z}'\boldsymbol{A}^{-2}\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{H}\boldsymbol{H}'\boldsymbol{A}^{-2}\boldsymbol{H}\boldsymbol{H}'\boldsymbol{z} \\
&= (\boldsymbol{z}'\boldsymbol{z})\tilde{A}_{11.2}^{-2}(1 + \tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-2}\tilde{\boldsymbol{A}}_{21}) = (\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{z})\tilde{A}_{11.2}^{-2}\left\{1 + \left(\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1/2}\right)\tilde{\boldsymbol{A}}_{22}^{-1}\left(\tilde{\boldsymbol{A}}_{22}^{-1/2}\tilde{\boldsymbol{A}}_{21}\right)\right\} \\
&= \frac{y_1 + (z_2 + \sqrt{m}\Delta)^2}{y_2^2}\left(1 + \frac{y_3}{y_4}\right).
\end{aligned}
$$

A.3 Calculation of $F_1(\Delta)$

The expansion of (T_1, T_2, T_3) up to the term of O_1 can be given by

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

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

$$
\begin{aligned}
t_{1,0} &= \frac{m\Delta^2}{f_2}, \quad t_{1,1} = \frac{1}{\sqrt{m}\Delta}\left(z_2 + z_3\sqrt{\frac{f_1 f_3}{f_4 f_5}}\right) - \frac{u_2}{\sqrt{f_2}}, \\
t_{1,2} &= \frac{z_3}{2\sqrt{m}\Delta}\sqrt{\frac{f_1 f_3}{f_4 f_5}}\left(\frac{u_1}{\sqrt{f_1}} + \frac{u_3}{\sqrt{f_3}} - \frac{u_4}{\sqrt{f_4}} - \frac{u_5}{\sqrt{f_5}}\right) + \frac{u_2^2}{f_2} - \frac{u_2}{\sqrt{m f_2}\Delta}\left(z_2 + z_3\sqrt{\frac{f_1 f_3}{f_4 f_5}}\right)
\end{aligned}
$$

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

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

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

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

$$
t_{3,0}=\frac{(f_1+m\Delta^2)(f_3+f_4)}{f_2^2f_4},\quad t_{3,1}=\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}-\frac{2u_2}{\sqrt{f_2}}+\frac{\sqrt{f_3}u_3+\sqrt{f_4}u_4}{f_3+f_4}-\frac{u_4}{\sqrt{f_4}},
$$

$$
\begin{aligned}
t_{3,2}=&\frac{z_2^2}{f_1+m\Delta^2}+\frac{3u_2^2}{f_2}-\frac{2\sqrt{f_1}u_1+4\sqrt{m}\Delta z_2}{\sqrt{f_2}(f_1+m\Delta^2)}+\frac{f_3u_4^2}{f_4(f_3+f_4)} \\
&-\frac{\sqrt{f_3}u_3u_4}{(f_3+f_4)\sqrt{f_4}}+\frac{\sqrt{f_3}u_3}{f_3+f_4}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta}\right) \\
&-\frac{2\sqrt{f_3}u_2u_3}{(f_3+f_4)\sqrt{f_2}}-\frac{f_3u_4}{(f_3+f_4)}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta}\right)+\frac{f_3u_2u_4}{(f_3+f_4)\sqrt{f_2f_4}},
\end{aligned}
$$

$$
\begin{aligned}
t_{3,3}=&\frac{3}{f_2}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}\right)-\frac{2u_2z_2}{(f_1+m\Delta^2)\sqrt{f_2}}-\frac{4u_2^3}{f_2\sqrt{f_2}}+\frac{\sqrt{f_3}u_3u_4^2}{f_4(f_3+f_4)}-\frac{f_3u_4^3}{(f_3+f_4)f_4\sqrt{f_4}} \\
&+\frac{f_3u_4^2}{f_4(f_3+f_4)}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta^2z_2}{f_1+m\Delta^2}\right)-\frac{2f_3u_2u_4^2}{f_4(f_3+f_4)\sqrt{f_2}} \\
&-\frac{\sqrt{f_3}u_3u_4}{(f_3+f_4)\sqrt{f_4}}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}\right)+\frac{2\sqrt{f_3}u_2u_3u_4}{(f_3+f_4)\sqrt{f_2f_4}} \\
&+\frac{\sqrt{f_3}u_3z_2^2}{(f_3+f_4)(f_1+m\Delta^2)}+\frac{3\sqrt{f_3}u_2^2u_3}{(f_3+f_4)f_2}-\frac{2\sqrt{f_3}u_2u_3}{(f_3+f_4)\sqrt{f_2}}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}\right) \\
&-\frac{f_3u_4z_2^2}{(f_3+f_4)(f_1+m\Delta^2)\sqrt{f_4}}-\frac{3f_3u_2^2u_4}{(f_3+f_4)f_2\sqrt{f_4}}+\frac{2f_3u_2u_4}{(f_3+f_4)\sqrt{f_2f_4}}\left(\frac{\sqrt{f_1}u_1+2\sqrt{m}\Delta z_2}{f_1+m\Delta^2}\right).
\end{aligned}
$$

Using these expressions, $T_{(j)}$'s in (2.4) can be written as

$$
\begin{aligned}
T_{(0)} &= a_1+a_2 \\
T_{(1)} &= -\frac{T_{(0)}}{2}t_{3,1}+a_1t_{1,1}+a_2t_{2,1}+\frac{1}{\sqrt{N}}z_1 \\
T_{(2)} &= T_{(0)}\left(\frac{3}{8}t_{3,1}^2-\frac{1}{2}t_{3,2}\right)+a_1\left(t_{1,2}-\frac{1}{2}t_{3,1}t_{1,1}\right)+a_2\left(t_{2,2}-\frac{1}{2}t_{3,1}t_{2,1}\right) \\
T_{(3)} &= T_{(0)}\left(-\frac{5}{16}t_{3,1}^3+\frac{3}{4}t_{3,1}t_{3,2}-\frac{1}{2}t_{3,3}\right) \\
&\quad +a_1\left(t_{1,3}-\frac{1}{2}t_{3,1}t_{1,2}+\frac{3}{8}t_{1,1}t_{3,1}^2-\frac{1}{2}t_{1,1}t_{3,2}\right)
\end{aligned}
$$

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

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 the following expectation

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

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

$$\mathrm{E}[t_{1,1}] = \mathrm{E}[t_{2,1}] = \mathrm{E}[t_{3,1}] = 0,\quad \mathrm{E}[t_{1,2}] = \frac{2}{f_2}, \mathrm{E}[t_{2,2}] = \frac{1}{f_1+m\Delta^2}+\frac{2}{f_2},$$
$$\mathrm{E}[t_{3,2}] = \frac{1}{f_1+m\Delta^2}+\frac{6}{f_2}+\frac{2f_3}{f_4(f_3+f_4)},\quad \mathrm{E}[t_{1,3}] = 0,\quad \mathrm{E}[t_{2,3}] = \mathrm{E}[t_{3,3}] = O_2$$
$$\mathrm{E}[t_{1,1}^2] = \frac{1}{m\Delta^2}\left(1+\frac{f_1f_3}{f_4f_5}\right)+\frac{2}{f_2},\quad \mathrm{E}[t_{2,1}^2] = \frac{2f_1+4m\Delta^2}{(f_1+m\Delta^2)^2}+\frac{2}{f_2},$$
$$\mathrm{E}[t_{3,1}^2] = \frac{2f_1+4m\Delta^2}{(f_1+m\Delta^2)^2}+\frac{8}{f_2}+\frac{2f_3}{f_4(f_3+f_4)},$$
$$\mathrm{E}[t_{1,1}t_{2,1}] = \frac{2}{f_1+m\Delta^2}+\frac{2}{f_2},\quad \mathrm{E}[t_{1,1}t_{3,1}] = \frac{2}{f_1+m\Delta^2}+\frac{4}{f_2},\quad \mathrm{E}[t_{2,1}t_{3,1}] = \frac{2f_1+4m\Delta^2}{(f_1+m\Delta^2)^2}+\frac{4}{f_2},$$

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

A.4 Calculation of $G_1(\Delta)$

$\hat{\Delta}^2$ can be expanded as

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

where v_1 is given by

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

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

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

$\hat{\nu}$ can be expanded as

$$\hat{\nu} = \nu(1+\nu_{(1)}+\nu_{(2)}+\nu_{(3)}) + O_2$$

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

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

where

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

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

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

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

References

[1] Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edision. Wiley, Hoboken, NJ.

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

[3] Fujikoshi, Y. (2000), 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**, 1-17.

[4] Fujikoshi, Y., Seo, T. (1998), Asymptotic approximations 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**, 269-280.

[5] Fujikoshi, Y.,Ulyanv, V. V. and Shimizu, R. (2010), *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, NJ.

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

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

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

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

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