Constrained linear discriminant rule for 2-groups via the Studentized classification statistic W for large dimension

Takayuki Yamada[1]

Institute for Comprehensive Education, Center of General Education, Kagoshima University, 1-21-30 Korimoto, Kagoshima 890-0065 Japan

Abstract

This paper is concerned with 2-group linear discriminant analysis for multivariate normal populations with unknown mean vectors and unknown common covariance matrix for the case in which the sample sizes N_1, N_2 and the dimension p are large. We give Studentized version of the W statistic under the high-dimensional asymptotic framework A1 that N_1, N_2, and p tend to infinity together under the condition that $p/(N_1 + N_2 - 2)$ converges to a constant in $(0, 1)$, and N_1/N_2 converges to a constant in $(0, \infty)$. Asymptotic expansion of the distribution for the conditional probability of misclassification(CPMC) of the Studentized W is derived under A1. By using this asymptotic expansion, we give the cut-off point such that the one of two CPMCs is less than the presetting value. Such the constrained discriminant rule is studied by Anderson (1973) and McLachlan (1977). Monte Carlo simulation revealed that the proposed method is more accurate than McLachlan (1977)'s method for the case in which p is relatively large.

AMS 2000 subject classification: primary 62H30; secondary 62H12
Key Words and Phrases: Discriminant analysis, W classification statistic, location and scale mixtured expression, CPMC, High-dimensional asymptotic results, Upper confidence bounds on CPMC.

1 Introduction

Let $\boldsymbol{x}_{ij}$ $(j = 1, \ldots, N_i, i = 1, 2)$ be the jth sample observation $(j = 1, \ldots, N_i)$ from the ith population Π_i $(i = 1, 2)$ with mean $\boldsymbol{\mu}_i$ and common covariance matrix $\boldsymbol{\Sigma}$. We consider the problem to allocate an observation vector $\boldsymbol{x}$ which is according to either Π_1 or Π_2. A commonly used rule is that

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\boldsymbol{S}^{-1}\left\{\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\right\} < c\ (> c) \Rightarrow \text{allocate } \boldsymbol{x} \text{ as } \Pi_2(\Pi_1),$$

which is called the linear discriminant rule, where c is a cut-off point, $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$ and $\boldsymbol{S}$ are the sample mean vectors and the pooled sample covariance matrix defined by

$$\bar{\boldsymbol{x}}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{x}_{ij}, \quad i = 1, 2, \quad \boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)',$$
$$n = N - 2 = N_1 + N_2 - 2.$$

There are two types of probability of misclassification. One is the probability of allocating $\boldsymbol{x}$ into Π_2 even though it is actually belonging to Π_1. The other is the probability that $\boldsymbol{x}$ is classified as Π_1 although it is actually belonging to Π_2. These two types of expected probabilities of misclassifications (EPMCs) for W-rule are expressed as

$$e_{2|1}(c) = P(W < c|\boldsymbol{x} \in \Pi_1) \quad \text{and} \quad e_{1|2}(c) = P(W > c|\boldsymbol{x} \in \Pi_2).$$

[1] E-mail address:yamada@gm.kagoshima-u.ac.jp

In general, it is hard to evaluate these expected probabilities of misclassification (EPMC) explicitly, but some asymptotic results including asymptotic expansions have been obtained. Anderson [1] derived an asymptotic expansion for Studentized W, and applied it to identify c such that

$$e_{2|1}(c) = 1 - \varepsilon + O(n^{-1}),$$

where $\varepsilon \in (0,1)$ is a presetting level which is given by experimenter. This discriminant rule is used to control one of EPMCs for the case in which one type of errors is generally regarded as more serious than the others such as medical applications associated with the diagnosis of diseases. Anderson [1]'s asymptotic expansion is obtained under the asymptotic framework A0:

$$\text{A0}: \; N_1 \to \infty, N_2 \to \infty, N_1/N_2 \to \gamma \in (0,\infty), p \text{ is fixed}.$$

For achieving the same aim with Anderson [1], McLachlan [10] proposed the cut-off point c such that

$$P\left(c_{2|1}(c) < \Xi_{\mathrm{L}}\right) = 1 - \varepsilon + O(n^{-1}),$$

where

$$c_{2|1}(c) = P(W < c | \boldsymbol{x} \in \Pi_1; \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \boldsymbol{S}),$$

$\varepsilon \in (0,1)$ is a presetting level which is set by experimenter and Ξ_{L} is an upper bound which is set by experimenter. These and some other asymptotic results were reviewed by Siotani [13], by McLachlan [11] and by Anderson [2].

Generally, the precision of asymptotic approximations under A0 gets worth as the dimension p becomes large. As an alternative approach to overcome this shortcoming, it has been considered to derive asymptotic distributions of discriminant functions in a high-dimensional situation where n and p tend to infinity together. Yamada et al. [14] derived an asymptotic expansion of Studentized W under the high-dimensional asymptotic framework A1 and the assumption C such that

$$\begin{aligned} \text{A1}: N_1 &\to \infty, \quad N_2 \to \infty, \quad N_1/N_2 \to \gamma \in (0,\infty), \\ p &\to \infty, \quad p/n \to \gamma_0 \in [0,1); \\ \text{C}: \Delta &\to \Delta_0 \in (0,\infty), \end{aligned}$$

where Δ^2 is the squared Mahalanobis distance defined as $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Using the asymptotic expansion, they proposed a cut-off point c such that

$$e_{2|1}(c) = 1 - \varepsilon + O_{\mathrm{A1,C}}(n^{-1}),$$

where the symbol $O_{\mathrm{A1,C}}(n^{-1})$ stands for the term such that $nO_{\mathrm{A1,C}}(n^{-1})$ converges to a constant as $n \to \infty$ under A1 and C. The usefulness such the high-dimensional asymptotic framework is mentioned in Fujikoshi et al. [6].

The aim of this paper is to obtain a cut-off point c_h which satisfies that

$$P\left[c_{2|1}(c_h) < \Xi_{\mathrm{H}}\right] = 1 - \varepsilon + O_{\mathrm{A1,C}}(n^{-1})$$

for the presetting values ε and Ξ_{H}. In order to derive this, we show an asymptotic expansion of the distribution for the statistic $c_{2|1}(c_h)$ under A1 and C. Since the distribution for $c_{1|2}(c)$ is the same as the one for $c_{2|1}(-c)$ with interchanging N_1 and N_2, we only derive asymptotic expansion of the distribution for $c_{2|1}(c)$.

This paper is organized as the following: Section 2 presents Studentization for W under A1. In Section 3, we derive the asymptotic distribution of $c_{2|1}(c)$ via the Studentized statistic W under A1. Asymptotic expansion for the Studentized $c_{2|1}(c)$ is derived by making use of a powerful method known as the method by the differential operator which was used by James [8], Okamoto [12], etc. In Section 4, we propose constrained linear discriminant rule for CPMC for large dimensional case. Simulation results are written in Section 5. We revealed that the proposed methods performs well for the case in which Ξ_{H} is not so small. In Section 6, concluding remarks are written. Some proofs and technical results are given in Appendix.

2 Studentization for W under A1

For $\boldsymbol{x} \in \Pi_i$, it follows from Lachenbruch [9] that

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} \left\{ \boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \right\} = V^{1/2} Z_i - U_i \quad (i = 1, 2), \tag{1}$$

where

$$\begin{aligned} V &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} \boldsymbol{\Sigma} \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\ Z_i &= V^{-1/2} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i), \\ U_i &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_i) - \frac{1}{2} D^2, \end{aligned}$$

and D^2 is the squared sample Mahalanobis distance defined by $D^2 = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Then, we find that V is a positive random variable and (U_i, V) are jointly independent of Z_i. Further, Z_i is distributed as $N(0, 1)$. This normality follows by considering the conditional distribution of Z_i when $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$, and $\boldsymbol{S}$ are given. In this case, W is called a location and scale mixture of the standard normal distribution. It can be expressed that

$$\begin{aligned} E[U_i] &= \frac{(-1)^i}{2} \frac{n}{m-1} \Delta^2 - \frac{1}{2} \frac{p}{m-1} \left(\frac{n}{N_2} - \frac{n}{N_1} \right), \\ E[V] &= \frac{n^2(n+1)}{(m-1)^2(m+2)} \left(\Delta^2 + \frac{Np}{N_1 N_2} \right), \end{aligned}$$

where $m = n - p$. The analytic expressions for $\mathrm{Var}(U_i)$ and for $\mathrm{Var}(V)$, which are provided by Fujikoshi [3], show that $\mathrm{Var}(U_i) \to 0$ and $\mathrm{Var}(V) \to 0$ under A1 and C. It implies that the limiting distribution of W under A1 and C is normal with mean $-u_{0i} = -\lim_{\mathrm{A1}} E[U_i]$ and variance $v_0 = \lim_{\mathrm{A1}} \{E[V]\}$. The natural estimate for $(E[U_i], E[V])$ is obtained by replacing Δ^2 with the following unbiased estimator $\widehat{\Delta^2}$:

$$\widehat{\Delta^2} = \frac{m-1}{n} D^2 - \frac{Np}{N_1 N_2},$$

and we write it as $(\widehat{E[U_i]}, \widehat{E[V]})$. The unbiasedness for $(\widehat{E[U_i]}, \widehat{E[V]})$ also holds. We can show that $(\widehat{E[U_i]}, \widehat{E[V]})$ has consistency under A1 and C. From Slutsky's theorem,

$$\frac{W + \widehat{E[U_i]}}{\sqrt{\widehat{E[V]}}} \xrightarrow{\mathcal{D}} N(0, 1) \quad (\text{for } \boldsymbol{x} \in \Pi_i),$$

where the symbol $\xrightarrow{\mathcal{D}}$ stands for the convergence in distribution. From the technical reason, instead of using $(\widehat{E[U_i]}, \widehat{E[V]})$, we use (U_{0i}, V_0) for the Studentization of W in this paper, which is defined as follows.

$$\begin{aligned} U_{0i} &= \widehat{E[U_i]} + \frac{(-1)^i 2(n-1)}{(m+1)(m-1)} \frac{n}{N_i} \\ &= \frac{(-1)^i}{2} \frac{n}{m-1} \widehat{\Delta^2} - \frac{1}{2} \frac{p}{m-1} \left(\frac{n}{N_2} - \frac{n}{N_1} \right) + \frac{(-1)^i 2(n-1)}{(m+1)(m-1)} \frac{n}{N_i}, \\ V_0 &= \frac{m-1}{m+1} \widehat{E[V]} \\ &= \frac{n^2(n+1)}{(m-1)(m+1)(m+2)} \left(\widehat{\Delta^2} + \frac{Np}{N_1 N_2} \right). \end{aligned}$$

It is noted that (U_{0i}, V_0) is not the unbiased estimator of $(E[U_i], E[V])$, but has has the consistency under A1 and C. We can also show that

$$W_i^* = \frac{W + U_{0i}}{\sqrt{V_0}} \xrightarrow{\mathcal{D}} N(0, 1) \quad (\text{for } \boldsymbol{x} \in \Pi_i).$$

3 Asymptotic distribution for the Studentized CPMC

Let c_j be the cut-off point for W_j^* and let $C_{i|j}$ denote the conditional probability of misclassification of W_j^* misallocating an observation from Π_j, where $i \neq j$. Then $C_{i|j}$ is given by

$$
\begin{aligned}
C_{i|j} &= c_{i|j}\left(\sqrt{V_0}c_j - U_{0j}\right) \\
&= P\left((-1)^i W < (-1)^i\left(\sqrt{V_0}c_j - U_{0j}\right) | \boldsymbol{x} \in \Pi_j, \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \boldsymbol{S}\right) \\
&= \Phi\left((-1)^i \frac{\sqrt{V_0}c_j - U_{0j} + U_j}{\sqrt{V}}\right) \\
&= \begin{cases} \Phi\left(\sqrt{\dfrac{V_0}{V}}(-c_2) - \sqrt{\dfrac{1}{V}}\left\{-(U_{02} - U_2)\right\}\right) & (i = 1, j = 2) \\ \Phi\left(\sqrt{\dfrac{V_0}{V}}c_1 - \sqrt{\dfrac{1}{V}}\left\{(U_{01} - U_1)\right\}\right) & (i = 2, j = 1). \end{cases}
\end{aligned}
$$

From Lemma 2 and 3 in Appendix A, the distribution for $C_{2|1}$ is identical to $C_{1|2}$ if $c_1 = -c_2$. From that reason, we only deal with $C_{2|1}$. Asymptotic expansion of the distribution for $C_{1|2}$ can be obtained by the one for $C_{2|1}$ by replacing (N_1, N_2, c_1) with $(N_2, N_1, -c_2)$. Hereafter, we set $U_0 = U_{01}$, $U = U_1$ and $c = c_1$, unless making confusion.

3.1 Stochastic expression for CPMC

Now, we consider to express $C_{2|1}$ as the function of simple variables. Let

$$
\begin{aligned}
\boldsymbol{u}_1 &= \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1/2} \boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\
\boldsymbol{u}_2 &= \frac{1}{\sqrt{N}}\boldsymbol{\Sigma}^{-1/2}(N_1\bar{\boldsymbol{x}}_1 + N_2\bar{\boldsymbol{x}}_2 - N_1\boldsymbol{\mu}_1 - N_2\boldsymbol{\mu}_2), \\
\boldsymbol{B} &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}.
\end{aligned}
$$

Then $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{B}$ are independent. In addition, $\boldsymbol{u}_1 \sim N_p((1/N_1 + 1/N_2)^{-1/2}\boldsymbol{\delta}, \boldsymbol{I}_p)$ and $\boldsymbol{u}_2 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, where $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. It also holds that $n\boldsymbol{B}$ is distributed as a Wishart distribution with n degrees of freedom and covariance matrix $\boldsymbol{I}_p$, which is denoted as $W_p(n, \boldsymbol{I}_p)$. Substituting them, we have

$$
\begin{aligned}
U &= -\frac{1}{2}\left(\frac{n}{N_2} - \frac{n}{N_1}\right)\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_1}{n} + \frac{n}{\sqrt{N_1N_2}}\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_2}{n} - \sqrt{\frac{nN_2}{NN_1}}\frac{\boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}_1}{\sqrt{n}}, \\
V &= \frac{Nn}{N_1N_2}\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-2}\boldsymbol{u}_1}{n}.
\end{aligned}
$$

It is also described that

$$
\widehat{\Delta^2} = \frac{N(m-1)}{N_1N_2}\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_1}{n} - \frac{Np}{N_1N_2}.
$$

Using this expression, we can write

$$
\begin{aligned}
U_0 &= -\frac{1}{2}\frac{Nn}{N_1N_2}\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_1}{n} + \frac{p-2}{m+1}\frac{n}{N_1}, \\
V_0 &= \frac{n(n+1)}{(m+1)(m+2)}\frac{Nn}{N_1N_2}\frac{\boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_1}{n}.
\end{aligned}
$$

The following lemma, which is given by Yamada et al. [14], enables to see the functions of the independent standard normal and chi-squared variables for U, V, U_0 and V_0.

Lemma 1. *Let* $\boldsymbol{v}_1 \sim N_p(\boldsymbol{a}, \boldsymbol{I}_p)$, $\boldsymbol{v}_2 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $\boldsymbol{W} \sim W_p(n, \boldsymbol{I}_p)$, *and* $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ *and* $\boldsymbol{W}$ *are independent. Then*

$$\begin{pmatrix} \boldsymbol{a}'\boldsymbol{W}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_2'\boldsymbol{W}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_1'\boldsymbol{W}^{-1}\boldsymbol{v}_1 \\ \boldsymbol{v}_1'\boldsymbol{W}^{-2}\boldsymbol{v}_1 \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \dfrac{\sqrt{\boldsymbol{a}'\boldsymbol{a}}}{\mathcal{X}_1}\left(Z_1 + \sqrt{\boldsymbol{a}'\boldsymbol{a}} - \sqrt{\dfrac{\mathcal{X}_2}{\mathcal{X}_3}} Z_2\right) \\ \sqrt{\dfrac{1}{\mathcal{X}_1^2}\left(1+\dfrac{\mathcal{X}_2}{\mathcal{X}_3}\right)\{(Z_1+\sqrt{\boldsymbol{a}'\boldsymbol{a}})^2 + Z_2^2 + \mathcal{X}_4\} Z_3} \\ \dfrac{1}{\mathcal{X}_1}\{(Z_1+\sqrt{\boldsymbol{a}'\boldsymbol{a}})^2 + Z_2^2 + \mathcal{X}_4\} \\ \dfrac{1}{\mathcal{X}_1^2}\left(1+\dfrac{\mathcal{X}_2}{\mathcal{X}_3}\right)\{(Z_1+\sqrt{\boldsymbol{a}'\boldsymbol{a}})^2 + Z_2^2 + \mathcal{X}_4\} \end{pmatrix},$$

where $\mathcal{X}_i \sim \chi^2_{f_i}$, $i = 1, 2, 3, 4$; $Z_i \sim N(0, 1)$, $i = 1, 2, 3$; *all the seven variables* $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{X}_3$. $\mathcal{X}_4$, Z_1, Z_2, Z_3 *are independent;*

$$f_1 = n - p + 1, \ \ f_2 = p - 1, \ \ f_3 = n - p + 2, \ \ f_4 = p - 2.$$

Similar results to Lemma 1 was treated in Fujikoshi and Seo [5], Fujikoshi [4], and Hyodo and Kubokawa [7].

Put

$$\begin{aligned} b_1 &= b_1(w_1, w_2, w_3, z_1, z_2) = \frac{n}{f_1}\frac{\Delta}{1+w_1}\left(z_1 + \sqrt{\frac{N_1N_2}{Nn}}\Delta - \sqrt{\frac{f_2}{f_3}}\sqrt{t} z_2\right), \\ b_2 &= b_2(w_1, w_2, w_3, w_4, z_1, z_2, z_3) = \frac{n}{f_1}\frac{1}{1+w_1}\sqrt{\left(1+\frac{f_2}{f_3}t\right) s} z_3, \\ q_1 &= q_1(w_1, w_4, z_1, z_2) = \frac{n}{f_1}\frac{1}{1+w_1} s, \\ q_2 &= q_2(w_1, w_2, w_3, w_4, z_1, z_2) = \left(\frac{n}{f_1}\right)^2 \left(\frac{1}{1+w_1}\right)^2 \left(1 + \frac{f_2}{f_3}t\right) s, \end{aligned}$$

where

$$\begin{aligned} s &= s(w_4, z_1, z_2) = \left(z_1 + \sqrt{\frac{N_1N_2}{Nn}}\Delta\right)^2 + z_2^2 + \frac{f_4}{n}(1+w_4), \\ t &= t(w_2, w_3) = \frac{1+w_2}{1+w_3}; \end{aligned}$$

f_1, f_2, f_3, f_4 are defined in Lemma 1. Then we have

$$\begin{aligned} \begin{pmatrix} B_1 \\ B_2 \\ Q_1 \\ Q_2 \end{pmatrix} &\stackrel{\text{def}}{=} \frac{1}{n}\begin{pmatrix} \boldsymbol{\delta}'\boldsymbol{B}^{-1}\boldsymbol{u}_1 \\ \boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_2 \\ \boldsymbol{u}_1'\boldsymbol{B}^{-1}\boldsymbol{u}_1 \\ \boldsymbol{u}_1'\boldsymbol{B}^{-2}\boldsymbol{u}_1 \end{pmatrix} \\ &\stackrel{\mathcal{D}}{=} \begin{pmatrix} b_1(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, Z_1/\sqrt{n}, Z_2/\sqrt{n}) \\ b_2(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}) \\ q_1(\sqrt{2/f_1}W_1, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}) \\ q_2(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}) \end{pmatrix}, \end{aligned} \tag{2}$$

where $W_i = \sqrt{f_i/2}(\mathcal{X}_i/f_i - 1)$, $i = 1, \ldots, 4$. This implies that

$$\begin{pmatrix} U - U_0 \\ V_0 \\ V \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} -\alpha_4 & \alpha_3 & \alpha_1 & 0 \\ 0 & 0 & \beta_1 & 0 \\ 0 & 0 & 0 & \beta_2 \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ Q_1 \\ Q_2 \end{pmatrix} - \begin{pmatrix} \alpha_1\alpha_2 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where

$$\alpha_1 = \frac{1}{2}\left\{\frac{Nn}{N_1N_2} - \left(\frac{n}{N_2} - \frac{n}{N_1}\right)\right\} = \frac{n}{N_1}, \alpha_2 = \frac{p-2}{m+1}, \alpha_3 = \frac{n}{\sqrt{N_1N_2}}, \alpha_4 = \sqrt{\frac{nN_2}{NN_1}},$$
$$\beta_1 = \frac{n(n+1)}{(m+1)(m+2)}\frac{Nn}{N_1N_2}, \beta_2 = \frac{Nn}{N_1N_2}.$$

Letting

$$\Psi(w_1, w_2, w_3, w_4, z_1, z_2, z_3) = \Phi\left(\frac{\sqrt{\beta_1 q_1}c + \alpha_1(q_1 - \alpha_2) + \alpha_3 b_2 - \alpha_4 b_1}{\sqrt{\beta_2 q_2}}\right),$$

we can express that

$$\begin{aligned} C_{2|1} &= \Phi\left(\frac{\sqrt{V_0}c - U_0 + U}{\sqrt{V}}\right) \\ &\stackrel{\mathcal{D}}{=} \Psi\left(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}\right). \end{aligned} \tag{3}$$

Hereafter, we set $\boldsymbol{v}$ as the variable vector and $\boldsymbol{y}$ as the random variable vector, which are defined by

$$\boldsymbol{v} = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \end{pmatrix}' = \begin{pmatrix} w_1 & w_2 & w_3 & w_4 & z_1 & z_2 & z_3 \end{pmatrix}',$$
$$\boldsymbol{y} = \begin{pmatrix} Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 \end{pmatrix}' = \begin{pmatrix} W_1 & W_2 & W_3 & W_4 & Z_1 & Z_2 & Z_3 \end{pmatrix},$$

and

$$\Psi(\boldsymbol{v}) = \Psi(w_1, w_2, w_3, w_4, z_1, z_2, z_3), \tag{4}$$

unless making confusion.

3.2 Studentization for CPMC under A1

It can be expressed that

$$\begin{aligned} &b_1(0,0,0,0,0) = \frac{n}{f_1}\sqrt{\frac{N_1N_2}{Nn}}\Delta^2 = b_{1,0}, \\ &b_2(0,0,0,0,0,0,0) = 0, \\ &q_1(0,0,0,0) = \frac{n}{f_1}\left(\frac{N_1N_2}{Nn}\Delta^2 + \frac{f_4}{n}\right) = q_{1,0}, \\ &q_2(0,0,0,0,0,0) = \left(\frac{n}{f_1}\right)^2\left(1 + \frac{f_2}{f_3}\right)\left(\frac{N_1N_2}{Nn}\Delta^2 + \frac{f_4}{n}\right) = q_{2,0}. \end{aligned}$$

So, we have

$$\Psi(\boldsymbol{0}) = \Phi\left(\frac{\sqrt{\beta_1 q_{1,0}}c + \alpha_1(q_{1,0} - \alpha_2) - \alpha_4 b_{1,0}}{\sqrt{\beta_2 q_{2,0}}}\right) = \Phi(c).$$

It is noted that $\Psi(\boldsymbol{v})$ is the smooth function on $(-1,\infty)\times(-1,\infty)\times(-1,\infty)\times(-1,\infty)\times\mathbb{R}^3$. We will expand

$$\Psi\left(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}\right)$$

at $(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}) = (0,0,0,0,0,0,0)$. Let $\boldsymbol{\psi}_1 = \boldsymbol{\psi}_1(c, \Delta^2)$ be the vector valued function in $\mathbb{R}^7$ defined by

$$\boldsymbol{\psi}_1 = \boldsymbol{D}\frac{\partial}{\partial \boldsymbol{v}}\Psi(\boldsymbol{v})_{|0}, \tag{5}$$

where $\boldsymbol{D} = \mathrm{diag}(\sqrt{2n/f_1}, \sqrt{2n/f_2}, \sqrt{2n/f_3}, \sqrt{2n/f_4}, 1, 1, 1)$, $\Psi(\boldsymbol{v})$ is defined as (4), and the notation "$|_0$" stands for the value at the point that $\boldsymbol{v} = \boldsymbol{0}$. We can express that $\boldsymbol{\psi}_1 = \phi(c)\boldsymbol{p}_1$, where

$$\begin{aligned}\boldsymbol{p}_1 &= \boldsymbol{p}_1(c, \delta^2) \\ &= \frac{c}{2}\left(\sqrt{\tfrac{2n}{f_1}} \quad -\sqrt{\tfrac{2n}{f_2}}\tfrac{f_2}{f_2+f_3} \quad \sqrt{\tfrac{2n}{f_3}}\tfrac{f_2}{f_2+f_3} \quad 0 \quad 0 \quad 0 \quad 0\right)' \\ &\quad + \frac{1}{\sqrt{q_{2,0}\beta_2}}\left(-\sqrt{\tfrac{2n}{f_1}}\tfrac{f_4}{f_1}\alpha_1 \quad 0 \quad 0 \quad \sqrt{\tfrac{2n}{f_4}}\tfrac{f_4}{f_1}\alpha_1 \quad \tfrac{\alpha_1}{\sqrt{\beta_2}}\tfrac{n}{f_1}\Delta \quad \sqrt{\tfrac{f_2}{f_3}}\tfrac{n}{f_1}\alpha_4\Delta \quad \alpha_3\sqrt{q_{2,0}}\right)'. \end{aligned} \tag{6}$$

In addition, let $\boldsymbol{\Psi}_2 = \boldsymbol{\Psi}_2(c, \Delta^2)$ be 7×7 matrix valued function defined by

$$\boldsymbol{\Psi}_2 = \boldsymbol{D}\frac{\partial^2}{\partial\boldsymbol{v}\partial\boldsymbol{v}'}\Psi(\boldsymbol{v})\boldsymbol{D}_{|0}.$$

We can show that

$$\boldsymbol{D}\boldsymbol{\Psi}_2\boldsymbol{D} = \phi(c)\boldsymbol{P},$$

where $\boldsymbol{P} = \boldsymbol{P}(c, \Delta^2)$ is the 7×7 matrix valued function defined as $\boldsymbol{P} = \boldsymbol{P}_0 + c\boldsymbol{P}_1 + c^2\boldsymbol{P}_2 + c^3\boldsymbol{P}_3$ with $\boldsymbol{P}_i = \boldsymbol{P}_i(\Delta^2)$ $(i = 0, 1, 2, 3)$ being the 7×7 symmetric matrix valued function. The analytic form for $\boldsymbol{P}_i$ $(i = 0, 1, 2, 3)$ are given in Appendix C.

It can be expressed that

$$\begin{aligned}&\Psi\left(\sqrt{2/f_1}W_1, \sqrt{2/f_2}W_2, \sqrt{2/f_3}W_3, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}, Z_3/\sqrt{n}\right) \\ &= \Phi(c) + \frac{1}{\sqrt{n}}\boldsymbol{\psi}_1'\boldsymbol{y} + \frac{1}{n}\boldsymbol{y}'\boldsymbol{\Psi}_2\boldsymbol{y} + \frac{1}{n^{3/2}}R_1,\end{aligned} \tag{7}$$

where R_1 is a remainder term consisting of a homogeneous polynomial of order 3 in the elements of $\boldsymbol{y}$ of which the coefficients are $O(1)$ as $n \to \infty$ under A1 and C, plus $n^{-1/2}$ times a homogeneous polynomial of order 4, plus a remainder term that is $O(n^{-1})$ under A1 and C for fixed $\boldsymbol{y}$.

By virtue of (7) with combined the use of the formula (3), we have

$$\frac{C_{2|1} - \Phi(c)}{\sqrt{\boldsymbol{\psi}_1'\boldsymbol{\psi}_1/n}} \overset{\mathcal{D}}{=} \frac{1}{\sqrt{\boldsymbol{\psi}_1'\boldsymbol{\psi}_1}}\left(\boldsymbol{\psi}_1'\boldsymbol{y} + \frac{1}{\sqrt{n}}\boldsymbol{y}'\boldsymbol{\Psi}_2\boldsymbol{y} + \frac{1}{n}R_1\right). \tag{8}$$

It follows from the definition of $\boldsymbol{y}$ that

$$\boldsymbol{y} \xrightarrow{\mathcal{D}} N_7(\boldsymbol{0}, \boldsymbol{I}_7)$$

under A1 and C, which implies the following theorem.

Theorem 1. *Under the high-dimensional asymptotic framework* A1 *and the assumption* C*,*

$$\frac{C_{2|1} - \Phi(c)}{\phi(c)\sqrt{\rho(c, \Delta^2)/n}} \xrightarrow{\mathcal{D}} N(0, 1),$$

where

$$\begin{aligned}\rho(c, \Delta^2) &= \boldsymbol{\psi}_1'\boldsymbol{\psi}_1/\{\phi(c)\}^2 \\ &= \frac{c^2}{2}\left(\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2 + f_3}\right) - 2c\frac{n}{f_1}\frac{f_4}{f_1}\frac{\alpha_1}{\sqrt{\beta_2}}\frac{1}{\sqrt{q_{2,0}}} \\ &\quad + \frac{1}{\beta_2 q_{2,0}}\left\{2\left(\frac{n}{f_1} + \frac{n}{f_4}\right)\left(\frac{f_4}{f_1}\right)^2\alpha_1^2 + \left(1 + \frac{f_2}{f_3}\right)\frac{\alpha_1^2}{\beta_2}\left(\frac{n}{f_1}\right)^2\Delta^2 + \alpha_3^2 q_{2,0}\right\}.\end{aligned} \tag{9}$$

Since $\rho(c, \Delta^2)$ is unknown parameter, it is needed to estimate for Studentization. The natural estimate $\rho(c, \widehat{\Delta^2})$ can not be used. The reason is that $\{q_{2,0}(\Delta^2)\}^{-1/2}$ which is included in $\rho(c, \Delta^2)$ can not by defined for the case in which $0 < D^2 < \{2/(m-1)\}(Nn)/(N_1N_2)$ since $q_{2,0}(\widehat{\Delta^2})$ takes negative value. Instead of using the unbiased estimator $\widehat{\Delta^2}$, we use

$$\widehat{\Delta_A^2} = \frac{m+1}{n} D^2 - \frac{N(p-2)}{N_1 N_2}, \tag{10}$$

which is not the unbiased estimate of Δ^2. It can be expressed that

$$\begin{aligned}\rho(c, \Delta^2) &= \frac{1}{2}\left(\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}\right)\left(c - \frac{2}{\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}}\frac{1}{\sqrt{\beta_2 q_{2,0}}}\frac{n}{f_1}\frac{f_4}{f_1}\alpha_1\right)^2 \\ &\quad + \frac{1}{\beta_2 q_{2,0}}\left(\frac{n}{f_1}\right)^2 \alpha_1^2 \tau(\Delta^2) + \frac{\alpha_3^2}{\beta_2},\end{aligned} \tag{11}$$

where

$$\tau(\Delta^2) = -\frac{2}{\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}}\left(\frac{f_4}{f_1}\right)^2 + 2\left(\frac{n}{f_1} + \frac{n}{f_4}\right)\left(\frac{f_4}{n}\right)^2 + \left(1 + \frac{f_2}{f_3}\right)\frac{\Delta^2}{\beta_2}.$$

It is sufficient to show the positiveness for $\tau(\widehat{\Delta_A^2})$ to ensure that $\rho(c, \widehat{\Delta_A^2}) > 0$. We can express that

$$\begin{aligned}\tau(\widehat{\Delta_A^2}) &= -\frac{2}{\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}}\left(\frac{f_4}{f_1}\right)^2 + 2\left(\frac{n}{f_1} + \frac{n}{f_4}\right)\left(\frac{f_4}{n}\right)^2 + \left(1 + \frac{f_2}{f_3}\right)\frac{1}{\beta_2}\left(\frac{f_1}{n}D^2 - \frac{f_4}{n}\beta_2\right) \\ &= \frac{f_1}{n}\left(1 + \frac{f_2}{f_3}\right)\frac{D^2}{\beta_2} + 2\left(\frac{n}{f_1} + \frac{n}{f_4}\right) - \frac{2}{\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}}\left(\frac{n}{f_1}\right)^2 - \left(1 + \frac{f_2}{f_3}\right)\frac{f_4}{n} \\ &= \left(\frac{f_1}{n}\right)^2 q_{2,0}(\widehat{\Delta_A^2}) + \left[2 + \frac{(f_3 - f_2)\{f_3(f_2+f_3) + f_1 f_2\}}{f_2 f_3 f_4}\right]\frac{f_2 f_4^2}{f_1 f_3 (f_2+f_3)}\frac{1}{\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2+f_3}},\end{aligned} \tag{12}$$

where the last equality follows from the fact that

$$q_{2,0}(\widehat{\Delta_A^2}) = \frac{n}{f_1}\left(1 + \frac{f_2}{f_3}\right)\frac{D^2}{\beta_2}. \tag{13}$$

The non-negativeness for $\tau(\widehat{\Delta_A^2})$ follows from that

$$\begin{aligned}&2f_2f_3f_4 + (f_3 - f_2)\{f_3(f_2+f_3) + f_1f_2\} \\ &= (n-p)^3 + (5+p)(n-p)^2 + (11+p)(n-p) + (p-2)^2 + 7 \\ &> 0.\end{aligned}$$

Note that

$$\frac{\widehat{\Delta_A^2}}{\Delta^2} \xrightarrow{p} 1$$

under A1 and C. From this rate consistency, we obtain that

$$\frac{\rho(c, \widehat{\Delta_A^2})}{\rho(c, \Delta^2)} \xrightarrow{p} 1.$$

By Theorem 1 and Slutsky's theorem,

$$\frac{C_{2|1} - \Phi(c)}{\phi(c)\sqrt{\rho(c, \widehat{\Delta_A^2})/n}} \xrightarrow{\mathcal{D}} N(0,1). \tag{14}$$

3.3 Asymptotic expansion for the distribution of the proposed Studentized statistic for CPMC under A1

In this section, we derive an asymptotic expansion for the distribution for the Studentized $C_{2|1}$ to improve the convergence rate in (14).

Firstly, we give a general result for the cumulative distribution function of the random variable T which has the form:

$$T = \frac{1}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\boldsymbol{h}'\boldsymbol{y} + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\boldsymbol{y}'\boldsymbol{H}\boldsymbol{y} + \frac{1}{n}R \tag{15}$$

for $\boldsymbol{h} \in \mathbb{R}^7$ and the symmetric matrix $\boldsymbol{H}$, where R is the term consisting of a homogeneous polynomial of order 3 in the elements of $\boldsymbol{y}$ of which the coefficients are $O(1)$ under A1 and C, plus $n^{-1/2}$ times a homogeneous polynomial of order 4, plus a remainder term that is $O(n^{-1})$ under A1 and C for fixed $\boldsymbol{y}$.

Theorem 2. *The cumulative distribution function of T which is described as* (15) *can be expressed as*

$$P(T \le x) = \Phi(x) - \frac{1}{\sqrt{n}}(s_1 H_0(x) + s_2 H_2(x))\phi(x) + O(n^{-1}),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, $\phi(\cdot)$ is the derivative of $\Phi(\cdot)$, and $H_k(x)$ denotes the Hermite polynomial of degree k, especially, $H_0(x) = 1$, $H_2(x) = x^2 - 1$. Here,

$$s_1 = \frac{\operatorname{tr}\boldsymbol{H}}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}, \quad s_2 = \frac{1}{(\boldsymbol{h}'\boldsymbol{h})^{3/2}}\left\{\frac{\sqrt{2}}{3}\sum_{k=1}^{4}\sqrt{\frac{n}{f_k}}h_k^3 + \boldsymbol{h}'\boldsymbol{H}\boldsymbol{h}\right\}$$

for $\begin{pmatrix} h_1 & \cdots & h_7 \end{pmatrix}' = \boldsymbol{h}$.

The proof of Theorem 2 is given in Appendix B.

Now, we consider to express the proposed Studentized statistic as the form (15). By virtue of (13) combined with the use of the formula (2) and the fact that $Q_1 = \{N/(N_1N_2)\}D^2$, we have

$$q_{2,0}(\widehat{\Delta^2}) \overset{\mathcal{D}}{=} \frac{n}{f_1}\left(1 + \frac{f_2}{f_3}\right) q_1(\sqrt{2/f_1}W_1, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}). \tag{16}$$

Put

$$\Omega(w_1, w_4, z_1, z_2) = \left\{\alpha_5\left(c - \frac{\alpha_6}{\sqrt{\frac{n}{f_1}\left(1 + \frac{f_2}{f_3}\right)q_1}}\right)^2 + \alpha_7 + \frac{\alpha_8}{\frac{n}{f_1}\left(1 + \frac{f_2}{f_3}\right)q_1}\right\}^{-1/2},$$

where $q_1 = q_1(w_1, w_4, z_1, z_2)$,

$$\begin{aligned}
\alpha_5 &= \frac{1}{2}\left(\frac{n}{f_1} + \frac{f_2}{f_3}\frac{n}{f_2 + f_3}\right), \\
\alpha_6 &= \frac{1}{\alpha_5}\frac{n}{f_1}\frac{f_4}{f_1}\frac{\alpha_1}{\sqrt{\beta_2}}, \\
\alpha_7 &= \frac{\alpha_1^2 + \alpha_3^2}{\beta_2}, \\
\alpha_8 &= \left[2 + \frac{(f_3 - f_2)\{f_3(f_2 + f_3) + f_1f_2\}}{f_2f_3f_4}\right]\frac{n^2f_2f_4^2}{f_1^3f_3(f_2 + f_3)}\frac{\alpha_1^2}{2\alpha_5\beta_2}.
\end{aligned}$$

Without making confusion, we express

$$\Omega(\boldsymbol{v}_1) = \Omega(w_1, w_4, z_1, z_2).$$

for $\boldsymbol{v}_1 = \begin{pmatrix} w_1 & w_4 & z_1 & z_2 \end{pmatrix}$. From the expressions (11), (12) and (16), we have

$$\Omega(\sqrt{2/f_1}W_1, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}) \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{\rho(c, \widehat{\Delta_A^2})}}.$$

By taking into consideration that $q_{2,0} = (n/f_1)(1 + f_2/f_3)q_1(0, 0, 0, 0)$, it is easy to see that

$$\Omega(\mathbf{0}) = \frac{1}{\sqrt{\rho(c, \Delta^2)}}.$$

Since $\Omega(\boldsymbol{v}_1)$ is the smooth function on $(-1, \infty) \times (-1, \infty) \times \mathbb{R}^2$, Taylor series expansion at $(\sqrt{2/f_1} W_1, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}) = (0, 0, 0, 0)$ gives

$$\Omega(\sqrt{2/f_1}W_1, \sqrt{2/f_4}W_4, Z_1/\sqrt{n}, Z_2/\sqrt{n}) = \frac{1}{\sqrt{\rho(c, \Delta^2)}} + \frac{1}{\sqrt{n}}\tilde{\boldsymbol{\omega}}_1'\boldsymbol{y}_1 + \frac{1}{n}R_2,$$

where $\boldsymbol{y}_1 = (W_1, W_4, Z_1, Z_2)$, $\tilde{\boldsymbol{\omega}}_1 = \tilde{\boldsymbol{\omega}}_1(c, \Delta^2)$ is the vector valued function in $\mathbb{R}^4$ defined by

$$\tilde{\boldsymbol{\omega}}_1 = \boldsymbol{D}_1 \frac{\partial}{\partial \boldsymbol{v}_1}\Omega(\boldsymbol{v}_1)_{|0} \tag{17}$$

with being that $\boldsymbol{D}_1 = \mathrm{diag}(\sqrt{2n/f_1}, \sqrt{2n/f_4}, 1, 1)$, and R_2 is the residue term of which the property is similar to R_1. We can express that

$$\tilde{\boldsymbol{\omega}}_1 = \frac{1}{2}\frac{1}{\{\rho(c, \Delta^2)\}^{3/2}}\tilde{\boldsymbol{p}}_2,$$

where

$$\begin{aligned}\tilde{\boldsymbol{p}}_2 &= \tilde{\boldsymbol{p}}_2(c, \Delta^2) \\ &= \left(\frac{\alpha_5\alpha_6}{\sqrt{q_{1,0}}}c - \frac{\alpha_8 + \alpha_5\alpha_6^2}{q_{1,0}}\right) \cdot \begin{pmatrix} \sqrt{\frac{2n}{f_1}} & -\frac{f_4}{f_1}\sqrt{\frac{2n}{f_4}}\frac{1}{q_{1,0}} & -2\frac{n}{f_1}\sqrt{\frac{N_1N_2}{Nn}}\frac{\Delta}{q_{1,0}} & 0 \end{pmatrix}'.\end{aligned} \tag{18}$$

Let $\boldsymbol{\omega}_1$ be the extension for $\tilde{\boldsymbol{\omega}}_1$ defined as

$$\boldsymbol{\omega}_1 = \begin{pmatrix} \tilde{\omega}_{11} & 0 & 0 & \tilde{\omega}_{12} & \tilde{\omega}_{13} & \tilde{\omega}_{14} & 0 \end{pmatrix}'$$

for $\tilde{\boldsymbol{\omega}}_1 = \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} & \tilde{\omega}_{13} & \tilde{\omega}_{14} \end{pmatrix}'$. Then we have

$$\sqrt{\frac{\rho(c, \Delta^2)}{\rho(c, \widehat{\Delta_A^2})}} \stackrel{\mathcal{D}}{=} 1 + \frac{\sqrt{\rho(c, \Delta^2)}}{n}\boldsymbol{\omega}_1'\boldsymbol{y} + \frac{1}{n}R_3, \tag{19}$$

where R_3 is the residue term of which the property is similar to R_1. Combining (8) and (19), we have

$$\frac{C_{2|1} - \Phi(c)}{\phi(c)\sqrt{\rho(c, \widehat{\Delta_A^2})/n}} \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{\boldsymbol{p}_1'\boldsymbol{p}_1}}\boldsymbol{p}_1'\boldsymbol{y} + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{\boldsymbol{p}_1'\boldsymbol{p}_1}}\boldsymbol{y}'\left\{\boldsymbol{P} + \frac{1}{2\rho(c, \Delta^2)}(\boldsymbol{p}_1\boldsymbol{p}_2' + \boldsymbol{p}_2\boldsymbol{p}_1')\right\}\boldsymbol{y} + \frac{1}{n}R_4,$$

where $\boldsymbol{p}_2$ is the extension for $\tilde{\boldsymbol{p}}_2$ of which the definition is the same as $\boldsymbol{\omega}_1$, and R_4 is the residue term of which the property is similar to R_1.

Summarizing the above results, asymptotic expansion for the conditional probability of misclassification is obtained, which is given as the following theorem.

Theorem 3.

$$P\left(\frac{c_{2|1}\left(\sqrt{V_0}c - U_0\right) - \Phi(c)}{\phi(c)\sqrt{\rho(c,\widehat{\Delta^2_A})/n}} \le x\right) = \Phi(x) - \frac{1}{\sqrt{n}}\left(s_1 H_0(x) + s_2 H_2(x)\right)\phi(x) + O_{A,C}(n^{-1}),$$

where

$$s_1 = s_1(c,\Delta^2) = \frac{1}{\sqrt{\rho(c,\Delta^2)}}\,\mathrm{tr}\left\{\boldsymbol{P} + \frac{1}{2}\frac{1}{\rho(c,\Delta^2)}(\boldsymbol{p}_1\boldsymbol{p}_2' + \boldsymbol{p}_2\boldsymbol{p}_1')\right\},$$
$$s_2 = s_2(c,\Delta^2) = \frac{1}{\{\rho(c,\Delta^2)\}^{3/2}}\left\{\frac{\sqrt{2}}{3}\sum_{k=1}^{4}\sqrt{\frac{n}{f_k}}p_{1,k}^3 + \boldsymbol{p}_1'\boldsymbol{P}\boldsymbol{p}_1 + \boldsymbol{p}_1'\boldsymbol{p}_2\right\}.$$

Analytic forms for s_1 and s_2 are complicated, so we omit to describe them. We notice that s_1 and s_2 contain only the term in which the power of Δ is the even number.

4 Constrained linear discriminant rule for CPMC

In this section, we give a constrained linear discriminant rule for 2-groups of which one of the two conditional misclassification probabilities does not exceed the presetting value Ξ_{H} with the confidence level α.

Suppose that

$$c_{\mathrm{H}_1} = \xi_{\mathrm{H}_1} - \frac{1}{\sqrt{n}}\sqrt{\rho(\xi_{\mathrm{H}_1},\widehat{\Delta^2_A})}z_{1-\varepsilon}, \tag{20}$$

where $\xi_{\mathrm{H}_1} = \Phi^{-1}(\Xi_{\mathrm{H}_1})$ for $\Xi_{\mathrm{H}_1} \in (0,1)$. By virtue of (14) combined with Slutsky's theorem, we have

$$\lim_{\mathrm{A1}} P\left(c_{2|1}\left(\sqrt{V_0}c_{\mathrm{H}_1} - U_0\right) < \Xi_{\mathrm{H}_1}\right) = 1-\varepsilon$$

under the assumption C.

As an extension of (20), we obtain the following result.

Theorem 4. *Let*

$$c_{\mathrm{H}_2} = \xi_{\mathrm{H}_2} - \frac{1}{\sqrt{n}}h_1 - \frac{1}{n}h_2,$$

where

$$h_1 = \sqrt{\rho(\xi_{\mathrm{H}_2},\widehat{\Delta^2_A})}z_{1-\varepsilon},$$
$$h_2 = \left(\frac{\xi_{\mathrm{H}_2}}{2}\rho(\xi_{\mathrm{H}_2},\widehat{\Delta^2_A}) + \frac{nf_4}{f_1^2}\frac{\alpha_1}{\sqrt{\beta_2}}\sqrt{\frac{1}{q_{2,0}(\widehat{\Delta^2_A})}}\right)z_{1-\varepsilon}^2$$
$$+ \sqrt{\rho(\xi_{\mathrm{H}_2},\widehat{\Delta^2_A})}\left(b_1(\xi_{\mathrm{H}_2},\widehat{\Delta^2_A})H_0(z_{1-\varepsilon}) + b_2(\xi_{\mathrm{H}_2},\widehat{\Delta^2_A})H_2(z_{1-\varepsilon})\right)$$

with being that $\xi_{\mathrm{H}_2} = \Phi^{-1}(\Xi_{\mathrm{H}_2})$ *for* $\Xi_{\mathrm{H}_2} \in (0,1)$. *Then*

$$P\left(c_{2|1}\left(\sqrt{V_0}c_{\mathrm{H}_2} - U_0\right) < \Xi_{\mathrm{H}_2}\right) = 1-\varepsilon + O_{\mathrm{A1,C}}(n^{-1}).$$

The proof of Theorem 4 is similar to the one of Theorem 2 in McLachran [10], and so we omit to describe it.

5 Simulation result

Simulation experiments were performed to confirm the asymptotic result of Theorem 4. We also compared the accuracies with the asymptotic result of Theorem 2 in McLachlan [10] for the case in which $N_1 = N_2 = 250$, $p = 10, 30, 50, 70$, $\Delta = 1, 2, 3$, $\varepsilon = 0.05$, $\Xi = \Phi(-\Delta/2)$, where the settings of Δ and Ξ are followed to McLachlan [10]. When we treat the distributions of W-rule, without loss of generality from invariant property of the distribution for the orthogonal transformation of observation vector, we may assume that two given normal populations with the same covariance matrix are

$$\Pi_1 : N_p((\Delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p), \quad \Pi_2 : N_p(-(\Delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p),$$

where $\boldsymbol{e}_1 = (1, 0, \dots, 0)'$. To compute misclassification probability, generate 10^4 training samples. For each training samples, we generate 10^4 test samples in which observation vectors are i.i.d. as $N_p((\Delta/2)\boldsymbol{e}_1, \boldsymbol{I}_p)$. The value of the conditional misclassification probability was calculated by

$$\mathrm{sim}_k = \frac{\text{number of misclassification}}{10^4} \quad (k = 1, \dots, 10^4)$$

in each training samples. We took the average of $I(\mathrm{sim}_1 < \Xi), \dots, I(\mathrm{sim}_{10^4} < \Xi)$, where $I(.)$ denotes the indicator function, and wrote it as the value for the actual level in row "Y" in Tables 1. The same value for McLachlan [10]'s approximation was written in row "Mc".

Table 1: Actual levels of confidences that the conditional error probabilities are less than Ξ when the nominal level is $1 - \varepsilon = 0.95$.

		$p = 10$	$p = 30$	$p = 50$	$p = 70$
$\Delta = 1$	Y	0.95	0.95	0.96	0.95
	Mc	0.93	0.91	0.85	0.78
$\Delta = 2$	Y	0.95	0.95	0.95	0.95
	Mc	0.93	0.90	0.84	0.75
$\Delta = 3$	Y	0.94	0.94	0.93	0.93
	Mc	0.92	0.87	0.79	0.67

From Tables 1, we can see that our proposed asymptotic approximation has good accuracy when $\Delta = 1$. The actual level of confidence becomes small as the dimension gets large for the case in which $\Delta = 3$. We can check that McLachlan [10]'s result does not work well for our settings. Extra simulation results which does not written in this paper reveals that the actual confidence level gets small from the nominal level as the dimension becomes close to sample size for the case in which Ξ is small.

6 Concluding remarks

In this paper, we derived Studentized statistic for the conditional probability of misclassification for the Studentized W, and derived its asymptotic expansion for distribution up to the term of $O(n^{-1/2})$ under the high-dimensional asymptotic framework A1. It may be noted that the order of its error is $O(n^{-1})$. Based on the derived asymptotic expansion, we gave the cut-off point for the linear discriminant rule that the one of two conditional error probabilities is less than the presetting value. Simulation results revealed that our proposed rule is superior than McLachlan [10]'s result.

Unfortunately, our proposed rule did not work well for the case in which Ξ is small. The modification should be considered and is being a future problem.

A Equality in distributions for proposed statistics

In this section, firstly, we mention the equality in distributions for $U_{01} - U_1$ and $-(U_{02} - U_2)$, which is given as the following lemma,

Lemma 2. *The distribution for $U_{01} - U_1$ is the same as the one for $-(U_{02} - U_2)$ with exchanging N_1 for N_2.*

Proof. Set $S_1(N_1, N_2) = U_{01} - U_1$, and set $S_2(N_1, N_2) = -(U_{02} - U_2)$. In addition, put

$$\bar{\boldsymbol{x}}_i \stackrel{\mathcal{D}}{=} \boldsymbol{\mu}_i + \frac{1}{\sqrt{N_i}}\boldsymbol{z}_i \quad (i = 1, 2), \quad \boldsymbol{S} \stackrel{\mathcal{D}}{=} \frac{1}{n}W, \quad \boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

where $\boldsymbol{z}_1$,$\boldsymbol{z}_2$, $\boldsymbol{W}$ are independent; $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are distributed as $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$; $\boldsymbol{W}$ is distributed as $W_p(n, \boldsymbol{I}_p)$. Then, we have

$$\begin{aligned}
S_1(N_1, N_2) \stackrel{\mathcal{D}}{=} & -\frac{1}{2}\frac{n}{m-1}\left\{n\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right) - \frac{Np}{N_1N_2}\right\} \\
& - \frac{1}{2}\frac{p}{m-1}\left(\frac{n}{N_2} - \frac{n}{N_1}\right) - \frac{2(n-1)}{(m-1)(m+1)}\frac{n}{N_1} \\
& - \frac{n}{2}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 + \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right), \\
S_2(N_1, N_2) \stackrel{\mathcal{D}}{=} & -\frac{1}{2}\frac{n}{m-1}\left\{n\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right) - \frac{Np}{N_1N_2}\right\} \\
& + \frac{1}{2}\frac{p}{m-1}\left(\frac{n}{N_2} - \frac{n}{N_1}\right) - \frac{2(n-1)}{(m-1)(m+1)}\frac{n}{N_2} \\
& + \frac{n}{2}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\boldsymbol{z}_1 + \frac{1}{\sqrt{N_2}}\boldsymbol{z}_2 + \boldsymbol{\delta}\right).
\end{aligned}$$

By interchanging N_1 and N_2,

$$\begin{aligned}
S_2(N_2, N_1) \stackrel{\mathcal{D}}{=} & -\frac{1}{2}\frac{n}{m-1}\left\{n\left(\frac{1}{\sqrt{N_2}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_1}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_2}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_1}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right) - \frac{Np}{N_2N_1}\right\} \\
& + \frac{1}{2}\frac{p}{m-1}\left(\frac{n}{N_1} - \frac{n}{N_2}\right) - \frac{2(n-1)}{(m-1)(m+1)}\frac{n}{N_1} \\
& + \frac{n}{2}\left(\frac{1}{\sqrt{N_2}}\boldsymbol{z}_1 - \frac{1}{\sqrt{N_1}}\boldsymbol{z}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_2}}\boldsymbol{z}_1 + \frac{1}{\sqrt{N_1}}\boldsymbol{z}_2 + \boldsymbol{\delta}\right) \\
= & -\frac{1}{2}\frac{n}{m-1}\left\{n\left(\frac{1}{\sqrt{N_1}}\tilde{\boldsymbol{z}}_1 - \frac{1}{\sqrt{N_2}}\tilde{\boldsymbol{z}}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\tilde{\boldsymbol{z}}_1 - \frac{1}{\sqrt{N_2}}\tilde{\boldsymbol{z}}_2 - \boldsymbol{\delta}\right) - \frac{Np}{N_1N_2}\right\} \\
& - \frac{1}{2}\frac{p}{m-1}\left(\frac{n}{N_2} - \frac{n}{N_1}\right) - \frac{2(n-1)}{(m-1)(m+1)}\frac{n}{N_1} \\
& - \frac{n}{2}\left(\frac{1}{\sqrt{N_1}}\tilde{\boldsymbol{z}}_1 - \frac{1}{\sqrt{N_2}}\tilde{\boldsymbol{z}}_2 - \boldsymbol{\delta}\right)'\boldsymbol{W}^{-1}\left(\frac{1}{\sqrt{N_1}}\tilde{\boldsymbol{z}}_1 + \frac{1}{\sqrt{N_2}}\tilde{\boldsymbol{z}}_2 - \boldsymbol{\delta}\right) \\
\stackrel{\mathcal{D}}{=} & \; S_1(N_1, N_2),
\end{aligned}$$

where $\tilde{\boldsymbol{z}}_1 = -\boldsymbol{z}_2$ and $\tilde{\boldsymbol{z}}_2 = -\boldsymbol{z}_1$. □

Next, we show the equality in distribution for V_0 and for V, which is given as the following lemma.

Lemma 3. *Each of the distributions for V_0 and for V is the same as the one with exchanging N_1 for N_2.*

We omit to write the proof of Lemma 3 since it is similar to Lemma 2.

B Proof of Theorem 2

In this section, we gave the proof of Theorem 2. Firstly, we gave the following two lemmas which is to be used.

Lemma 4. *Suppose that $a \in \mathbb{R}$ and $g(.)$ is a polynomial function. Let Z and Y are random variables; Z is distributed as the standard normal distribution; Y is distributed as the chi-square distribution with f degrees of freedom. Then,*

$$E\left[g(Z)e^{itaZ}\right] = \exp\left(\frac{a^2}{2}(it)^2\right) E[g(Z+ita)],$$
$$E\left[g(W)e^{itaW}\right] = \left(1 - ita\sqrt{\frac{2}{f}}\right)^{-f/2} \exp\left(-ita\sqrt{\frac{f}{2}}\right) E\left[g\left(\frac{W+ita}{1-ita\sqrt{2/a}}\right)\right],$$

where $i = \sqrt{-1}$, and $W = \sqrt{f/2}(Y/f - 1)$.

It is easy to prove Lemma 4, so we omit to write the proof.

Proof of Theorem 2. From the assumption for T given in (15), the characteristic function can be expanded as

$$E[\exp(itT)] = E\left[T_0 + \frac{it}{\sqrt{n}}\frac{1}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}T_1\right] + O(n^{-1}),$$

where

$$T_0 = \exp\left(it\frac{1}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\boldsymbol{h}'\boldsymbol{y}\right), \quad T_1 = \boldsymbol{y}'\boldsymbol{H}\boldsymbol{y}\exp\left(it\frac{1}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\boldsymbol{h}'\boldsymbol{y}\right).$$

From Lemma 4, we have

$$\begin{aligned} E[T_0] &= \prod_{k=1}^{4}\left(1 - it\frac{h_k}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\sqrt{\frac{2}{f_k}}\right)^{-f_k/2} \exp\left(-it\frac{h_k}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}\sqrt{\frac{f_k}{2}}\right) \cdot \prod_{k=5}^{7}\exp\left(\frac{1}{2}(it)^2\frac{h_k^2}{\boldsymbol{h}'\boldsymbol{h}}\right) \\ &= \left[1 + \frac{1}{\sqrt{n}}\left\{\frac{\sqrt{2}}{3}(it)^3\sum_{k=1}^{4}\sqrt{\frac{n}{f_k}}\frac{h_k^3}{(\boldsymbol{h}'\boldsymbol{h})^{3/2}}\right\}\right] e^{(it)^2/2} + O(n^{-1}) \end{aligned}$$

under A1. It can be expressed that

$$\begin{aligned} E[T_1] = &\sum_{k=1}^{7} h_{kk}E\left[Y_k^2\exp\left(it\frac{h_k}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}Y_k\right)\right] E\left[\exp\left(it\sum_{\substack{\ell=1\\ \ell\neq k}}^{7}\frac{h_k}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}Y_k\right)\right] \\ &+ \sum_{k=1}^{7}\sum_{\substack{\ell=1\\ \ell\neq k}}^{7} h_{k\ell}E\left[Y_kY_\ell\exp\left(it\frac{h_k}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}Y_k\right)\exp\left(it\frac{h_\ell}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}Y_\ell\right)\right] E\left[\exp\left(it\sum_{\substack{\alpha=1\\ \alpha\neq k,\alpha\neq\ell}}^{7}\frac{h_\alpha}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}}Y_\alpha\right)\right] \end{aligned}$$

for $(h_{k\ell}) = \boldsymbol{H}$. From Lemma 2 again, we have

$$\begin{aligned}
E[T_1] &= E[T_0]\left[\sum_{k=1}^{4} h_{kk}\frac{1+(it)^2 h_k^2/\boldsymbol{h}'\boldsymbol{h}}{\{1-it(h_k/\sqrt{\boldsymbol{h}'\boldsymbol{h}})\sqrt{2/f_k}\}^2} + \sum_{k=5}^{7} h_{kk}\{1+(it)^2 h_k^2/\boldsymbol{h}'\boldsymbol{h}\}\right] \\
&\quad + E[T_0]\left\{\sum_{k=1}^{4}\sum_{\substack{\ell=1\\ \ell\neq k}}^{4} h_{k\ell}\frac{ith_k/\sqrt{\boldsymbol{h}'\boldsymbol{h}}}{1-it(h_k/\sqrt{\boldsymbol{h}'\boldsymbol{h}})\sqrt{2/f_k}}\frac{ith_\ell/\sqrt{\boldsymbol{h}'\boldsymbol{h}}}{1-it(h_\ell/\sqrt{\boldsymbol{h}'\boldsymbol{h}})\sqrt{2/f_k}}\right. \\
&\quad \left. +2\sum_{k=1}^{4}\sum_{\ell=5}^{7} h_{k\ell}\frac{ith_k/\sqrt{\boldsymbol{h}'\boldsymbol{h}}}{1-it(h_k/\sqrt{\boldsymbol{h}'\boldsymbol{h}})\sqrt{2/f_k}} it\frac{h_\ell}{\sqrt{\boldsymbol{h}'\boldsymbol{h}}} + \sum_{k=5}^{7}\sum_{\substack{\ell=5\\ \ell\neq k}}^{7} h_{k\ell}(it)^2\frac{h_k h_\ell}{\boldsymbol{h}'\boldsymbol{h}}\right\} \\
&= E[T_0]\left\{\operatorname{tr}\boldsymbol{H} + (it)^2\frac{\boldsymbol{h}'\boldsymbol{H}\boldsymbol{h}}{\boldsymbol{h}'\boldsymbol{h}} + O(n^{-1/2})\right\}
\end{aligned}$$

under A1. The desired result now follows by formally inverting the expansion for the characteristic function. □

C Analytic forms for $\boldsymbol{P}_0$, $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$

In this section, we give the analytic form for $\boldsymbol{P}_i$ $(i = 0, 1, 2, 3)$. The derivation is straightforward, and so is omitted.

$$
\boldsymbol{P}_0 = \frac{1}{\sqrt{q_{2,0}\beta_2}}
\begin{pmatrix}
0 & \frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_2}}\alpha_1 & -\frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_3}}\alpha_1 & \left(\frac{f_4}{f_1}\right)^2\frac{n}{\sqrt{f_1 f_4}}\frac{\alpha_1}{q_{1,0}} & \sqrt{2}\frac{f_4}{f_1}\left(\frac{n}{f_1}\right)^{3/2}\frac{\alpha_1}{\sqrt{\beta_2}}\frac{\Delta}{q_{1,0}} & 0 & 0 \\
* & 0 & 0 & -\frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_2 f_4}}\alpha_1 & -\frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_2}}\frac{\alpha_1}{\sqrt{\beta_2}}\Delta & \frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_3}}\alpha_4\Delta & 0 \\
* & * & 0 & \frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_3 f_4}}\alpha_1 & \frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_3}}\frac{\alpha_1}{\sqrt{\beta_2}}\Delta & -\frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\frac{\sqrt{2nf_2}}{f_3}\alpha_4\Delta & 0 \\
* & * & * & -\frac{2nf_4}{f_1^2}\frac{\alpha_1}{q_{1,0}} & -\frac{3}{2}\frac{n}{f_1}\frac{f_4}{f_1}\sqrt{\frac{2n}{f_4}}\frac{\alpha_1}{\sqrt{\beta_2}}\frac{\Delta}{q_{1,0}} & -\frac{1}{2}\frac{n}{f_1}\frac{f_4}{f_1}\sqrt{\frac{2nf_2}{f_3 f_4}}\alpha_4\frac{\Delta}{q_{1,0}} & 0 \\
* & * & * & * & 2\frac{n}{f_1}\frac{f_4}{f_1}\frac{\alpha_1}{q_{1,0}} & -\left(\frac{n}{f_1}\right)^2\sqrt{\frac{f_2}{f_3}}\frac{\alpha_4}{\sqrt{\beta_2}}\frac{\Delta^2}{q_{1,0}} & 0 \\
* & * & * & * & * & 2\frac{n}{f_1}\alpha_1 & 0 \\
* & * & * & * & * & * & 0
\end{pmatrix},
$$

$$
\boldsymbol{P}_1 =
\begin{pmatrix}
-\frac{2n}{f_1}\left\{\left(\frac{f_4}{f_1}\right)^2\frac{\alpha_1^2}{\beta_2}\frac{1}{q_{2,0}}+\frac{1}{4}\right\} & -\frac{1}{2}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_2}} & \frac{1}{2}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_3}} & \left(\frac{f_4}{f_1}\right)^2\frac{2n}{\sqrt{f_1 f_4}}\frac{\alpha_1^2}{\beta_2}\frac{1}{q_{2,0}} & \sqrt{2}\frac{f_4}{f_1}\left(\frac{n}{f_1}\right)^{3/2}\frac{\alpha_1^2}{\beta_2^{3/2}}\frac{\Delta}{q_{2,0}} & \frac{n}{f_1}\frac{f_4}{f_1}\sqrt{\frac{2nf_2}{f_1 f_3}}\frac{\alpha_1\alpha_4}{\beta_2}\frac{\Delta}{q_{2,0}} & \frac{f_4}{f_1}\sqrt{\frac{2n}{f_1}}\frac{\alpha_1\alpha_3}{\beta_2}\frac{1}{\sqrt{q_{2,0}}} \\
* & \frac{3}{2}\frac{n}{f_2}\left(\frac{f_2}{f_2+f_3}\right)^2 & \left\{-\frac{3}{2}\left(\frac{f_2}{f_2+f_3}\right)^2+\frac{f_2}{f_2+f_3}\right\}\frac{n}{\sqrt{f_2 f_3}} & 0 & 0 & 0 & 0 \\
* & * & \left\{\frac{3}{4}\left(\frac{f_2}{f_2+f_3}\right)^2-\frac{f_2}{f_2+f_3}\right\}\frac{2n}{f_3} & 0 & 0 & 0 & 0 \\
* & * & * & -\frac{2nf_4}{f_1^2}\frac{\alpha_1^2}{\beta_2}\frac{1}{q_{2,0}} & -\frac{n}{f_1}\frac{\sqrt{2nf_4}}{f_1}\frac{\alpha_1^2}{\beta_2^{3/2}}\frac{\Delta}{q_{2,0}} & -\frac{n}{f_1}\frac{f_4}{f_1}\sqrt{\frac{2nf_2}{f_3 f_4}}\frac{\alpha_1\alpha_4}{\beta_2}\frac{\Delta}{q_{2,0}} & -\frac{f_4}{f_1}\sqrt{\frac{2n}{f_4}}\frac{\alpha_1\alpha_3}{\beta_2}\frac{1}{\sqrt{q_{2,0}}} \\
* & * & * & * & -\left(\frac{n}{f_1}\right)^2\frac{\alpha_1^2}{\beta_2^2}\frac{\Delta^2}{q_{2,0}} & -\left(\frac{n}{f_1}\right)^2\sqrt{\frac{f_2}{f_3}}\frac{\alpha_1\alpha_4}{\beta_2^{3/2}}\frac{\Delta^2}{q_{2,0}} & -\frac{n}{f_1}\frac{\alpha_1\alpha_3}{\beta_2^{3/2}}\frac{\Delta}{\sqrt{q_{2,0}}} \\
* & * & * & * & * & -\left(\frac{n}{f_1}\right)^2\frac{f_2}{f_3}\frac{\alpha_4^2}{\beta_2}\frac{\Delta^2}{q_{2,0}} & -\frac{n}{f_1}\sqrt{\frac{f_2}{f_3}}\frac{\alpha_3\alpha_4}{\beta_2}\frac{\Delta}{\sqrt{q_{2,0}}} \\
* & * & * & * & * & * & -\frac{\alpha_3^2}{\beta_2}
\end{pmatrix},
$$

$$
\boldsymbol{P}_2 = \frac{1}{\sqrt{\beta_2 q_{2,0}}} \begin{pmatrix}
\frac{2nf_4}{f_1^2}\alpha_1 & -\frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_2}}\alpha_1 & \frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_3}}\alpha_1 & -\frac{f_4}{f_1}\frac{n}{\sqrt{f_1 f_4}}\alpha_1 & -\frac{1}{\sqrt{2}}\left(\frac{n}{f_1}\right)^{3/2}\frac{\alpha_1}{\sqrt{\beta_2}}\Delta & -\frac{1}{\sqrt{2}}\left(\frac{n}{f_1}\right)^{3/2}\sqrt{\frac{f_2}{f_3}}\alpha_4\Delta & -\frac{1}{2}\sqrt{\frac{2n}{f_1}}\alpha_3\sqrt{q_{2,0}} \\
* & 0 & 0 & \frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_2 f_4}}\alpha_1 & \frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_2}}\frac{\alpha_1}{\sqrt{\beta_2}}\Delta & \frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_3}}\alpha_4\Delta & \frac{1}{2}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_2}}\alpha_3\sqrt{q_{2,0}} \\
* & * & 0 & -\frac{f_4}{f_1}\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_3 f_4}}\alpha_1 & -\frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_3}}\frac{\alpha_1}{\sqrt{\beta_2}}\Delta & -\frac{1}{2}\frac{n}{f_1}\frac{f_2}{f_2+f_3}\frac{\sqrt{2nf_2}}{f_3}\alpha_4\Delta & -\frac{1}{2}\frac{f_2}{f_2+f_3}\sqrt{\frac{2n}{f_3}}\alpha_3\sqrt{q_{2,0}} \\
* & * & * & 0 & 0 & 0 & 0 \\
* & * & * & * & 0 & 0 & 0 \\
* & * & * & * & * & 0 & 0 \\
* & * & * & * & * & * & 0
\end{pmatrix},
$$

$$
\boldsymbol{P}_3 = -\frac{1}{2} \begin{pmatrix}
\frac{n}{f_1} & -\frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_2}} & \frac{f_2}{f_2+f_3}\frac{n}{\sqrt{f_1 f_3}} & 0 & 0 & 0 & 0 \\
* & \left(\frac{f_2}{f_2+f_3}\right)^2\frac{n}{f_2} & -\left(\frac{f_2}{f_2+f_3}\right)^2\frac{n}{\sqrt{f_2 f_3}} & 0 & 0 & 0 & 0 \\
* & * & \left(\frac{f_2}{f_2+f_3}\right)^2\frac{n}{f_3} & 0 & 0 & 0 & 0 \\
* & * & * & 0 & 0 & 0 & 0 \\
* & * & * & * & 0 & 0 & 0 \\
* & * & * & * & * & 0 & 0 \\
* & * & * & * & * & * & 0
\end{pmatrix}.
$$

References

[1] Anderson, T.W. (1973). An asymptotic expansion of the distribution of the Studentized classification statistic W. *Ann. Statist.*, **1**, 964–972.

[2] Anderson, T.W. (2003). *An Introduction of Multivariate Statistics 3rd ed.* Wiley, Hoboken.

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

[4] Fujikoshi, Y. (2002). Selection of variables for discriminant analysis in a high-dimensional case. *Sankhyā*, **64**, 256–267.

[5] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for 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.

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

[7] Hyodo, M. and Kubokawa, T. (2014). A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data. *J. Multivariate Anal.*, **123**, 364–379.

[8] James, G.S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, **41**, 19-43.

[9] Lachenbruch, P.A. (1968). On expected probability of misclassification in discriminant analysis, secessary sample size, and a relation with the multiple correlation coefficients. *Biometrics*, **24**, 823–834. Wiley, Hoboken, NJ.

[10] McLachlan, G. J. (1977). Constrained sample discrimination with the Studentized classification statistic W, *Comm. Statist. -Theory and Methods*, **6**, 575–583.

[11] McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*, Wiley, NY.

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

[13] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistic. *Handbook of Statistics* (P. R. Krishnaiah and L. N. Kanal ed.), North-Holland, 47-60.

[14] Yamada, T., Himeno, T. and Sakurai, T. (2017). Asymptotic cut-off point in linear discriminant rule to adjust the misclassification probability for large dimensions. *Hiroshima Math. J.*, **37**, 319–348.