Asymptotic expansions relating to discrimination based on 2-step monotone missing samples

Nobumichi Shutoh[a,*]

[a] *Department of Mathematical Information Science, Graduate School of Science, Tokyo University of Science*
1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, JAPAN

Abstract

This paper provides two asymptotic expansions derived in Okamoto (1963) and McLachlan (1973) based on 2-step monotone missing samples, i.e., certain extensions of Okamoto (1963) and McLachlan (1973) up to the first order. These asymptotic expansions are useful for approximating the probabilities of misclassification and considering of their interval estimation. For investigating the performances in approximating the probabilities, simulation studies for asymptotic expansion of the type of Okamoto (1963) are also given under the selected parameters.

Keywords: discriminant analysis, asymptotic expansion, probabilities of misclassification, asymptotic approximation, monotone missing samples.
2000 MSC: 62H30, 62H10.

---

*Corresponding author, *Telephone Number:*+81-3-3260-4271,
Fax Number:+81-3-3260-4293.
Email address: j1409702@ed.kagu.tus.ac.jp (Nobumichi Shutoh)

1. Introduction

In discriminant analysis, it is important to discuss probabilities of misclassification in discrimination. Although the probabilities are unknown in both of linear discriminant function (LDF) and maximum likelihood discriminant function based on sample vectors, that is, including estimators, the asymptotic distributions have been found for two groups $\Pi^{(g)}$: $N_p(\boldsymbol{\mu}^{(g)}, \Sigma)$ for $g = 1, 2$. First, as introduction in this paper, we present the settings, LDF and its probabilities of misclassification.

Suppose that we can obtain p dimensional sample vectors:

$$\boldsymbol{x}_j^{(g)} = \begin{pmatrix} \boldsymbol{x}_{1j}^{(g)} \\ \boldsymbol{x}_{2j}^{(g)} \end{pmatrix} \sim N_p(\boldsymbol{\mu}^{(g)}, \Sigma) \tag{1}$$

for $g = 1, 2,\ j = 1, \ldots, N_1^{(g)}$ from $\Pi^{(g)}$ respectively, where $\boldsymbol{x}_{\ell j}^{(g)}$ is p_ℓ dimensional partitioned vector of $\boldsymbol{x}_j^{(g)}$ and $p = p_1 + p_2$. Now we prepare similar partitions of $\boldsymbol{\mu}^{(g)}$ and Σ:

$$\boldsymbol{\mu}^{(g)} = \begin{pmatrix} \boldsymbol{\mu}_1^{(g)} \\ \boldsymbol{\mu}_2^{(g)} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\boldsymbol{\mu}_\ell^{(g)}$ denotes p_ℓ dimensional partitioned vector of $\boldsymbol{\mu}^{(g)}$ and $\Sigma_{\ell m}$ denotes $p_\ell \times p_m$ partitioned matrix of Σ for $\ell = 1, 2,\ m = 1, 2$. Then LDF can be constructed as

$$W = (\overline{\boldsymbol{x}}_F^{(1)} - \overline{\boldsymbol{x}}_F^{(2)})' S^{-1} \left[\boldsymbol{x} - \frac{1}{2}(\overline{\boldsymbol{x}}_F^{(1)} + \overline{\boldsymbol{x}}_F^{(2)}) \right], \tag{2}$$

where $\boldsymbol{x}$ is p dimensional sample vector arising from $\Pi^{(1)}$ or $\Pi^{(2)}$,

$$\overline{\boldsymbol{x}}_F^{(g)} = \begin{pmatrix} \overline{\boldsymbol{x}}_{1F}^{(g)} \\ \overline{\boldsymbol{x}}_{2F}^{(g)} \end{pmatrix} = \frac{1}{N_1^{(g)}} \sum_{j=1}^{N_1^{(g)}} \begin{pmatrix} \boldsymbol{x}_{1j}^{(g)} \\ \boldsymbol{x}_{2j}^{(g)} \end{pmatrix},$$

S denotes pooled sample covariance matrix with well known correction of coefficient:

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \frac{1}{n_1} \sum_{g=1}^{2} \sum_{j=1}^{N_1^{(g)}} (\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}_F^{(g)})(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}_F^{(g)})',$$

$S_{\ell m}$ denotes $p_\ell \times p_m$ partitioned matrix of S and $n_1 = N_1^{(1)} + N_1^{(2)} - 2$. They are the maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}^{(g)}$ and Σ respectively. Then the discrimination rule for (2) is as follows: $\boldsymbol{x}$ is assigned to $\Pi^{(1)}$ if $W > c$, $\boldsymbol{x}$ is assigned to $\Pi^{(2)}$ otherwise, where c is cut-off point depends on a priori probabilities $\boldsymbol{x}$ arises from $\Pi^{(1)}$ or $\Pi^{(2)}$ and the risk of discrimination for (2). The probabilities of misclassification can be described:

$$\begin{aligned} e(2|1) &\equiv \Pr\left[W \le c | \boldsymbol{x} \in \Pi^{(1)}\right], \\ e(1|2) &\equiv \Pr\left[W > c | \boldsymbol{x} \in \Pi^{(2)}\right] \\ &= 1 - \Pr\left[W \le c | \boldsymbol{x} \in \Pi^{(2)}\right]. \end{aligned}$$

It is hard to evaluate the above probabilities exactly since LDF includes estimates. However asymptotic distribution of W has been well known: W is asymptotically distributed as $N((-1)^{g-1}(1/2)\Delta^2, \Delta^2)$ under $\boldsymbol{x} \in \Pi^{(g)}$ as $N_1^{(1)}$ and $N_1^{(2)}$ tend to infinity, where Δ^2 denotes Mahalanobis squared distance $(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})'\Sigma^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$. Therefore asymptotic expansions have been provided.

Okamoto (1963) (with correction, Okamoto (1968)) provided the results up to the terms of the second order with respect to $(N_1^{(1)^{-1}}, N_1^{(2)^{-1}}, n_1^{-1})$. The results up to the terms of the first order with respect to $(N_1^{(1)^{-1}}, N_1^{(2)^{-1}}, n_1^{-1})$ have following forms:

$$P_{(2|1)}(v, N_1^{(1)}, N_1^{(2)}, \Delta) \equiv \Pr\left[\frac{W - \frac{1}{2}\Delta^2}{\Delta} \le v | \boldsymbol{x} \in \Pi^{(1)}\right] \tag{3}$$

$$= \Phi(v) - \frac{1}{2}\left[\frac{a_1}{N_1^{(1)}\Delta^2} + \frac{a_2}{N_1^{(2)}\Delta^2} + \frac{a_3}{2n_1}\right]\phi(v)$$
$$+ \mathrm{O}(N_1^{(1)^{-2}}, N_1^{(2)^{-2}}, n_1^{-2})$$

and $\Pr\left[(W + (1/2)\Delta^2)\Delta^{-1} \le v | \boldsymbol{x} \in \Pi^{(2)}\right] = 1 - P_{(2|1)}(-v, N_1^{(2)}, N_1^{(1)}, \Delta)$ respectively, where $\Phi(\cdot)$ denotes cumulative distribution function, $\phi(\cdot)$ denotes probability density function of standard normal distribution,

$$\begin{aligned} a_1 &= v^3 + (p-3)v - p\Delta, \\ a_2 &= v^3 + 2\Delta v^2 + (p - 3 + \Delta^2)v + (p-2)\Delta, \\ a_3 &= 4v^3 + 4\Delta v^2 + (6p - 6 + \Delta^2)v + 2(p-1)\Delta. \end{aligned}$$

Besides Anderson (1973) derived an asymptotic expansion of the case that Δ^2 is replaced by D^2, where D^2 denotes sample Mahalanobis squared distance based on $N_1^{(1)}+N_1^{(2)}$ sample vectors. Also in the discrimination based on maximum likelihood, the result similar to Okamoto (1963) was given by Memon and Okamoto (1971) by using asymptotic normality of discriminant function, which is asymptotically distributed as $N((-1)^g\Delta^2, 4\Delta^2)$ under $\boldsymbol{x} \in \Pi^{(g)}$ for $g = 1, 2$.

Especially, in both of discrimination rules with cut-off point $c = 0$, $\Phi(-(1/2)D)$ is considerable as an estimator for the probabilities of misclassification. Therefore McLachlan (1973) expanded the expectation of that up to the terms of the second order:

$$\mathrm{E}\left[\Phi\left(-\frac{1}{2}D\right)\right] = \Phi\left(-\frac{1}{2}\Delta\right) + b_1 + \mathrm{O}(N_1^{(1)^{-2}}, N_1^{(2)^{-2}}, n_1^{-2}), \qquad (4)$$

where

$$b_1 = \frac{1}{16}\phi\left(-\frac{1}{2}\Delta\right)\left[\left(\frac{1}{N_1^{(1)}} + \frac{1}{N_1^{(2)}}\right)\{\Delta - \frac{4}{\Delta}(p-1)\}\right.$$

$$+ \frac{\Delta}{2n_1}\{\Delta^2 - 4(2p+1)\}\Bigg].$$

On the other asymptotic approximations, Lachenbruch (1968) derived an approximation by asymptotic normality of LDF. Fujikoshi and Seo (1998) proposed that in both of LDF and discriminant function based on maximum likelihood by expressing discriminant functions with chi-square and noncentral chi-square random variables and making use of their asymptotic properties.

There are two types of asymptotic approximations. Type I approximations are ones under the framework such that $N_1^{(1)} \to \infty$, $N_1^{(2)} \to \infty$, $N_1^{(2)}/N_1^{(1)} \to$ positive const. and type II approximations are ones under the framework such that $N_1^{(1)} \to \infty$, $N_1^{(2)} \to \infty$, $p \to \infty$, $n_1 - p \to \infty$, $N_1^{(2)}/N_1^{(1)} \to$ positive const. In general, asymptotic expansions are regarded as type I approximations. Fujikoshi and Seo (1998) considered one of type II approximations. Besides Lachenbruch (1968) can be regarded as both of type I and type II approximation. Error bounds for them have been reviewed in Fujikoshi, Ulyanov and Shimizu (2010).

Recently extensions of Lachenbruch (1968) and Anderson (1973) in the case of 2-step monotone missing samples have been given. In addition to (1), the authors assumed that $N_2^{(g)} = N^{(g)} - N_1^{(g)}$ sample vectors from $\Pi^{(g)}$ are also obtained:

$$\boldsymbol{x}_{1j}^{(g)} \sim N_{p_1}(\boldsymbol{\mu}_1^{(g)}, \Sigma_{11})$$

for $g = 1, 2,\ j = N_1^{(g)} + 1, \ldots, N^{(g)}$. Shutoh, Hyodo and Seo (2009) gave that for Lachenbruch (1968) and unbiased estimates of Δ^2 and δ_{11}^2, where $\delta_{11}^2 = (\boldsymbol{\mu}_1^{(1)} - \boldsymbol{\mu}_1^{(2)})'\Sigma_{11}^{-1}(\boldsymbol{\mu}_1^{(1)} - \boldsymbol{\mu}_1^{(2)})$. Shutoh and Seo (2010) gave that for

Anderson (1973) by perturbation method.

In this paper, by making use of Shutoh and Seo (2010), we extend the result of Okamoto (1963) and McLachlan (1973) up to the terms of the first order with respect to $(N_1^{(1)^{-1}}, N_1^{(2)^{-1}}, n_1^{-1})$ to the case of 2-step monotone missing samples. Also we prepare relevant conditions in deriving asymptotic expansion and nonsingularity of estimators such that

$$\rho_i \equiv \frac{n_i}{n} \to \text{positive const.},\ n_1 - p_1 > 0,\ n_1 - p_2 > 0,$$

where $n = N^{(1)} + N^{(2)} - 2$ and $n_i = N_i^{(1)} + N_i^{(2)} - 2$ for large $N_1^{(g)}$, $N_2^{(g)}$ $(g = 1, 2,\ i = 1, 2)$. Thus we find $\rho \equiv \rho_1 + \rho_2 = 1$ easily in asymptotic sense. Besides asymptotic expansions provided in this paper are regarded as type I approximations for the probabilities of misclassification for 2-step monotone missing samples under the framework such that

$$N_1^{(g)} \to \infty,\ N_2^{(g)} \to \infty,\ \frac{N_1^{(2)}}{N_1^{(1)}} \to \text{positive const.},\ \frac{N^{(2)}}{N^{(1)}} \to \text{positive const.}$$

The organization of this paper is as follows. In Section 2, we introduce the MLEs of 2-step monotone missing samples with their properties and construct LDF. In Section 3, we show the derivations of asymptotic expansions. We give some useful Lemmas for our purpose in Section 4. Also simulation studies are presented in Section 5. Finally we present conclusion in this paper.

2. MLEs and LDF based on 2-step monotone missing samples

In this paper, we need MLEs based on 2-step monotone missing samples for two groups. These can be derived similar to Anderson and Olkin (1985) which gave the results for one group. As concerns details of derivation of the MLEs, see Shutoh, Hyodo and Seo (2009). The MLEs are

$$\widehat{\boldsymbol{\mu}}^{(g)} = \begin{pmatrix} \widehat{\boldsymbol{\mu}}_1^{(g)} \\ \widehat{\boldsymbol{\mu}}_2^{(g)} \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{x}}_{1T}^{(g)} \\ \overline{\boldsymbol{x}}_{2F}^{(g)} - \widehat{\Psi}_{21}(\overline{\boldsymbol{x}}_{1F}^{(g)} - \overline{\boldsymbol{x}}_{1T}^{(g)}) \end{pmatrix},$$

$$\widehat{\Sigma} = \begin{pmatrix} \widehat{\Sigma}_{11} & \widehat{\Sigma}_{12} \\ \widehat{\Sigma}_{21} & \widehat{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \widehat{\Psi}_{11} & \widehat{\Psi}_{11}\widehat{\Psi}_{12} \\ \widehat{\Psi}_{21}\widehat{\Psi}_{11} & \widehat{\Psi}_{22} + \widehat{\Psi}_{21}\widehat{\Psi}_{11}\widehat{\Psi}_{12} \end{pmatrix},$$

where

$$\begin{aligned}
\widehat{\Psi}_{11} &= \frac{1}{n}\left[n_1 S_{11} + n_2 S^{(2)} + \sum_{g=1}^{2}\left\{\frac{N_1^{(g)}N_2^{(g)}}{N^{(g)}}(\overline{\boldsymbol{x}}_{1F}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})(\overline{\boldsymbol{x}}_{1F}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})'\right\}\right], \\
\widehat{\Psi}_{12} &= S_{11}^{-1}S_{12}, \\
\widehat{\Psi}_{22} &= S_{22\cdot 1}, \\
S^{(2)} &= \frac{1}{n_2}\sum_{g=1}^{2}\sum_{j=N_1^{(g)}+1}^{N^{(g)}}(\boldsymbol{x}_{1j}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})(\boldsymbol{x}_{1j}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})',
\end{aligned}$$

and $S_{\ell m}$ denotes $p_\ell \times p_m$ partitioned matrix of S. Also random matrices which construct $\widehat{\Sigma}$ have the following distributions.

$$\begin{aligned}
n_1 S &\sim W_p(n_1, \Sigma), \\
n_1 S_{11} &\sim W_{p_1}(n_1, \Sigma_{11}), \\
n\widehat{\Psi}_{11} &\sim W_{p_1}(n, \Sigma_{11}), \\
n_1\widehat{\Psi}_{22} &\sim W_{p_2}(n_1 - p_1, \Sigma_{22\cdot 1}),
\end{aligned}$$

where $W_d(m, \Omega)$ denotes Wishart distribution with the parameters m and Ω. These results are derived in Lemma 2.1 of Shutoh, Hyodo and Seo (2009). Now we consider of the conditional distribution of LDF W_m. By the MLEs, W_m is constructed as

$$W_m = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}\left[\boldsymbol{x} - \frac{1}{2}(\widehat{\boldsymbol{\mu}}^{(1)} + \widehat{\boldsymbol{\mu}}^{(2)})\right].$$

The probabilities of misclassification in W_m are

$$e_m(2|1) \equiv \Pr\left[W_m \le c|\boldsymbol{x} \in \Pi^{(1)}\right], \tag{5}$$

$$\begin{aligned} e_m(1|2) &\equiv \Pr\left[W_m > c|\boldsymbol{x} \in \Pi^{(2)}\right] \\ &= 1 - \Pr\left[W_m \le c|\boldsymbol{x} \in \Pi^{(2)}\right]. \end{aligned} \tag{6}$$

If we define

$$\begin{aligned} D_m^2 &= (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}), \\ F_m &= (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \boldsymbol{\mu}^{(1)}), \\ V_m &= (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}), \\ Z_m &= V_m^{-\frac{1}{2}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}^{(1)}), \end{aligned}$$

the probability (5) can be written by

$$\begin{aligned} &\Pr\left[\frac{W_m - \frac{1}{2}\Delta^2}{\Delta} \le u|\boldsymbol{x} \in \Pi^{(1)}\right] = \\ &\qquad \Pr\left[Z_m \le \left(u\Delta + F_m - \frac{1}{2}(D_m^2 - \Delta^2)\right) V_m^{-\frac{1}{2}}|\boldsymbol{x} \in \Pi^{(1)}\right], \end{aligned} \tag{7}$$

where $u = (c - (1/2)\Delta^2)/\Delta$. Under $\boldsymbol{x} \in \Pi^{(1)}$, the conditional distribution of Z_m given $\widehat{\boldsymbol{\mu}}^{(1)}, \widehat{\boldsymbol{\mu}}^{(2)}$ and $\widehat{\Sigma}$ is standard normal distribution. Therefore, the expectation of (5) can be expressed as

$$\mathrm{E}\left[\Phi\left(\left\{u\Delta + F_m - \frac{1}{2}(D_m^2 - \Delta^2)\right\} V_m^{-\frac{1}{2}}\right)|\boldsymbol{x} \in \Pi^{(1)}\right],$$

where $\mathrm{E}(\cdot)$ denotes expectation with respect to $\widehat{\boldsymbol{\mu}}^{(1)}, \widehat{\boldsymbol{\mu}}^{(2)}$ and $\widehat{\Sigma}$. Similarly, probability (6) is related to

$$\Pr\left[\frac{W_m + \frac{1}{2}\Delta^2}{\Delta} \le u|\boldsymbol{x} \in \Pi^{(2)}\right]. \tag{8}$$

Since we can achieve the purpose in this paper to consider of only (7), we show derivation for (7) mainly.

On the other hand, we can also consider the expectation of both estimated probabilities of misclassification under $c = 0$ which is similar to McLachlan (1973), i.e.,

$$\mathrm{E}\left[\Phi\left(-\frac{1}{2}D_m\right)\right]. \tag{9}$$

We will show derivations asymptotic expansions of (7) and (9) in the next section.

3. Derivation of asymptotic expansions

We prepare the following expressions of random vectors:

$$\boldsymbol{y}_{1T}^{(g)} = \sqrt{n}(\overline{\boldsymbol{x}}_{1T}^{(g)} - \boldsymbol{\mu}_1^{(g)}),\ \boldsymbol{y}_{\ell F}^{(g)} = \sqrt{n_1}(\overline{\boldsymbol{x}}_{\ell F}^{(g)} - \boldsymbol{\mu}_\ell^{(g)}),\ \boldsymbol{y}_{1L}^{(g)} = \sqrt{n_2}(\overline{\boldsymbol{x}}_{1L}^{(g)} - \boldsymbol{\mu}_1^{(g)}),$$

$$\boldsymbol{z}_{1T} = \sqrt{n}(\overline{\boldsymbol{x}}_{1T}^{(1)} - \overline{\boldsymbol{x}}_{1T}^{(2)} - \boldsymbol{\delta}_1),\quad \boldsymbol{z}_{\ell F} = \sqrt{n_1}(\overline{\boldsymbol{x}}_{\ell F}^{(1)} - \overline{\boldsymbol{x}}_{\ell F}^{(2)} - \boldsymbol{\delta}_\ell),$$

$$\boldsymbol{y}_F^{(g)} = \begin{pmatrix} \boldsymbol{y}_{1F}^{(g)} \\ \boldsymbol{y}_{2F}^{(g)} \end{pmatrix} = \sqrt{n_1}(\overline{\boldsymbol{x}}_F^{(g)} - \boldsymbol{\mu}^{(g)}),\quad \boldsymbol{z}_F = \begin{pmatrix} \boldsymbol{z}_{1F} \\ \boldsymbol{z}_{2F} \end{pmatrix} = \sqrt{n_1}(\overline{\boldsymbol{x}}_F^{(1)} - \overline{\boldsymbol{x}}_F^{(2)} - \boldsymbol{\delta}),$$

Similarly, we consider of random matrices:

$$T^{(1)} = \sqrt{n_1}(S - \Sigma), T_{\ell m}^{(1)} = \sqrt{n_1}(S_{\ell m} - \Sigma_{\ell m}), T^{(2)} = \sqrt{n_2}(S^{(2)} - \Sigma_{11}),$$

for $g = 1, 2$, $\ell = 1, 2$ and $m = 1, 2$. Also we can rewrite the MLEs of covariance matrix as

$$\begin{aligned}
\widehat{\Psi}_{11} &= \rho\Sigma_{11} + \frac{1}{\sqrt{n}}\left(\sqrt{\rho_1}T_{11}^{(1)} + \sqrt{\rho_2}T^{(2)}\right) + \frac{1}{n}\sum_{g=1}^{2}\boldsymbol{v}^{(g)}\boldsymbol{v}^{(g)\prime}, \\
\widehat{\Psi}_{12} &= \left(\Sigma_{11} + \frac{1}{\sqrt{n_1}}T_{11}^{(1)}\right)^{-1}\left(\Sigma_{12} + \frac{1}{\sqrt{n_1}}T_{12}^{(1)}\right),
\end{aligned}$$

$$
\begin{aligned}
\widehat{\Psi}_{22} &= \Sigma_{22\cdot 1} + \frac{1}{\sqrt{n_1}} T^{(1)}_{22\cdot 1}, \\
S &= \Sigma + \frac{1}{\sqrt{n_1}} T^{(1)},
\end{aligned}
$$

where

$$
\boldsymbol{v}^{(g)} = \sqrt{\frac{N_1^{(g)} N_2^{(g)}}{n N^{(g)}}} \left(\frac{1}{\sqrt{\rho_1}} \boldsymbol{y}^{(g)}_{1F} - \frac{1}{\sqrt{\rho_2}} \boldsymbol{y}^{(g)}_{1L} \right).
$$

By using

$$
\begin{aligned}
\left(I + \frac{1}{\sqrt{m}} A \right)^{-1} &= I + \sum_{j=1}^{\infty} (-1)^j m^{-\frac{j}{2}} A^j, \\
\left(I + \frac{1}{\sqrt{m}} A \right)^{-2} &= I + \sum_{j=1}^{\infty} (-1)^j (j+1) m^{-\frac{j}{2}} A^j,
\end{aligned}
$$

where A is a matrix with finite eigenvalues, D_m^2, F_m and V_m are expressed as

$$
\begin{aligned}
D_m^2 &\equiv \Delta_1^2 + \frac{1}{\sqrt{n}} D_{m1} + \frac{1}{n} D_{m2} + \frac{1}{n\sqrt{n}} R_d + \mathrm{O}_p(n^{-2}), && (10) \\
F_m &\equiv \frac{1}{\sqrt{n}} F_{m1} + \frac{1}{n} F_{m2} + \frac{1}{n\sqrt{n}} R_f + \mathrm{O}_p(n^{-2}), && (11) \\
V_m &\equiv \Delta_2^2 + \frac{1}{\sqrt{n}} V_{m1} + \frac{1}{n} V_{m2} + \frac{1}{n\sqrt{n}} R_v + \mathrm{O}_p(n^{-2}),
\end{aligned}
$$

where

$$
\Delta_1^2 = \Delta^2 - \left(1 - \frac{1}{\rho} \right) \delta_{11}^2, \ \Delta_2^2 = \Delta^2 - \left(1 - \frac{1}{\rho^2} \right) \delta_{11}^2, \ \delta_{11}^2 = \boldsymbol{\delta}_1' \Sigma_{11}^{-1} \boldsymbol{\delta}_1
$$

and R_d, R_f and R_v denotes homogeneous polynomial of degree 3 in the elements of random vectors and matrices. D_{mi}, F_{mi} and V_{mi} $(i = 1, 2)$ are presented in Appendix A. By using the above results, we derive two asymptotic expansions for discrimination based on 2-step monotone missing samples. We review the results derived by Shutoh and Seo (2010) in subsection 3.1. The main results in this paper are provided in subsections 3.2 and 3.3.

3.1. Asymptotic expansion for the studentized LDF

Shutoh and Seo (2010) considered an asymptotic expansion for the distributions of

$$\Pr\left[\frac{W_m - \frac{1}{2}D_m^2}{D_m} \le w|\boldsymbol{x} \in \Pi^{(1)}\right] \tag{12}$$

and

$$\Pr\left[\frac{W_m + \frac{1}{2}D_m^2}{D_m} \le w'|\boldsymbol{x} \in \Pi^{(2)}\right], \tag{13}$$

where $w = (c - (1/2)D_m^2)/D_m$. Therefore (12) can be rewritten as

$$\mathrm{E}\left[\Phi\left((wD_m + F_m)V_m^{-\frac{1}{2}}\right)\right].$$

Besides D_m and $V_m^{-\frac{1}{2}}$ have the following forms respectively:

$$\begin{aligned} D_m &= \Delta_1 + \frac{1}{\sqrt{n}}\left(\frac{1}{2\Delta_1}D_{m1}\right) + \frac{1}{n}\left(\frac{1}{2\Delta_1}D_{m2} - \frac{1}{8\Delta_1^3}D_{m1}^2\right) \\ &\quad + \frac{1}{n\sqrt{n}}R'_d + \mathrm{O}_p(n^{-2}) \end{aligned} \tag{14}$$

and

$$\begin{aligned} V_m^{-\frac{1}{2}} &= \frac{1}{\Delta_2} + \frac{1}{\sqrt{n}}\left(-\frac{1}{2\Delta_2^3}V_{m1}\right) + \frac{1}{n}\left(-\frac{1}{2\Delta_2^3}V_{m2} + \frac{3}{8\Delta_2^5}V_{m1}^2\right) \\ &\quad + \frac{1}{n\sqrt{n}}R'_v + \mathrm{O}_p(n^{-2}), \end{aligned} \tag{15}$$

where R'_d has the property same as R_d and R'_v has the property same as R_v. By Taylor series expansion of Φ and the results of expectations, Shutoh and Seo (2010) gave the following theorem.

Theorem 1. *The distributions for the studentized discriminant function of (12) and (13) can be expanded as*

$$\Phi(w) + \frac{1}{nr_1}\phi(w)\left[\frac{\Delta_{11}^2 + p_2 - 1}{\Delta}(1+k_1) - \left\{\left(p - \frac{1}{4} + \frac{k_1}{2}\right) + \frac{7}{4}\Delta_{11}^4 \right.\right.$$
$$\left.\left. - \left(p_1 + \frac{3}{2} + \frac{k_1}{2}\right)\Delta_{11}^2\right\}w - \frac{1}{4}(1-\Delta_{11}^4)w^3\right]$$
$$+\frac{1}{n}\phi(w)\left[\frac{p_1 - \Delta_{11}^2}{\Delta}(1+k) - \left\{\left(p_1 + \frac{1}{2} + \frac{k}{2}\right)\Delta_{11}^2 - \frac{7}{4}\Delta_{11}^4\right\}w\right.$$
$$\left. - \frac{1}{4}\Delta_{11}^4 w^3\right] + \mathrm{O}(n^{-2})$$

and

$$\Phi(w') - \frac{1}{nr_1}\phi(w')\left[\frac{\Delta_{11}^2 + p_2 - 1}{\Delta}\left(1 + \frac{1}{k_1}\right) + \left\{\left(p - \frac{1}{4} + \frac{1}{2k_1}\right)\right.\right.$$
$$\left.\left. + \frac{7}{4}\Delta_{11}^4 - \left(p_1 + \frac{3}{2} + \frac{1}{2k_1}\right)\Delta_{11}^2\right\}w' + \frac{1}{4}(1-\Delta_{11}^4)w'^3\right]$$
$$-\frac{1}{n}\phi(w')\left[\frac{p_1 - \Delta_{11}^2}{\Delta}\left(1 + \frac{1}{k}\right) + \left\{\left(p_1 + \frac{1}{2} + \frac{1}{2k}\right)\Delta_{11}^2 - \frac{7}{4}\Delta_{11}^4\right\}w'\right.$$
$$\left. + \frac{1}{4}\Delta_{11}^4 w'^3\right] + \mathrm{O}(n^{-2})$$

respectively, where $\Delta_{11} = \delta_{11}/\Delta$.

3.2. Asymptotic expansion for LDF

In this subsection, we consider the asymptotic expansion for the distribution of linear discriminant function based on 2-step monotone missing samples, i.e., the versions for 2-step monotone missing samples of Okamoto (1963). By (10), (11) and (15), we can obtain

$$\left(u\Delta + F_m - \frac{1}{2}(D_m^2 - \Delta^2)\right) V_m^{-\frac{1}{2}}$$
$$= u^* + \frac{1}{\sqrt{n}}g_1 + \frac{1}{n}g_2 + \frac{1}{n\sqrt{n}}R_1 + \mathrm{O}_p(n^{-2}),$$

where R_1 denotes homogeneous polynomial of degree 3 in the elements of random vectors and matrices,

$$\begin{aligned} u^* &= \left(u\Delta + \frac{1}{2}\left(1 - \frac{1}{\rho}\right)\delta_{11}^2\right)\Delta_2^{-1}, \\ g_1 &= \frac{1}{\Delta_2}F_{m1} - \frac{1}{2\Delta_2}D_{m1} - \frac{1}{2\Delta_2^2}u^*V_{m1}, \\ g_2 &= \frac{1}{\Delta_2}F_{m2} - \frac{1}{2\Delta_2}D_{m2} - \frac{1}{2\Delta_2^2}u^*V_{m2} + \frac{3}{8\Delta_2^4}u^*V_{m1}^2 - \frac{1}{2\Delta_2^3}F_{m1}V_{m1} \\ &\quad + \frac{1}{4\Delta_2^3}D_{m1}V_{m1}. \end{aligned}$$

Thus, by Taylor series expansion of Φ, we can rewrite (7) as

$$\Phi(u^*) + \phi(u^*)\left[\frac{1}{\sqrt{n}}\mathrm{E}(g_1) + \frac{1}{n}\left(\mathrm{E}(g_2) - \frac{1}{2}u^*\mathrm{E}(g_1^2)\right) + \frac{1}{n\sqrt{n}}\mathrm{E}(R_2)\right] + \mathrm{O}(n^{-2}),$$

where R_2 has the property same as R_1. By making use of expectations derived by Lemmas in Section 4, we can obtain the following theorem. Besides the terms included in $\mathrm{E}(R_2)/(n\sqrt{n})$ are either 0 or $\mathrm{O}(n^{-2})$. For details of expectations, see Shutoh and Seo (2010).

Theorem 2. *The distributions of linear discriminant function based on 2-step monotone missing samples described in (7) and (8) can be expanded as*

$$\begin{aligned} &P_{(2|1)}^m(u, N_1^{(1)}, N_1^{(2)}, N_2^{(1)}, N_2^{(2)}, \Delta, \delta_{11}) \\ &= \Phi(u) - \frac{1}{2}\left[\frac{a_{11}}{N_1^{(1)}\Delta^2} + \frac{a_{12}}{N_1^{(2)}\Delta^2} + \frac{a_{13}}{2n_1} + \frac{a_{21}}{N^{(1)}\Delta^2} + \frac{a_{22}}{N^{(2)}\Delta^2} + \frac{a_{23}}{2n}\right]\phi(u) \\ &\quad + \mathrm{O}_2, \end{aligned}$$

and $1 - P_{(2|1)}^m(-u, N_1^{(2)}, N_1^{(1)}, N_2^{(2)}, N_2^{(1)}, \Delta, \delta_{11})$ *respectively, where* O_2 *denotes*

$\mathrm{O}(N_1^{(1)^{-2}}, N_1^{(2)^{-2}}, n_1^{-2}, N^{(1)^{-2}}, N^{(2)^{-2}}, n^{-2})$,

$$\begin{aligned}
a_{11} &= (1-\Delta_{11}^2)u^3 + (p_2 - 3(1-\Delta_{11}^2))u - p_2\Delta,\\
a_{12} &= (1-\Delta_{11}^2)u^3 + 2(\Delta - \delta_{11}\Delta_{11})u^2 + (p_2 - 3(1-\Delta_{11}^2) + \Delta^2 - \delta_{11}^2)u\\
&\quad +(p_2-2)\Delta + 2\delta_{11}\Delta_{11},\\
a_{13} &= 4(1-\Delta_{11}^4)u^3 + 4(\Delta - \delta_{11}\Delta_{11}^3)u^2\\
&\quad +(6p - 6 - 6(p_1+1)\Delta_{11}^2 + 12\Delta_{11}^4 + \Delta^2 - \delta_{11}^2\Delta_{11}^2)u\\
&\quad +2(p-1)\Delta - 2(p_1+1)\delta_{11}\Delta_{11} + 4\delta_{11}\Delta_{11}^3,\\
a_{21} &= \Delta_{11}^2u^3 + (p_1 - 3\Delta_{11}^2)u - p_1\Delta,\\
a_{22} &= \Delta_{11}^2u^3 + 2\delta_{11}\Delta_{11}u^2 + (p_1 - 3\Delta_{11}^2 + \delta_{11}^2)u + p_1\Delta - 2\delta_{11}\Delta_{11},\\
a_{23} &= 4\Delta_{11}^4u^3 + 4\delta_{11}\Delta_{11}^3u^2 + (2(3p_1-1) - 12\Delta_{11}^2 + \delta_{11}^2)\Delta_{11}^2u\\
&\quad +((p_1-1)\delta_{11} - 2\delta_{11}\Delta_{11}^2)2\Delta_{11},\\
\Delta_{11} &= \delta_{11}/\Delta.
\end{aligned}$$

If Ψ_{11} is estimated by the simple estimator

$$\widetilde{\Psi}_{11} = \frac{1}{n_{12}}(n_1S_{11} + n_2S^{(2)}), \quad n_{12} = n_1 + n_2,$$

the distributions of (7) and (8) are expanded as

$$\begin{aligned}
&P_{(2|1)}^m(u, N_1^{(1)}, N_1^{(2)}, N_2^{(1)}, N_2^{(2)}, \Delta, \delta_{11}) \qquad (16)\\
&= \Phi(u) - \frac{1}{2}\left[\frac{a_{11}}{N_1^{(1)}\Delta^2} + \frac{a_{12}}{N_1^{(2)}\Delta^2} + \frac{a_{13}}{2n_1} + \frac{a_{21}}{N^{(1)}\Delta^2} + \frac{a_{22}}{N^{(2)}\Delta^2} + \frac{a_{23}^*}{2n_{12}}\right]\phi(u)\\
&\quad + \mathrm{O}_2^*,
\end{aligned}$$

and $1 - P_{(2|1)}^m(-u, N_1^{(2)}, N_1^{(1)}, N_2^{(2)}, N_2^{(1)}, \Delta, \delta_{11})$ respectively, where O_2^* denotes $\mathrm{O}(N_1^{(1)^{-2}}, N_1^{(2)^{-2}}, n_1^{-2}, N^{(1)^{-2}}, N^{(2)^{-2}}, n_{12}^{-2})$ and

$$\begin{aligned}
a_{23}^* &= 4\Delta_{11}^4u^3 + 4\delta_{11}\Delta_{11}^3u^2 + (6(p_1+1) - 12\Delta_{11}^2 + \delta_{11}^2)\Delta_{11}^2u\\
&\quad +((p_1+1)\delta_{11} - 2\delta_{11}\Delta_{11}^2)2\Delta_{11}.
\end{aligned}$$

Then we can obtain the following corollary.

Corollary 3. *The results obtained in (16) can be reduced to (3) up to the terms of the first order of* $(N_1^{(1)^{-1}}, N_1^{(2)^{-1}}, n_1^{-1})$ *by putting* $N^{(g)} = N_1^{(g)}$*, i.e.,* $N_2^{(g)} = 0$*.*

3.3. Asymptotic expansion for the expectation of estimated probabilities of misclassification

In this subsection, we derive the asymptotic expansion for the expectation of estimated probabilities of misclassification, i.e., the versions for 2-step monotone missing samples of McLachlan (1973). By (14), we can obtain

$$-\frac{1}{2}D_m = -\frac{1}{2}\Delta_1 + \frac{1}{\sqrt{n}}h_1 + \frac{1}{n}h_2 + \frac{1}{n\sqrt{n}}R_3 + \mathrm{O}_p(n^{-2}),$$

where

$$\begin{aligned} h_1 &= -\frac{1}{4\Delta_1}D_{m1}, \\ h_2 &= -\frac{1}{4\Delta_1}D_{m2} + \frac{1}{16\Delta_1^3}D_{m1}^2 \end{aligned}$$

and R_3 denotes the terms which have the property same as R_1. Therefore, by Taylor series expansion of Φ, we can rewrite (9) as

$$\begin{aligned} &\Phi\left(-\frac{1}{2}\Delta_1\right) + \phi\left(-\frac{1}{2}\Delta_1\right)\left[\frac{1}{\sqrt{n}}\mathrm{E}(h_1)\right. \\ &\left. + \frac{1}{n}\left(\mathrm{E}(h_2) + \frac{1}{4}\Delta_1\mathrm{E}(h_1^2)\right) + \frac{1}{n\sqrt{n}}\mathrm{E}(R_4)\right] + \mathrm{O}(n^{-2}), \end{aligned}$$

where R_4 denotes the terms which have the property same as R_1. Since we find

$$h_2 + \frac{1}{4}\Delta_1 h_1^2 = -\frac{1}{4\Delta_1}D_{m2} + \frac{1}{64\Delta_1^3}(\Delta_1^2 + 4)D_{m1}^2$$

and the terms included in $\mathrm{E}(R_4)/(n\sqrt{n})$ are either 0 or $\mathrm{O}(n^{-2})$, we can also obtain the following theorem by making use of expectations derived in Section 4.

Theorem 4. *The expectation of estimator of the probabilities of misclassification based on 2-step monotone missing samples similar to McLachlan (1973) is expanded as follows:*

$$\mathrm{E}\left[\Phi\left(-\frac{1}{2}D_m\right)\right] = \Phi\left(-\frac{1}{2}\Delta\right) + b_{m1} + \mathrm{O}_2,$$

where

$$\begin{aligned} b_{m1} &= \frac{1}{16}\phi\Big(-\frac{1}{2}\Delta\Big)\Big[\Big(\frac{1}{N_1^{(1)}} + \frac{1}{N_1^{(2)}}\Big)\{\Delta - \frac{4}{\Delta}(p_2-1) - \frac{\Delta^2+4}{\Delta}\Delta_{11}^2\} \\ &\quad + \Big(\frac{1}{N^{(1)}} + \frac{1}{N^{(2)}}\Big)\{-\frac{4}{\Delta}p_1 + \frac{\Delta^2+4}{\Delta}\Delta_{11}^2\} \\ &\quad + \frac{\Delta}{2n_1}\{\Delta^2 - 4(2p+1) + 8(p_1+1)\Delta_{11}^2 - (\Delta^2+4)\Delta_{11}^4\} \\ &\quad + \frac{\Delta}{2n}\{-8(p_1-1)\Delta_{11}^2 + (\Delta^2+4)\Delta_{11}^4\}\Big], \\ \Delta_{11} &= \delta_{11}/\Delta. \end{aligned}$$

If Ψ_{11} is estimated by $\widetilde{\Psi}_{11}$,

$$\mathrm{E}\left[\Phi\left(-\frac{1}{2}D_m\right)\right] = \Phi\left(-\frac{1}{2}\Delta\right) + b_{m1}^* + \mathrm{O}_2^*, \tag{17}$$

where

$$\begin{aligned} b_{m1}^* &= \frac{1}{16}\phi\Big(-\frac{1}{2}\Delta\Big)\Big[\Big(\frac{1}{N_1^{(1)}} + \frac{1}{N_1^{(2)}}\Big)\{\Delta - \frac{4}{\Delta}(p_2-1) - \frac{\Delta^2+4}{\Delta}\Delta_{11}^2\} \\ &\quad + \Big(\frac{1}{N^{(1)}} + \frac{1}{N^{(2)}}\Big)\{-\frac{4}{\Delta}p_1 + \frac{\Delta^2+4}{\Delta}\Delta_{11}^2\} \\ &\quad + \frac{\Delta}{2n_1}\{\Delta^2 - 4(2p+1) + 8(p_1+1)\Delta_{11}^2 - (\Delta^2+4)\Delta_{11}^4\} \\ &\quad + \frac{\Delta}{2n_{12}}\{-8(p_1+1)\Delta_{11}^2 + (\Delta^2+4)\Delta_{11}^4\}\Big]. \end{aligned}$$

Also we can find the following corollary.

Corollary 5. *The results obtained in (17) can be reduced to (4) up to the terms of the first order of* $(N_1^{(1)^{-1}}, N_1^{(2)^{-1}}, n_1^{-1})$ *by putting* $N^{(g)} = N_1^{(g)}$*, i.e.,* $N_2^{(g)} = 0$*.*

4. On the results for obtaining asymptotic expansions

For obtaining asymptotic expansions, we need some results of inverse of partitioned matrix and expectations of random vectors or matrices. First we give a Lemma for inverse of partitioned matrix.

Lemma 6. *Let* Ω *be the following partitioned matrix:*

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

where Ω *and* $\Omega_{\ell m}$ *denotes* $d \times d$ *matrix and* $d_\ell \times d_m$ *partitioned matrix of* Ω*, i.e.,* $d = d_1 + d_2$*. If* Ω_{11} *and* $\Omega_{22\cdot 1}$ *are nonsingular, then* Ω^{-1} *is*

$$\Omega^{-1} = \begin{pmatrix} \Omega_{11}^{-1} + \Omega_{11}^{-1}\Omega_{12}\Omega_{22\cdot 1}^{-1}\Omega_{21}\Omega_{11}^{-1} & -\Omega_{11}^{-1}\Omega_{12}\Omega_{22\cdot 1}^{-1} \\ -\Omega_{22\cdot 1}^{-1}\Omega_{21}\Omega_{11}^{-1} & \Omega_{22\cdot 1}^{-1} \end{pmatrix}$$

and it holds that

$$\Omega^{-1} \begin{pmatrix} \Omega_{11} \\ \Omega_{21} \end{pmatrix} \Omega_{11}^{-1} = \begin{pmatrix} \Omega_{11}^{-1} \\ O_{21} \end{pmatrix},$$

where $O_{\ell m}$ *denotes* $d_\ell \times d_m$ *matrix with* 0*'s.*

Also, by making use of expectations and the conditional distribution for partitioned vectors which are distributed as normal distribution, Shutoh and Seo (2010) showed the following results concerning partitioned Wishart matrix.

Lemma 7. *Suppose that G has $W_d(m, \Omega)$. Then the following expectations can be obtained:*

$$
\begin{aligned}
(i)\quad & \mathrm{E}(G_{12}C_{21}G_{11}) = m^2\Omega_{12}C_{21}\Omega_{11} + m\Omega_{11}C_{21}'\Omega_{21} + m\mathrm{tr}(C_{21}\Omega_{12})\Omega_{11}, \\
(ii)\quad & \mathrm{E}(G_{12}C_{22}G_{21}) = m^2\Omega_{12}C_{22}\Omega_{21} + m\Omega_{12}C_{22}'\Omega_{21} \\
& \qquad + m\mathrm{tr}(C_{22}\Omega_{22\cdot 1})\Omega_{11} + m\mathrm{tr}(C_{22}\Omega_{21}\Omega_{11}^{-1}\Omega_{12})\Omega_{11}, \\
(iii)\quad & \mathrm{E}(G_{22}C_{21}G_{11}) = m^2\Omega_{22\cdot 1}C_{21}\Omega_{11} + m^2\Omega_{21}\Omega_{11}^{-1}\Omega_{12}C_{21}\Omega_{11} \\
& \qquad + m\Omega_{21}C_{21}'\Omega_{21} + m\mathrm{tr}(C_{21}\Omega_{12})\Omega_{21}, \\
(iv)\quad & \mathrm{E}(G_{21}C_{11}G_{11}) = m^2\Omega_{21}C_{11}\Omega_{11} + m\Omega_{21}C_{11}'\Omega_{11} + m\mathrm{tr}(C_{11}\Omega_{11})\Omega_{21},
\end{aligned}
$$

where $G_{\ell m}$ denotes $d_\ell \times d_m$ partitioned matrix of G, $C_{\ell m}$ denotes $d_\ell \times d_m$ constant matrix and $d = d_1 + d_2$ respectively.

For details of the derivation, see Shutoh and Seo (2010). By the expectations with limit of ρ, we can obtain the asymptotic expansions shown in Theorem 2 and Theorem 4 with noting $\Delta_1 = \Delta_2 = \Delta$ and $u^* = u$ in asymptotic sense.

5. Simulation studies

We can obtain two asymptotic expansions in Section 3. Therefore we evaluate accuracy in approximating the probabilities of misclassification by Monte Carlo simulations with selected parameters under $\boldsymbol{x} \in \Pi^{(1)}$ and $c = 0$.

In simulations, especially, we compare the following results:

- The result of subsection 3.2 denoted by O–type,
- The result of Shutoh, Hyodo and Seo (2009) denoted by L–type.

Besides Shutoh, Hyodo and Seo (2009) is the version for 2-step monotone missing samples of Lachenbruch (1968) and regarded as both type I and type II approximation.

All the values are compared with expected probabilities of misclassification (EPMC) of W_m calculated by simulations under various settings of sample sizes and dimensionality. As concerns Δ, we are interested in the situation under small Δ since the approximations of this type play an important role if the exact probabilities of misclassification are higher. Thus, in these simulations, we set the value of Δ as 1.05. Then the exact probabilities of misclassification are nearly equal to 0.2998.

In all simulations, for $i = 1, 2$, the sample sizes $N_i^{(1)}$ and $N_i^{(2)}$ are equal and they are denoted by M_i respectively. Also M_1 and M_2 have same order. Since two types of approximations include unknown Δ and δ_{11}, we use estimators of them.

Tables 1–3 give the results under $p = 3, 5, 9$ and $\delta_{11} = \Delta$ respectively. In approximating EPMC, we can see that O–type provides better approximation for $p = 3$ and $p = 5$. Also we can see large dimensionality makes O–type poorer. For this case, L–type has good accuracy. Compare the case of $M_2 = 0$ and that of $M_2 > 0$ which have equal M_1, e.g., the first row and the second row for all the three tables. It seems that both of two approximations derived by Okamoto (1963) and Lachenbruch (1968) are improved by the versions for 2-step monotone missing samples.

In Tables 1–3, the results under $\delta_{11} = \Delta$ are presented. However, under $M_2 > 0$, the result of O–type given in Theorem 2 depends on ratio of δ_{11} to Δ. Therefore, in Table 4, the simulation results under $\Delta_{11} =$

$0.2, 0.4, 0.6, 0.7, 0.8, 1.0$ and $p = 3$ are presented. Note that $\Delta_{11} = 0$ implies p_1 variables are redundant for discrimination. We see that larger Δ_{11} makes the values of approximation proposed in Theorem 2 lower and EPMC are not influenced by Δ_{11}. By these simulations, our result is useful for large Δ_{11} in approximating EPMC.

6. Conclusion and future problems

This paper has provided extensions of Okamoto (1963) and McLachlan (1973) in the case of 2-step monotone missing samples up to the terms of the first order. Also the type of Okamoto (1963) given in subsection 3.2 has been evaluated numerically by comparing with Shutoh, Hyodo and Seo (2009). In these numerical evaluations, with small dimensionality and large Δ_{11}, we have seen that our result have better accuracy. However, in approximating EPMC, approximations for the type of Okamoto (1963) are lower than EPMC. The two results provided in this paper may lead to discussion for interval estimations for the probabilities of misclassification mentioned by McLachlan (1975) in the case of 2-step monotone missing samples and it is one of the future problems. On the other approximation for EPMC in 2-step monotone missing samples, the approximation similar to Fujikoshi and Seo (1998) will be needed since the result have good accuracy for large dimensionality. As the other future problems, the discussions for maximum likelihood discrimination will be considerable. Also, by making use of Fujikoshi, Ulyanov and Shimizu (2010), we may consider the error bounds for the results of this paper, Shutoh, Hyodo and Seo (2009) and Shutoh and Seo (2010) when Ψ_{11} is estimated by $\widetilde{\Psi}_{11}$.

Appendix A. Details for the terms

The terms included in D_m^2, F_m and V_m are as follows:

$$
\begin{aligned}
D_{m1} &= \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{z}_F - \frac{1}{\sqrt{\rho_1}}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{\delta} - \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
&\quad + \frac{1}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 + \frac{2}{\rho}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{z}_{1T} \\
&\quad - \frac{\sqrt{\rho_1}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 - \frac{\sqrt{\rho_2}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1, \\
D_{m2} &= \frac{1}{\rho_1}\boldsymbol{z}_F'\Sigma^{-1}\boldsymbol{z}_F - \frac{2}{\rho_1}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{z}_F + \frac{1}{\rho_1}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{\delta} \\
&\quad - \frac{1}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}\boldsymbol{z}_{1F} + \frac{2}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
&\quad - \frac{1}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 + \frac{1}{\rho}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}\boldsymbol{z}_{1T} \\
&\quad - \frac{2\sqrt{\rho_1}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1T} - \frac{2\sqrt{\rho_2}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{z}_{1T} \\
&\quad + \frac{\rho_1}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 + \frac{2\sqrt{\rho_1\rho_2}}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
&\quad + \frac{\rho_2}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 - \frac{1}{\rho^2}\sum_{g=1}^{2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{v}^{(g)}\boldsymbol{v}^{(g)\prime}\Sigma_{11}^{-1}\boldsymbol{\delta}_1, \\
F_{m1} &= \frac{1}{\sqrt{\rho_1}}\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{y}_F^{(1)} - \frac{1}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{y}_{1F}^{(1)} + \frac{1}{\rho}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{y}_{1T}^{(1)}, \\
F_{m2} &= \frac{1}{\rho_1}\boldsymbol{z}_F'\Sigma^{-1}\boldsymbol{y}_F^{(1)} - \frac{1}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}\boldsymbol{y}_{1F}^{(1)} - \frac{1}{\rho_1}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{y}_F^{(1)} \\
&\quad + \frac{1}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{y}_{1F}^{(1)} + \frac{1}{\rho}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}\boldsymbol{y}_{1T}^{(1)} - \frac{\sqrt{\rho_1}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{y}_{1T}^{(1)} \\
&\quad - \frac{\sqrt{\rho_2}}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{y}_{1T}^{(1)}, \\
V_{m1} &= \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{z}_F - \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{\delta} - \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
&\quad + \frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{2}{\sqrt{\rho_1}}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& -\frac{2}{\sqrt{\rho_1}}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& -\frac{2}{\sqrt{\rho_1}}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2 \\
& +\frac{2}{\sqrt{\rho_1}}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2 \\
& +\frac{2}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{z}_{1T}-\frac{2\sqrt{\rho_1}}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& -\frac{2\sqrt{\rho_2}}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1, \\
V_{m2} \;=\; & \frac{1}{\rho_1}\boldsymbol{z}_F'\Sigma^{-1}\boldsymbol{z}_F-\frac{4}{\rho_1}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{z}_F+\frac{3}{\rho_1}\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{\delta} \\
& -\frac{1}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}\boldsymbol{z}_{1F}+\frac{4}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
& -\frac{3}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& +\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
& -\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
& -\frac{2}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
& +\frac{2}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F} \\
& +\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& -\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& +\frac{2}{\rho_1}\left(1-\frac{2}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1 \\
& +\frac{2}{\rho\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1
\end{aligned}
$$

$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2+\frac{2}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{z}_{2F}$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{z}_{2F}$$
$$-\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$+\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$+\frac{1}{\rho^2}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}\boldsymbol{z}_{1T}-\frac{4\sqrt{\rho_1}}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1T}$$
$$-\frac{4\sqrt{\rho_2}}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{z}_{1T}$$
$$+\frac{3\rho_1}{\rho^4}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1+\frac{6\sqrt{\rho_1\rho_2}}{\rho^4}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{3\rho_2}{\rho^4}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho\sqrt{\rho_1}}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho\sqrt{\rho_1}}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$

$$
\begin{aligned}
&-\frac{2}{\rho^2}\boldsymbol{\delta}'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1\\
&+\frac{2}{\rho^2}\sqrt{\frac{\rho_2}{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1\\
&-\frac{2}{\rho^2}\sqrt{\frac{\rho_2}{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1\\
&+\frac{2}{\rho\sqrt{\rho_1}}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&-\frac{2}{\rho\sqrt{\rho_1}}\boldsymbol{z}_{1T}'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&-\frac{2}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&+\frac{2}{\rho^2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&-\frac{2}{\rho^2}\sqrt{\frac{\rho_2}{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&+\frac{2}{\rho^2}\sqrt{\frac{\rho_2}{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}\boldsymbol{\delta}_2\\
&-\frac{2}{\rho^3}\sum_{g=1}^{2}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\boldsymbol{v}^{(g)}\boldsymbol{v}^{(g)\prime}\Sigma_{11}^{-1}\boldsymbol{\delta}_1.
\end{aligned}
$$

Acknowledgements

The author is grateful to Professor Takashi Seo of Tokyo University of Science for many helpful suggestions.

References

Anderson, T. W., 1973. An asymptotic expansion of the distribution of the studentized classification statistic W. Ann. Statist. 1, 964–972.

Anderson, T. W. and Olkin, I., 1985. Maximum-likelihood estimation of the parameters of a multivariate normal distribution. Linear Algebra Appl. 70, 147–171.

Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R., 2010. *Multivariate Statistics High-Dimensional and Large-Sample Approximations*. John Wiley & Sons, Inc., Hoboken, New Jersey, pp.467–479.

Fujikoshi, Y. and Seo, T., 1998. Asymptotic approximations for EPMC's of the linear and quadratic discriminant functions when the sample sizes and the dimension are large. Random Oper. Stochastic Equations 6, 269–280.

Lachenbruch, P. A., 1968. On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficient. Biometrics 24, 823–834.

McLachlan, G. J., 1973. An asymptotic expansion of the expectation of the estimated error rate in discriminant analysis. Austral. J. Statist. 15, 210–214.

McLachlan, G. J., 1975. Confidence intervals for the conditional probability of misallocation in discriminant analysis. Biometrics. 31, 161–167.

Memon, A. Z. and Okamoto, M., 1971. Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. J. Multivariate Anal. 1, 294–307.

Okamoto, M., 1963. An asymptotic expansion for the distribution of the linear discriminant function. Ann. Math. Statist. 34, 1268–1301.

Okamoto, M., 1968. Correction to "An asymptotic expansion for the distribution of the linear discriminant function". Ann. Math. Statist. 39, 1358–1359.

Shutoh, N., Hyodo, M. and Seo, T., 2009. Expected probabilities of misclassification in linear discriminant analysis based on 2-Step monotone missing samples. Technical Report 09–01, Hiroshima Statistical Research Group, Hiroshima University.

Shutoh, N. and Seo, T., 2010. Asymptotic expansion of the distribution of the studentized linear discriminant function with 2-Step monotone missing data. Technical Report 10–01, Hiroshima Statistical Research Group, Hiroshima University.

Table 1 The values of approximations and EPMC when

$p = 3\ (p_1 = 2,\ p_2 = 1)$.

M_1	M_2	O–type	L–type	EPMC
10	0	0.3258	0.3960	0.3470
10	10	0.3239	0.3660	0.3344
15	0	0.3161	0.3643	0.3309
15	10	0.3150	0.3461	0.3236
15	15	0.3146	0.3416	0.3224
20	0	0.3115	0.3478	0.3234
20	20	0.3105	0.3303	0.3162
40	40	0.3049	0.3145	0.3083

Table 2 The values of approximations and EPMC when

$p = 5\ (p_1 = 3,\ p_2 = 2)$.

M_1	M_2	O–type	L–type	EPMC
10	0	0.3354	0.4189	0.3702
10	10	0.3338	0.3981	0.3585
15	0	0.3218	0.3829	0.3511
15	10	0.3210	0.3665	0.3427
15	15	0.3207	0.3627	0.3403
20	0	0.3159	0.3635	0.3397
20	20	0.3150	0.3461	0.3311
40	40	0.3072	0.3224	0.3157

Table 3 The values of approximations and EPMC when

$p = 9\ (p_1 = 5,\ p_2 = 4)$.

M_1	M_2	O–type	L–type	EPMC
20	0	0.3232	0.3885	0.3658
20	20	0.3228	0.3743	0.3562
40	40	0.3112	0.3372	0.3304

Table 4 The values of approximations of O–type and EPMC when $p = 3$ $(p_1 = 2,\ p_2 = 1)$.

M_1	M_2	$\Delta_{11} = 0.2$	EPMC
10	10	0.3269	0.3337
15	10	0.3164	0.3232
15	15	0.3161	0.3202
20	20	0.3118	0.3154
40	40	0.3054	0.3072
M_1	M_2	$\Delta_{11} = 0.4$	EPMC
10	10	0.3264	0.3334
15	10	0.3162	0.3233
15	15	0.3159	0.3204
20	20	0.3116	0.3155
40	40	0.3053	0.3071
M_1	M_2	$\Delta_{11} = 0.6$	EPMC
10	10	0.3256	0.3335
15	10	0.3159	0.3234
15	15	0.3155	0.3209
20	20	0.3113	0.3162
40	40	0.3052	0.3075
M_1	M_2	$\Delta_{11} = 0.7$	EPMC
10	10	0.3252	0.3336
15	10	0.3157	0.3232
15	15	0.3153	0.3212
20	20	0.3111	0.3163
40	40	0.3051	0.3077
M_1	M_2	$\Delta_{11} = 0.8$	EPMC
10	10	0.3248	0.3338
15	10	0.3154	0.3234
15	15	0.3150	0.3215
20	20	0.3110	0.3162
40	40	0.3050	0.3079
M_1	M_2	$\Delta_{11} = 1.0$	EPMC
10	10	0.3239	0.3344
15	10	0.3150	0.3236
15	15	0.3146	0.3224
20	20	0.3105	0.3162
40	40	0.3049	0.3083