Asymptotic expansion of the distribution of the studentized linear discriminant function with 2-Step monotone missing data

Nobumichi Shutoh*[,a], Takashi Seo[b]

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

Abstract

In discriminant analysis, it is important to evaluate probabilities of misclassification. Therefore some asymptotic approximations have been proposed in case that all the sample vectors do not have missing data. Also it is known that asymptotic expansions of the distribution of discriminant function are useful techniques for considering asymptotic approximation in sample vectors with small dimension. By using perturbation method, an extension for Anderson (1973) in the case of 2-Step monotone missing samples is given. Finally, numerical evaluations by Monte Carlo simulations are given.

Key words: linear 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 multivariate analysis, discriminant analysis has been considered as a statistical method for discriminating p dimensional sample vector $\boldsymbol{x}$ which arises from any of some groups. In this paper, we discuss linear discriminant analysis under the situation of discriminating $\boldsymbol{x}$ which comes from one of two groups $\Pi^{(1)}$ and $\Pi^{(2)}$ having multivariate normality, i.e., $\Pi^{(g)} : N_p(\boldsymbol{\mu}^{(g)}, \Sigma)$.

If the parameters are unknown, linear discriminant function (LDF) is constructed as following form with estimators:

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

where $\overline{\boldsymbol{x}}_F^{(g)}$ denotes p dimensional sample mean vector based on $N_1^{(g)}$ sample vectors from $\Pi^{(g)}$, S denotes unbiased pooled sample covariance matrix for two groups $\Pi^{(1)}$ and $\Pi^{(2)}$ and they are based on maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}^{(g)}$ and Σ. With cut-off point c which depends on a priori probabilities of drawing an observation from $\Pi^{(g)}$ and costs of discrimination, $\boldsymbol{x}$ may be assigned to $\Pi^{(1)}$ if $W > c$, or $\Pi^{(2)}$ otherwise. Therefore the probabilities of misclassification in linear discriminant analysis can be considered as following expressions:

$$e(2|1) \equiv \Pr\left[W \le c | \boldsymbol{x} \in \Pi^{(1)} \right] \tag{1}$$

and

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

It is difficult to obtain $e(2|1)$ and $e(1|2)$ since W includes estimators. However we can find asymptotic distributions of W as $N_1^{(1)}$ and $N_1^{(2)}$ tends to infinity: W is distributed as $N((-1)^{g-1}(1/2)\Delta^2, \Delta^2)$ asymptotically where $\Delta^2 = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \Sigma^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ under $\boldsymbol{x} \in \Pi^{(g)}$.

Okamoto (1963) (with correction, Okamoto (1968)) gave the asymptotic expansions for the distributions of $(W - (-1)^{g-1}(1/2)\Delta^2)/\Delta$ under $\boldsymbol{x} \in \Pi^{(g)}$ for each g by using differential operator. With positive constant k_1 which is limit of $N_1^{(2)}/N_1^{(1)}$ for large $N_1^{(1)}$ and $N_1^{(2)}$, Anderson (1973) considered that for $(W - (-1)^{g-1}(1/2)D^2)/D$ by using perturbation method where D^2 denotes sample mahalanobis distance between $\Pi^{(1)}$ and $\Pi^{(2)}$ based on $N_1^{(1)}$

and $N_1^{(2)}$ sample vectors:

$$\Pr\left[\frac{W-\frac{1}{2}D^2}{D} \le v|\boldsymbol{x} \in \Pi^{(1)}\right] \tag{3}$$
$$= \Phi(v) + \frac{1}{n_1}\phi(v)\left[\frac{p-1}{\Delta}(1+k_1) - \left(p - \frac{1}{4} + \frac{1}{2}k_1\right)v - \frac{1}{4}v^3\right]$$
$$+ \mathrm{O}(n_1^{-2})$$

and

$$\Pr\left[\frac{W+\frac{1}{2}D^2}{D} \le v'|\boldsymbol{x} \in \Pi^{(2)}\right] \tag{4}$$
$$= \Phi(v') - \frac{1}{n_1}\phi(v')\left[\frac{p-1}{\Delta}\left(1+\frac{1}{k_1}\right) + \left(p - \frac{1}{4} + \frac{1}{2k_1}\right)v' + \frac{1}{4}v'^3\right]$$
$$+ \mathrm{O}(n_1^{-2}),$$

where $v = (c - (1/2)D^2)/D$ in (3), $n_1 = N_1^{(1)} + N_1^{(2)} - 2$, $\Phi(\cdot)$ denotes the cumulative density function of standard normal distribution and $\phi(\cdot)$ denotes the probability density function of that. They can be regarded as type I approximation for (1) and (2). Type I approximation is that under the framework $N_1^{(1)} \to \infty$, $N_1^{(2)} \to \infty$ and $N_1^{(2)}/N_1^{(1)} \to$ positive constant k_1. In maximum likelihood rule, also the results similar to the above have been derived. For details, see Memon and Okamoto (1971) and Fujikoshi and Kanazawa (1976).

In addition, without asymptotic expansions, some researchers have derived asymptotic approximations for the probabilities of (1) and (2). For instance, Lachenbruch (1968) has considered by using asymptotic normality of W. Also, in Fujikoshi and Seo (1998), they have derived by rewriting W and maximum likelihood discriminant function with expressions of central and non–central chi–square random variables and making use of their asymptotic properties. Error bounds for the result of Lachenbruch (1968) has been given by Fujikoshi (2000) with explicit form. Fujikoshi and Seo (1998) gave that which is called type II approximation under the framework such that $N_1^{(1)} \to \infty$, $N_1^{(2)} \to \infty$, $p \to \infty$, $n_1 - p \to \infty$ and $N_1^{(2)}/N_1^{(1)} \to$ positive constant k_1. The framework of the approximation proposed in Lachenbruch (1968) can be regarded as both type I and type II approximation.

However the discussions for discriminant functions constructed by missing samples have not been enough. Recently Shutoh, Hyodo and Seo (2009) extended Lachenbruch (1968) to the case of 2-Step monotone missing samples and evaluated the accuracy of the result by Monte Carlo simulations. Shutoh, Hyodo and Seo (2009) concluded that the approximation of Lachenbruch (1968) can be improved by using 2-Step monotone missing samples.

In this paper, since Shutoh, Hyodo and Seo (2009) indicated that approximation by asymptotic expansions have good accuracy for small p in simulation studies, we consider of extending Anderson (1973) to the case of LDF based on 2-Step monotone missing samples by the method similar to Anderson (1973). The settings of sample vectors based on 2-Step monotone missing samples is as follows:

$$\begin{aligned} \boldsymbol{x}_j^{(g)} &\sim N_p(\boldsymbol{\mu}^{(g)}, \Sigma), \quad g = 1, 2,\ j = 1, \ldots, N_1^{(g)}. \\ \boldsymbol{x}_{1j}^{(g)} &\sim N_{p_1}(\boldsymbol{\mu}_1^{(g)}, \Sigma_{11}), \quad g = 1, 2,\ j = N_1^{(g)} + 1, \ldots, N^{(g)}, \end{aligned}$$

where $p = p_1 + p_2$, $\boldsymbol{x}_{\ell j}^{(g)}$, $\boldsymbol{\mu}_\ell^{(g)}$ and $\Sigma_{\ell m}$ are p_ℓ dimensional partitioned vectors of $\boldsymbol{x}_j^{(g)}$, $\boldsymbol{\mu}^{(g)}$ and $p_\ell \times p_m$ partitioned matrix of Σ respectively, that is, considering

$$\boldsymbol{x}_j^{(g)} = \begin{pmatrix} \boldsymbol{x}_{1j}^{(g)} \\ \boldsymbol{x}_{2j}^{(g)} \end{pmatrix}, \quad \boldsymbol{\mu}^{(g)} = \begin{pmatrix} \boldsymbol{\mu}_1^{(g)} \\ \boldsymbol{\mu}_2^{(g)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

If $N_2^{(g)} \equiv N^{(g)} - N_1^{(g)} = 0$, the settings can be reduced to complete data. Also we assume relevant conditions that n_1 is larger than both of p_1 and p_2,

$$\rho_i \equiv \frac{n_i}{n} \to r_i, \quad \frac{N_1^{(2)}}{N_1^{(1)}} \to k_1, \quad \frac{N^{(2)}}{N^{(1)}} \to k$$

as $N_1^{(g)} \to \infty$, $N_2^{(g)} \to \infty$ with nonsingularity of the estimator $\widehat{\Sigma}$ which is shown in Section 2 where $n_i = N_i^{(1)} + N_i^{(2)} - 2$, $n = n_1 + n_2 + 2$ for $g = 1, 2,\ i = 1, 2$. r_i $(i = 1, 2)$, k_1 and k denote positive constants respectively.

The organization of this paper is as follows. We review MLEs based on 2-Step monotone missing samples and their properties in Section 2. In Section 3, we discuss asymptotic distribution of LDF to be needed by deriving asymptotic expansions. In Section 4, we derive asymptotic expansions for the distribution of studentized LDF based on 2-Step monotone missing samples. Also we give some Lemmas and main result as a Theorem in this paper. Section 5 presents evaluations of main result with Anderson (1973) by simulation studies. At the end of this paper, we give conclusion in this paper.

2. Note on MLEs and their properties

As concerns details of derivation of MLEs, see Shutoh, Hyodo and Seo (2009). In this section, we introduce the MLEs with explicit expressions and their properties. 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

$$\begin{aligned}
\overline{\boldsymbol{x}}_F^{(g)} &= \begin{pmatrix} \overline{\boldsymbol{x}}_{1F}^{(g)} \\ \overline{\boldsymbol{x}}_{2F}^{(g)} \end{pmatrix} = \frac{1}{N_1^{(g)}} \sum_{j=1}^{N_1^{(g)}} \boldsymbol{x}_j^{(g)} = \frac{1}{N_1^{(g)}} \sum_{j=1}^{N_1^{(g)}} \begin{pmatrix} \boldsymbol{x}_{1j}^{(g)} \\ \boldsymbol{x}_{2j}^{(g)} \end{pmatrix}, \\
\overline{\boldsymbol{x}}_{1T}^{(g)} &= \frac{1}{N^{(g)}} \sum_{j=1}^{N^{(g)}} \boldsymbol{x}_{1j}^{(g)}, \quad \overline{\boldsymbol{x}}_{1L}^{(g)} = \frac{1}{N_2^{(g)}} \sum_{j=N_1^{(g)}+1}^{N^{(g)}} \boldsymbol{x}_{1j}^{(g)}, \\
S^{(2)} &= \frac{1}{n_2} \sum_{g=1}^{2} \sum_{j=N_1^{(g)}+1}^{N^{(g)}} (\boldsymbol{x}_{1j}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})(\boldsymbol{x}_{1j}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})', \\
\widehat{\Psi}_{11} &= \frac{1}{n}\left[n_1 S_{11} + n_2 S^{(2)} + \sum_{g=1}^{2} \left\{ \frac{N_1^{(g)} N_2^{(g)}}{N^{(g)}} (\overline{\boldsymbol{x}}_{1F}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})(\overline{\boldsymbol{x}}_{1F}^{(g)} - \overline{\boldsymbol{x}}_{1L}^{(g)})' \right\} \right], \\
\widehat{\Psi}_{12} &= S_{11}^{-1} S_{12}, \quad \widehat{\Psi}_{22} = S_{22\cdot 1}
\end{aligned}$$

and $S_{\ell m}$ denotes $p_\ell \times p_m$ partitioned matrix of S which is similar to Σ. These MLEs are one of extensions for Anderson and Olkin (1985). Also random matrices which construct $\widehat{\Sigma}$ have the following distributions:

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

where $W_d(m, \Omega)$ denotes Wishart distribution with the parameters m and Ω.

3. Asymptotic distribution of LDF

In this section, we consider of the conditional distribution of LDF. By making use of MLEs introduced in Section 2, LDF is constructed as

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

and probabilities of misclassification in W_m are

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

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

Define

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

Then the probability (5) can be written by

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

where $u = (c-(1/2)D_m^2)/D_m$. In (7), the conditional distribution of Z_m given $\widehat{\boldsymbol{\mu}}^{(1)}, \widehat{\boldsymbol{\mu}}^{(2)}$ and $\widehat{\Sigma}$ has standard normal distribution. Therefore, we consider

$$\mathrm{E}\left[\Phi\left((uD_m + F_m)\, V_m^{-\frac{1}{2}}\right)|\boldsymbol{x} \in \Pi^{(1)}\right]. \tag{8}$$

In (6), also we consider that

$$\Pr\left[\frac{W_m + \frac{1}{2}D_m^2}{D_m} \leq u'|\boldsymbol{x} \in \Pi^{(2)}\right]. \tag{9}$$

Since it is sufficient to consider only the case of (8) for achieving our purpose, we derive the result under $\boldsymbol{x} \in \Pi^{(1)}$ mainly.

4. Derivation of asymptotic expansion

Now we prepare the following the other expression of random vectors:

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

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

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

for $g = 1, 2,\ \ell = 1, 2$, where $\boldsymbol{\delta}_\ell = \boldsymbol{\mu}_\ell^{(1)} - \boldsymbol{\mu}_\ell^{(2)}$ and $\boldsymbol{\delta}' = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2')$. 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 $\ell = 1, 2$ and $m = 1, 2$. Also we can rewrite MLEs of covariance matrix as

$$\begin{aligned}
\widehat{\Psi}_{11} &= \rho\Sigma_{11} + \frac{1}{\sqrt{n}}\left(\sqrt{\rho_1}T_{11}^{(1)} + \sqrt{\rho_2}T^{(2)}\right) + \frac{1}{n^2}\sum_{g=1}^{2}\boldsymbol{v}^{(g)}\boldsymbol{v}^{(g)\prime}, \\
\widehat{\Psi}_{12} &= \left(\Sigma_{11} + \frac{1}{\sqrt{n_1}}T_{11}^{(1)}\right)^{-1}\left(\Sigma_{12} + \frac{1}{\sqrt{n_1}}T_{12}^{(1)}\right), \\
\widehat{\Psi}_{22} &= \Sigma_{22\cdot 1} + \frac{1}{\sqrt{n_1}}T_{22\cdot 1}^{(1)}, \\
S^{(1)} &= \Sigma + \frac{1}{\sqrt{n_1}}T^{(1)},
\end{aligned}$$

where

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

Note that $\rho \equiv \rho_1 + \rho_2 = 1$ in asymptotic sense. For large m, by using

$$\begin{aligned}
\left(I + \frac{1}{\sqrt{m}}A\right)^{-1} &= I - \frac{1}{\sqrt{m}}A + \frac{1}{m}A^2 - \frac{1}{m\sqrt{m}}A^3 + \frac{1}{m^2}A^4 \\
&\quad - \frac{1}{m^2\sqrt{m}}A^5\left(I + \frac{1}{\sqrt{m}}A\right)^{-1}, \\
\left(I + \frac{1}{\sqrt{m}}A\right)^{-2} &= I - \frac{2}{\sqrt{m}}A + \frac{3}{m}A^2 - \frac{4}{m\sqrt{m}}A^3 + \frac{5}{m^2}A^4 \\
&\quad - \frac{1}{m^2\sqrt{m}}\left(6A^5 + \frac{5}{\sqrt{m}}A^6\right)\left(I + \frac{1}{\sqrt{m}}A\right)^{-2},
\end{aligned}$$

where A is matrix, D_m^2, F_m and V_m are expressed as

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

where

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

R_d, R_f and R_v are homogeneous polynomial of degree 3 in the elements of random vectors and matrices. Therefore

$$\begin{aligned} D_m &= \Delta_1 + \frac{1}{\sqrt{n}}\left(\frac{1}{2\Delta_1} D_{m1}\right) + \frac{1}{n}\left(\frac{1}{2\Delta_1} D_{m2} - \frac{1}{8\Delta_1^3} D_{m1}^2\right) \\ &\quad + \frac{1}{n\sqrt{n}} R_d' + \mathrm{O}_p(n^{-2}), \end{aligned} \tag{11}$$

$$\begin{aligned} V_m^{-\frac{1}{2}} &= \frac{1}{\Delta_2} + \frac{1}{\sqrt{n}}\left(-\frac{1}{2\Delta_2^3} V_{m1}\right) + \frac{1}{n}\left(-\frac{1}{2\Delta_2^3} V_{m2} + \frac{3}{8\Delta_2^5} V_{m1}^2\right) \\ &\quad + \frac{1}{n\sqrt{n}} R_v' + \mathrm{O}_p(n^{-2}), \end{aligned} \tag{12}$$

where R_d' and R_v' have same properties as R_d and R_v respectively. D_{mi}, F_{mi} and V_{mi} $(i = 1, 2)$ will be presented in Appendix A. By (10), (11) and (12), we can obtain

$$(uD_m + F_m) V_m^{-\frac{1}{2}} = u^* + \frac{1}{\sqrt{n}} f_1 + \frac{1}{n} f_2 + \frac{1}{n\sqrt{n}} R_1 + \mathrm{O}_p(n^{-2}),$$

where

$$\begin{aligned} u^* &= u\frac{\Delta_1}{\Delta_2}, \quad f_1 = -\frac{1}{2\Delta_2^2} u^* V_{m1} + \frac{1}{2\Delta_1^2} u^* D_{m1} + \frac{1}{\Delta_2} F_{m1}, \\ f_2 &= -\frac{1}{4\Delta_1^2\Delta_2^2} u^* D_{m1} V_{m1} - \frac{1}{2\Delta_2^3} F_{m1} V_{m1} + \frac{1}{2\Delta_1^2} u^* D_{m2} - \frac{1}{8\Delta_1^4} u^* D_{m1}^2 \\ &\quad + \frac{1}{\Delta_2} F_{m2} - \frac{1}{2\Delta_2^2} u^* V_{m2} + \frac{3}{8\Delta_2^4} u^* V_{m1}^2 \end{aligned}$$

and R_1 denotes homogeneous polynomial of degree 3 in the elements of random vectors and random matrices. By Taylor series expansion of Φ, we can rewrite (8) under $\boldsymbol{x} \in \Pi^{(1)}$ as

$$\Phi(u^*) + \phi(u^*) \left[\frac{1}{\sqrt{n}} \mathrm{E}(f_1) + \frac{1}{n} \left(\mathrm{E}(f_2) - \frac{1}{2} u^* \mathrm{E}(f_1^2) \right) + \frac{1}{n\sqrt{n}} \mathrm{E}(R_2) \right] \qquad (13)$$
$$+\mathrm{O}(n^{-2}),$$

where R_2 denotes the term having same property as R_1 and depends on u. For obtaining asymptotic expansion, we need the results of expectations and show some useful Lemmas:

Lemma 1. *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 and Ω_{11} are nonsingular, then Ω^{-1} is

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

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

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

Lemma 2. *Suppose that $\boldsymbol{X} \sim N_d(\boldsymbol{\eta}, \Omega)$. Then the following expectations can be obtained.*

$$\begin{aligned} \mathrm{E}(\boldsymbol{X}\boldsymbol{c}'\boldsymbol{X}) &= \Omega\boldsymbol{c} + \boldsymbol{\eta}\boldsymbol{c}'\boldsymbol{\eta}, \\ \mathrm{E}(\boldsymbol{X}'C\boldsymbol{X}) &= \mathrm{tr}(C\Omega) + \boldsymbol{\eta}'C\boldsymbol{\eta}, \\ \mathrm{E}(\boldsymbol{X}\boldsymbol{X}') &= \Omega + \boldsymbol{\eta}\boldsymbol{\eta}', \end{aligned}$$

where $\boldsymbol{c}$ and C denotes $d \times 1$ constant vector and $d \times d$ constant matrix respectively.

PROOF. See Nel (1977). □

Lemma 3. *Suppose* $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')' \sim N_d(\boldsymbol{\eta}, \Omega)$, *where* $\boldsymbol{X}_1$ *is* d_1 *partitioned random vector and* $\boldsymbol{X}_2$ *is* d_2 *partitioned random vector. Then the conditional distribution of* $\boldsymbol{X}_2$ *given* $\boldsymbol{X}_1$ *is as follows.*

$$\boldsymbol{X}_2|\boldsymbol{X}_1 \ \sim \ N_{d_2}(\boldsymbol{\eta}_2 + \Omega_{21}\Omega_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\eta}_1), \Omega_{22\cdot 1}),$$

where $\boldsymbol{\eta}_\ell$ *is partitioned vector of* $\boldsymbol{\eta}$ *corresponding to* $\boldsymbol{X}_\ell$ *for* $\ell = 1, 2$, *i.e.,* $d = d_1 + d_2$ *and* Ω *is defined in Lemma 1.*

Proof. Consider the joint p.d.f of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$. □

Lemma 4. *Suppose that* G *has* $W_d(m, \Omega)$. *Then the following expectations can be obtained:*

$$\begin{aligned}
&(i) && \mathrm{E}(G) = m\Omega,\\
&(ii) && \mathrm{E}(GCG) = m^2\Omega C\Omega + m\Omega C'\Omega + m\mathrm{tr}(C\Omega)\Omega,\\
&(iii) && \mathrm{E}(G_{11}C_{11}G_{11}) = m^2\Omega_{11}C_{11}\Omega_{11} + m\Omega_{11}C_{11}'\Omega_{11} + m\mathrm{tr}(C_{11}\Omega_{11})\Omega_{11},\\
&(iv) && \mathrm{E}(G_{12}C_{21}G_{11}) = m^2\Omega_{12}C_{21}\Omega_{11} + m\Omega_{11}C_{21}'\Omega_{21} + m\mathrm{tr}(C_{21}\Omega_{12})\Omega_{11},\\
&(v) && \mathrm{E}(G_{12}C_{22}G_{21}) = m^2\Omega_{12}C_{22}\Omega_{21} + m\Omega_{12}C_{22}'\Omega_{21}\\
& && \qquad + m\mathrm{tr}(C_{22}\Omega_{22\cdot 1})\Omega_{11} + m\mathrm{tr}(C_{22}\Omega_{21}\Omega_{11}^{-1}\Omega_{12})\Omega_{11},\\
&(vi) && \mathrm{E}(G_{22}C_{21}G_{11}) = m^2\Omega_{22\cdot 1}C_{21}\Omega_{11} + m^2\Omega_{21}\Omega_{11}^{-1}\Omega_{12}C_{21}\Omega_{11}\\
& && \qquad + m\Omega_{21}C_{21}'\Omega_{21} + m\mathrm{tr}(C_{21}\Omega_{12})\Omega_{21},\\
&(vii) && \mathrm{E}(GC_{(12)1}G_{11}) = \begin{pmatrix} \mathrm{E}(G_{11}C_{11}G_{11}) + \mathrm{E}(G_{12}C_{21}G_{11}) \\ \mathrm{E}(G_{21}C_{11}G_{11}) + \mathrm{E}(G_{22}C_{21}G_{11}) \end{pmatrix},
\end{aligned}$$

where C *denotes* $d \times d$ *constant matrix,* $C_{\ell m}$ *denotes* $d_\ell \times d_m$ *constant matrix and* $C_{(12)1} = (C_{11}'C_{21}')'$ *respectively.*

Proof. These results will be shown in Appendix B.

These Lemmas provide many expectations to obtain asymptotic expansion. The results to obtain that will be presented in Appendix C. By making use of them, we can obtain main result with replacing $n_1/N_1^{(1)}$, $n/N^{(1)}$, ρ_1 and ρ by substituting their limits $1 + k_1$, $1 + k$, r_1 and 1.

Theorem 5. *The distributions for the studentized discriminant function presented in (7) and (9) can be expanded as*

$$\Phi(u) + \frac{1}{nr_1}\phi(u)\Bigg[\frac{\Delta_{11}^2 + p_2 - 1}{\Delta}(1 + k_1) - \Bigg\{\Bigg(p - \frac{1}{4} + \frac{k_1}{2}\Bigg) + \frac{7}{4}\Delta_{11}^4$$

$$
-\left(p_1+\frac{3}{2}+\frac{k_1}{2}\right)\Delta_{11}^2\Bigg\}u-\frac{1}{4}(1-\Delta_{11}^4)u^3\Bigg]
$$

$$
+\frac{1}{n}\phi(u)\left[\frac{p_1-\Delta_{11}^2}{\Delta}(1+k)-\left\{\left(p_1+\frac{3}{2}+\frac{k}{2}\right)\Delta_{11}^2-\frac{7}{4}\Delta_{11}^4\right\}u\right.
$$

$$
\left.-\frac{1}{4}\Delta_{11}^4u^3\right]+\mathrm{O}(n^{-2})
$$

and

$$
\Phi(u')-\frac{1}{nr_1}\phi(u')\left[\frac{\Delta_{11}^2+p_2-1}{\Delta}\left(1+\frac{1}{k_1}\right)+\left\{\left(p-\frac{1}{4}+\frac{1}{2k_1}\right)\right.\right.
$$

$$
\left.\left.+\frac{7}{4}\Delta_{11}^4-\left(p_1+\frac{3}{2}+\frac{1}{2k_1}\right)\Delta_{11}^2\right\}u'+\frac{1}{4}(1-\Delta_{11}^4)u'^3\right]
$$

$$
-\frac{1}{n}\phi(u')\left[\frac{p_1-\Delta_{11}^2}{\Delta}\left(1+\frac{1}{k}\right)+\left\{\left(p_1+\frac{3}{2}+\frac{1}{2k}\right)\Delta_{11}^2-\frac{7}{4}\Delta_{11}^4\right\}u'\right.
$$

$$
\left.+\frac{1}{4}\Delta_{11}^4u'^3\right]+\mathrm{O}(n^{-2})
$$

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

Proof. By making use of the results in Appendix C which are derived by Lemma 1–Lemma 4, we can find that $\mathrm{E}(V_{m1})=\mathrm{E}(D_{m1})=\mathrm{E}(F_{m1})=0$, i.e., $\mathrm{E}(f_1)=0$. Also we can obtain following results with replacing $n_1/N_1^{(1)}$, $n/N^{(1)}$, ρ_1 and ρ by substituting their limits $1+k_1$, $1+k$, r_1 and 1:

$$
\begin{aligned}
\mathrm{E}(D_{m1}V_{m1}) &= \frac{4}{r_1}\ell(k_1)(\Delta^2-\delta_{11}^2)+4\ell(k)\delta_{11}^2+\frac{4}{r_1}(\Delta^4-\delta_{11}^4)+4\delta_{11}^4,\\
\mathrm{E}(F_{m1}V_{m1}) &= \frac{2}{r_1}(1+k_1)(\Delta^2-\delta_{11}^2)+2(1+k)\delta_{11}^2,\\
\mathrm{E}(D_{m2}) &= \frac{p_2}{r_1}\ell(k_1)+\frac{p+1}{r_1}\Delta^2-\frac{p_1+1}{r_1}\delta_{11}^2+p_1\ell(k)+(p_1+1)\delta_{11}^2,\\
\mathrm{E}(D_{m1}^2) &= \frac{4}{r_1}\ell(k_1)(\Delta^2-\delta_{11}^2)+4\ell(k)\delta_{11}^2+\frac{2}{r_1}(\Delta^4-\delta_{11}^4)+2\delta_{11}^4,\\
\mathrm{E}(F_{m2}) &= \frac{p_2}{r_1}(1+k_1)+p_1(1+k),\\
\mathrm{E}(V_{m2}) &= \frac{p_2}{r_1}\ell(k_1)+\frac{3(p+1)}{r_1}\Delta^2-\frac{3(p_1+1)}{r_1}\delta_{11}^2
\end{aligned}
$$

$$
\begin{aligned}
& +p_1\ell(k)+3(p_1+1)\delta_{11}^2, \\
\mathrm{E}(V_{m1}^2) &= \frac{4}{r_1}\ell(k_1)(\Delta^2-\delta_{11}^2)+4\ell(k)\delta_{11}^2+\frac{8}{r_1}(\Delta^4-\delta_{11}^4)+8\delta_{11}^4,
\end{aligned}
$$

where $\ell(t) = 1 + t + (1/t) + 1$. Thus we can obtain

$$
\begin{aligned}
\mathrm{E}(f_2) &= \frac{1}{\Delta}\left\{\frac{1+k_1}{r_1}(p_2-(1-\Delta_{11}^2))+(1+k)(p_1-\Delta_{11}^2)\right\} \\
&\quad -u\left\{\frac{1}{r_1}\left((p+1)-(p_1+1)\Delta_{11}^2-\frac{7}{4}(1-\Delta_{11}^4)\right)\right. \\
&\quad \left.+\left((p_1+1)\Delta_{11}^2-\frac{7}{4}\Delta_{11}^4\right)\right\}.
\end{aligned}
$$

By noting that $2\mathrm{E}(F_{m1}^2) = \mathrm{E}(D_{m1}F_{m1}) = \mathrm{E}(F_{m1}V_{m1})$ with limits, we find

$$
\mathrm{E}(f_1^2) = \frac{1}{2}u^2\left\{\frac{1}{r_1}(1-\Delta_{11}^4)+\Delta_{11}^4\right\}+\left\{\frac{1+k_1}{r_1}(1-\Delta_{11}^2)+(1+k)\Delta_{11}^2\right\}.
$$

Note that $u^* = u$, $\Delta_1^2 = \Delta_2^2 = \Delta^2$ with limits and expectations of the terms included in R_2 are either 0 or $\mathrm{O}(n^{-2})$. By substituting $\mathrm{E}(f_1)$, $\mathrm{E}(f_2)$ and $\mathrm{E}(f_1^2)$ in (13), we can complete the proof. □

In checking on our result, we present the following corollary at the end of this section. This implies our result is an extension for Anderson (1973).

Corollary 6. *The results obtained in Theorem 5 can be reduced to the forms of (3) and (4) derived in Anderson (1973) by putting $N^{(g)} = N_1^{(g)}$, i.e., $N_2^{(g)} = 0$ respectively.*

This corollary can be found easily by noting that $k = k_1$, $r_1 = 1$ and $n = n_1$ under $N_2^{(g)} = 0$.

5. Simulation studies

In this section, we evaluate an extension for Anderson (1973) derived in Section 4 and compare that with Anderson (1973) by Monte Carlo simulations under selected parameters and cut-off point $c = 0$. Since the result of Anderson (1973) and ours have unknown parameter δ_{11} and Δ, we use estimators of them. That is, in the result of Anderson (1973), we use

$$
\frac{n_1-p-1}{n_1}D^2
$$

as estimator of Δ^2 and we also use

$$\frac{n-p_1-1}{n}d_m^2 \quad \text{and} \quad \frac{n_1-p-1}{n_1}D_m^2$$

in our result as estimators of δ_{11}^2 and Δ^2, where $d_m^2 = (\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)})'\widehat{\Sigma}_{11}^{-1}(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)})$.

At first, by simulations with using the above estimators, we evaluate Anderson (1973) and our result under $\boldsymbol{x} \in \Pi^{(1)}$ when $p = 3$ $(p_1 = 2,\ p_2 = 1)$, $\Delta = 1.05, 3.29,\ M_1 \equiv N_1^{(1)} = N_1^{(2)} = N_2^{(1)} = 10, 15, 20, 40$ and $N_2^{(2)}$ in our result depends on M_1 such as the two cases:

$$\begin{aligned} \text{Case (i)} \quad &: \quad N_2^{(2)} = M_1, \\ \text{Case (ii)} \quad &: \quad N_2^{(2)} = 3M_1. \end{aligned}$$

The results of Tables 1-1 and 1-2 imply that our result has lower error than Anderson (1973) under the case such that $N_2^{(1)}$ has the same order as $N_2^{(2)}$. Then also we see that both of them have better approximation for small Δ than that for large Δ.

Table 1-1 The values of approximations and simulations.

In the case of complete data when $p = 3$ and $\Delta = 1.05$		
M_1	Anderson	Simulation
10	0.3787	0.3470
15	0.3508	0.3312
20	0.3373	0.3234
40	0.3180	0.3114
In the case of missing data when $p = 3$ and $\Delta = 1.05$		
M_1	Shutoh and Seo Case (i)	Simulation Case (i)
10	0.3611	0.3344
15	0.3389	0.3224
20	0.3285	0.3162
40	0.3138	0.3083
M_1	Shutoh and Seo Case (ii)	Simulation Case (ii)
10	0.3625	0.3345
15	0.3397	0.3223
20	0.3293	0.3170
40	0.3141	0.3078

Table 1-2 The values of approximations and simulations.

In the case of complete data when $p = 3$ and $\Delta = 3.29$		
M_1	Anderson	Simulation
10	0.1040	0.06979
15	0.08458	0.06215
20	0.07540	0.05857
40	0.06238	0.05455
In the case of missing data when $p = 3$ and $\Delta = 3.29$		
M_1	Shutoh and Seo Case (i)	Simulation Case (i)
10	0.09170	0.06440
15	0.07620	0.05863
20	0.06912	0.05584
40	0.05921	0.05324
M_1	Shutoh and Seo Case (ii)	Simulation Case (ii)
10	0.08966	0.06321
15	0.07478	0.05764
20	0.06806	0.05604
40	0.05866	0.05236

Similarly to Tables 1-1 and 1-2, we evaluate Anderson (1973) and our result when $p = 3$ $(p_1 = 2,\ p_2 = 1)$, $\Delta = 1.05, 3.29,\ M_2 \equiv N_2^{(1)} = N_2^{(2)} = N_1^{(1)} = 10, 15, 20, 40$ and $N_1^{(2)}$ depends on M_2 such as the following case:

$$\text{Case (iii)} \quad : \quad N_1^{(2)} = 3M_2.$$

The simulation of this case is the result which have done under $N_1^{(1)}$ and $N_1^{(2)}$ are not equal but same order. Also in this case, we can see that our result have better approximation than Anderson (1973) and LDF constructed by 2-Step monotone missing samples has lower probabilities of misclassification than that by complete data.

Table 1-3 The values of approximations and simulations under Case (iii).

In the case of complete data when $p = 3$ and $\Delta = 1.05$		
M_1	Anderson	Simulation
10	0.3709	0.3520
15	0.3466	0.3336
20	0.3347	0.3260
40	0.3170	0.3130
In the case of missing data when $p = 3$ and $\Delta = 1.05$		
M_1	Shutoh and Seo	Simulation
10	0.3511	0.3370
15	0.3337	0.3241
20	0.3251	0.3187
40	0.3124	0.3091

Table 1-4 The values of approximations and simulations under Case (iii).

In the case of complete data when $p = 3$ and $\Delta = 3.29$		
M_1	Anderson	Simulation
10	0.08330	0.06263
15	0.07178	0.05834
20	0.06623	0.05599
40	0.05801	0.05313
In the case of missing data when $p = 3$ and $\Delta = 3.29$		
M_1	Shutoh and Seo	Simulation
10	0.07488	0.05946
15	0.06618	0.05631
20	0.06200	0.05438
40	0.05590	0.05239

The above simulations are under the data set which fits the framework of type I approximation. In four simulations, we can see usefulness of our result with missing samples and discriminating by W_m under some cases.

Although we show usefulness of our result by simulation in some cases, we also show the results of simulations in the cases except for the framework. By these simulations, we make sure the framework of the approximation proposed in this paper.

For instance, in case that $p = 7$ $(p_1 = 4,\ p_2 = 3)$, $p = 9$ $(p_1 = 5,\ p_2 = 4)$, $\Delta = 1.05$ and the settings of sample sizes are Case (i), we can find that high dimensionality of data set makes approximations of this type poorer. See Tables 2-1 and 2-2.

Table 2-1 The values of approximations and simulations.

In the case of complete data when $p = 7$ and $\Delta = 1.05$		
M_1	Anderson	Simulation
10	0.4607	0.3904
15	0.4067	0.3668
20	0.3808	0.3502
40	0.3415	0.3295
In the case of missing data when $p = 7$ and $\Delta = 1.05$		
	Shutoh and Seo	Simulation
M_1	Case (i)	Case (i)
10	0.4336	0.3813
15	0.3873	0.3568
20	0.3655	0.3442
40	0.3331	0.3229

Table 2-2 The values of approximations and simulations.

In the case of complete data when $p = 9$ and $\Delta = 1.05$		
M_1	Anderson	Simulation
10	0.4980	0.4065
15	0.4309	0.3806
20	0.3996	0.3658
40	0.3521	0.3378

In the case of missing data when $p = 9$ and $\Delta = 1.05$		
M_1	Shutoh and Seo Case (i)	Simulation Case (i)
10	0.4654	0.4011
15	0.4088	0.3717
20	0.3822	0.3562
40	0.3421	0.3304

Also we can not obtain good approximations derived by Anderson (1973) and our result when $N_2^{(2)}$(or $N_1^{(2)}$) has higher order than $N_2^{(1)}$(or $N_1^{(1)}$). Now we show the simulated results when $p = 3$ ($p_1 = 2,\ p_2 = 1$), $\Delta = 1.05$, $M_3 \equiv N_1^{(1)} = N_1^{(2)} = N_2^{(1)} = 10$ and $N_2^{(2)} = M_3, M_3^2, M_3^3$.

Table 3-1 The values of approximations and simulations.

In the case of missing data when $p = 3$ and $\Delta = 1.05$		
$N_2^{(2)}$	Shutoh and Seo	Simulation
M_3	0.3611	0.3344
M_3^2	0.3638	0.3349
M_3^3	0.3644	0.3349

Finally, similarly to Table 3-1, we also show the result of Anderson (1973) and our result when $p = 3$ ($p_1 = 2,\ p_2 = 1$), $\Delta = 1.05$, $M_4 \equiv N_1^{(1)} = N_2^{(1)} = N_2^{(2)} = 10$ and $N_1^{(2)} = M_4, M_4^2, M_4^3$.

Table 3-2 The values of approximations and simulations.

In the case of complete data when $p = 3$ and $\Delta = 1.05$		
$N_1^{(2)}$	Anderson	Simulation
M_4	0.3787	0.3470
M_4^2	0.3644	0.3530
M_4^3	0.3610	0.3530

In the case of missing data when $p = 3$ and $\Delta = 1.05$		
$N_1^{(2)}$	Shutoh and Seo	Simulation
M_4	0.3611	0.3344
M_4^2	0.3446	0.3374
M_4^3	0.3412	0.3371

Thus, in case that $N_i^{(1)}$ and $N_i^{(2)}$ does not have same order, Tables 3-1 and 3-2 show that these approximations are poorer.

6. Conclusion and future problems

We proposed asymptotic expansion of the distribution of studentized linear discriminant function constructed by 2-Step monotone missing samples by using perturbation method. As it turns out, we give an extension for the result of Anderson (1973).

By simulation studies, we compared the accuracy in our result and the result of Anderson (1973) under various settings of dimensionality and sample sizes. We showed both cases that our result and the result of Anderson (1973) provide good approximation and do not. We could see that asymptotic expansion are useful for the case p is small, $N_1^{(1)}/N_1^{(2)} = \mathrm{O}(1)$, $N_2^{(1)}/N_2^{(2)} = \mathrm{O}(1)$ and Δ is small. Also we found that our result is useful for obtaining more accurate approximations than that of complete data.

As left problems, for large p, we consider that it is needed for better approximation to be provided. Shutoh, Hyodo and Seo (2009) proposed approximation for this type similar to Lachenbruch (1968). However the approximation based on 2-Step monotone missing samples similar to Fujikoshi and Seo (1998) is needed since Fujikoshi and Seo (1998) has better approximation than Lachenbruch (1968) by simulation studies in Fujikoshi and Seo (1998).

A. Details for the terms included in (10)–(12)

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

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

$$
\begin{aligned}
& -\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} \\
& +\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 \\
& +\frac{\rho_2}{\rho^3}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T^{(2)}\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)} \\
& +\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)} \\
& -\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} \\
& +\frac{2}{\sqrt{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\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\cdot1}^{-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\cdot1}^{-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\cdot1}^{-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\cdot1}^{-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}
\end{aligned}
$$

$$-\frac{3}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{z}_{1F}$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{z}_{1F}$$
$$-\frac{2}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F}$$
$$+\frac{2}{\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{z}_{1F}$$
$$+\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho_1}\left(1-\frac{2}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho\rho_1}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}T_{21}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1$$
$$-\frac{2}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2+\frac{2}{\rho_1}\boldsymbol{z}_{1F}'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}\boldsymbol{z}_{2F}$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{z}_{2F}$$
$$-\frac{2}{\rho_1}\left(2-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$
$$+\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22\cdot 1}^{-1}T_{22\cdot 1}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2$$

$$
\begin{aligned}
&-\frac{2}{\rho_1}\left(1-\frac{1}{\rho}\right)\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot1}^{-1}T_{22\cdot1}^{(1)}\Sigma_{22\cdot1}^{-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\cdot1}^{-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\cdot1}^{-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\cdot1}^{-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\cdot1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\delta}_1\\
&-\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
\end{aligned}
$$

$$+\frac{2}{\rho^2}\sqrt{\frac{\rho_2}{\rho_1}}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T^{(2)}\Sigma_{11}^{-1}T_{12}^{(1)}\Sigma_{22\cdot 1}^{-1}\boldsymbol{\delta}_2.$$

B. Proof of Lemma 4

Since only last four formulae are not well known result, we show the proof of them. By definition of Wishart matrix, we can rewrite the terms of (iv) as

$$\left(\sum_{i=1}^{m}\boldsymbol{z}_{i1}\boldsymbol{z}_{i2}'\right)C_{21}\left(\sum_{j=1}^{m}\boldsymbol{z}_{j1}\boldsymbol{z}_{j1}'\right), \tag{14}$$

where $\boldsymbol{z}_i \equiv (\boldsymbol{z}_{i1}'\boldsymbol{z}_{i2}')' \sim N_d(\boldsymbol{0},\Omega)$ and $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are independent for $i \neq j$. By Lemma 3, $\boldsymbol{z}_{i2}|\boldsymbol{z}_{i1} \sim N_{d_2}(\Omega_{21}\Omega_{11}^{-1}\boldsymbol{z}_{i1},\Omega_{22\cdot 1})$.Thus we find that the conditional expectation of (14) given $\boldsymbol{z}_{i1}$ and $\boldsymbol{z}_{j1}$ can be expressed as

$$\sum_{i}\sum_{j}\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\Omega_{11}^{-1}\Omega_{12}C_{21}\boldsymbol{z}_{j1}\boldsymbol{z}_{j1}'.$$

Since $\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'$ and $\boldsymbol{z}_{j1}\boldsymbol{z}_{j1}'$ are distributed as $W_{d_1}(1,\Omega_{11})$ respectively, by (ii),

$$\mathrm{E}(\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\Omega_{11}^{-1}\Omega_{12}C_{21}\boldsymbol{z}_{j1}\boldsymbol{z}_{j1}') = \Omega_{11}C_{21}'\Omega_{21} + \mathrm{tr}(C_{21}\Omega_{12})\Omega_{11} + \Omega_{12}C_{21}\Omega_{11} \tag{15}$$

for $i = j$. On the other hand, for $i \neq j$, we find that

$$\mathrm{E}(\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\Omega_{11}^{-1}\Omega_{12}C_{21}\boldsymbol{z}_{j1}\boldsymbol{z}_{j1}') = \Omega_{12}C_{21}\Omega_{11} \tag{16}$$

by (i). The left-hand of (iv) includes m terms of (15) and $m(m-1)$ terms of (16). Therefore (iv) holds. Similarly, (v) can be shown easily by making use of the second formula of Lemma 2.

Next we show (vi). By definition of Wishart matrix, also we can rewrite the terms of (vi) as follows:

$$\left(\sum_{i=1}^{m}\boldsymbol{z}_{i2}\boldsymbol{z}_{i2}'\right)C_{21}\left(\sum_{i=1}^{m}\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\right).$$

By Lemma 3, we found the conditional distribution of $\boldsymbol{z}_{i2}$ given $\boldsymbol{z}_{i1}$. Therefore, by Lemma 2, also we find $\mathrm{E}(\boldsymbol{z}_{i2}\boldsymbol{z}_{i2}'|\boldsymbol{z}_{i1}) = \Omega_{22\cdot 1} + \Omega_{21}\Omega_{11}^{-1}\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\Omega_{11}^{-1}\Omega_{12}$. Thus we can obtain the fact that

$$\mathrm{E}\left(\sum_{i=1}^{m}\boldsymbol{z}_{i2}\boldsymbol{z}_{i2}'|\boldsymbol{z}_{11},\ldots,\boldsymbol{z}_{m1}\right) = m\Omega_{22\cdot 1} + \Omega_{21}\Omega_{11}^{-1}\left(\sum_{i=1}^{m}\boldsymbol{z}_{i1}\boldsymbol{z}_{i1}'\right)\Omega_{11}^{-1}\Omega_{12}.$$

This implies

$$\mathrm{E}\left(G_{22}C_{21}G_{11}|G_{11}\right) = m\Omega_{22\cdot 1}C_{21}G_{11} + \Omega_{21}\Omega_{11}^{-1}G_{11}\Omega_{11}^{-1}\Omega_{12}C_{21}G_{11}.$$

Thus, by (i), (ii), this can be shown.

Clearly (vii) holds and we can also find that

$$\mathrm{E}(G_{21}C_{11}G_{11}) = m^2\Omega_{21}C_{11}\Omega_{11} + m\Omega_{21}C_{11}'\Omega_{11} + m\mathrm{tr}(C_{11}\Omega_{11})\Omega_{21}. \qquad (17)$$

Therefore (vii) can be obtained by (iii), (iv), (vi) and (17). Thus all the expectations in this Lemma have been derived. □

C. Formulae for expectations

We can obtain following results which support obtaining asymptotic expansion formulae by given Lemmas:

$$\mathrm{E}(\boldsymbol{z}_F) = \begin{pmatrix} \boldsymbol{z}_{1F} \\ \boldsymbol{z}_{2F} \end{pmatrix} = \boldsymbol{0}, \quad \mathrm{E}(\boldsymbol{z}_{1T}) = \boldsymbol{0},$$

$$\mathrm{E}(\boldsymbol{y}_F^{(g)}) = \begin{pmatrix} \boldsymbol{y}_{1F}^{(g)} \\ \boldsymbol{y}_{2F}^{(g)} \end{pmatrix} = \boldsymbol{0}, \quad \mathrm{E}(\boldsymbol{y}_{1T}^{(g)}) = \boldsymbol{0},$$

$$\mathrm{E}(T^{(1)}) = \begin{pmatrix} T_{11}^{(1)} & T_{12}^{(1)} \\ T_{21}^{(1)} & T_{22}^{(1)} \end{pmatrix} = O, \quad \mathrm{E}(T^{(2)}) = O_{11},$$

$$\begin{aligned} \mathrm{E}(\boldsymbol{z}_F'C\boldsymbol{z}_F) &= \frac{n_1(N_1^{(1)} + N_1^{(2)})}{N_1^{(1)}N_1^{(2)}}\mathrm{tr}(C\Sigma), \\ \mathrm{E}(\boldsymbol{z}_{1F}'C_{11}\boldsymbol{z}_{1F}) &= \frac{n_1(N_1^{(1)} + N_1^{(2)})}{N_1^{(1)}N_1^{(2)}}\mathrm{tr}(C_{11}\Sigma_{11}), \\ \mathrm{E}(\boldsymbol{z}_{1T}'C_{11}\boldsymbol{z}_{1T}) &= \frac{n(N^{(1)} + N^{(2)})}{N^{(1)}N^{(2)}}\mathrm{tr}(C_{11}\Sigma_{11}), \end{aligned}$$

$$\mathrm{E}(\boldsymbol{z}_F\boldsymbol{c}'\boldsymbol{z}_F) = \frac{n_1(N_1^{(1)} + N_1^{(2)})}{N_1^{(1)}N_1^{(2)}}\Sigma\boldsymbol{c},$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{z}_{1F}\boldsymbol{c}_1'\boldsymbol{z}_{1F}) &= \frac{n_1(N_1^{(1)}+N_1^{(2)})}{N_1^{(1)}N_1^{(2)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{z}_{1T}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{n(N^{(1)}+N^{(2)})}{N^{(1)}N^{(2)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{z}_{1F}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{\sqrt{nn_1}(N^{(1)}+N^{(2)})}{N^{(1)}N^{(2)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{z}_{F}\boldsymbol{c}_1'\boldsymbol{z}_{1F}) &= \frac{n_1(N_1^{(1)}+N_1^{(2)})}{N_1^{(1)}N_1^{(2)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{z}_{F}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{\sqrt{nn_1}(N^{(1)}+N^{(2)})}{N^{(1)}N^{(2)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1,
\end{aligned}$$

$$\mathrm{E}(\boldsymbol{z}_F'C\boldsymbol{y}_F^{(1)}) = \frac{n_1}{N_1^{(1)}}\mathrm{tr}(C\Sigma),\quad \mathrm{E}(\boldsymbol{z}_{1F}'C_{11}\boldsymbol{y}_{1F}^{(1)}) = \frac{n_1}{N_1^{(1)}}\mathrm{tr}(C_{11}\Sigma_{11}),$$

$$\mathrm{E}(\boldsymbol{z}_{1T}'C_{11}\boldsymbol{y}_{1T}^{(1)}) = \frac{n}{N^{(1)}}\mathrm{tr}(\Sigma_{11}C_{11}),$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}'\boldsymbol{z}_F) &= \frac{n_1}{N_1^{(1)}}\Sigma\boldsymbol{c},\\
\mathrm{E}(\boldsymbol{y}_{1F}^{(1)}\boldsymbol{c}'\boldsymbol{z}_F) &= \frac{n_1}{N_1^{(1)}}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\end{pmatrix}\boldsymbol{c},\\
\mathrm{E}(\boldsymbol{y}_{1T}^{(1)}\boldsymbol{c}'\boldsymbol{z}_F) &= \frac{\sqrt{nn_1}}{N^{(1)}}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\end{pmatrix}\boldsymbol{c},\\
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1F}) &= \frac{n_1}{N_1^{(1)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{\sqrt{nn_1}}{N^{(1)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{y}_{1F}^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1F}) &= \frac{n_1}{N_1^{(1)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{y}_{1F}^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{\sqrt{nn_1}}{N^{(1)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{y}_{1T}^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1F}) &= \frac{\sqrt{nn_1}}{N^{(1)}}\Sigma_{11}\boldsymbol{c}_1,\\
\mathrm{E}(\boldsymbol{y}_{1T}^{(1)}\boldsymbol{c}_1'\boldsymbol{z}_{1T}) &= \frac{n}{N^{(1)}}\Sigma_{11}\boldsymbol{c}_1,
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}'\boldsymbol{y}_F^{(1)}) &= \frac{n_1}{N_1^{(1)}}\Sigma\boldsymbol{c}, \\
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}_1'\boldsymbol{y}_{1F}^{(1)}) &= \frac{n_1}{N_1^{(1)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1, \\
\mathrm{E}(\boldsymbol{y}_F^{(1)}\boldsymbol{c}_1'\boldsymbol{y}_{1T}^{(1)}) &= \frac{\sqrt{nn_1}}{N^{(1)}}\begin{pmatrix}\Sigma_{11}\\ \Sigma_{21}\end{pmatrix}\boldsymbol{c}_1, \\
\mathrm{E}(\boldsymbol{y}_{1F}^{(1)}\boldsymbol{c}_1'\boldsymbol{y}_{1F}^{(1)}) &= \frac{n_1}{N_1^{(1)}}\Sigma_{11}\boldsymbol{c}_1, \\
\mathrm{E}(\boldsymbol{y}_{1F}^{(1)}\boldsymbol{c}_1'\boldsymbol{y}_{1T}^{(1)}) &= \frac{\sqrt{nn_1}}{N^{(1)}}\Sigma_{11}\boldsymbol{c}_1, \\
\mathrm{E}(\boldsymbol{y}_{1T}^{(1)}\boldsymbol{c}_1'\boldsymbol{y}_{1T}^{(1)}) &= \frac{n}{N^{(1)}}\Sigma_{11}\boldsymbol{c}_1,
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(T^{(1)}CT^{(1)}) &= \Sigma C'\Sigma + \mathrm{tr}(\Sigma C)\Sigma, \\
\mathrm{E}(T_{11}^{(1)}C_{11}T_{11}^{(1)}) &= \mathrm{E}(T^{(2)}C_{11}T^{(2)}) \;=\; \Sigma_{11}C_{11}'\Sigma_{11} + \mathrm{tr}(\Sigma_{11}C_{11})\Sigma_{11}, \\
\mathrm{E}(T_{11}^{(1)}C_{11}T_{12}^{(1)}) &= \Sigma_{11}C_{11}'\Sigma_{12} + \mathrm{tr}(C_{11}\Sigma_{11})\Sigma_{12}, \\
\mathrm{E}(T_{11}^{(1)}C_{12}T_{21}^{(1)}) &= \Sigma_{12}C_{12}'\Sigma_{11} + \mathrm{tr}(C_{12}\Sigma_{21})\Sigma_{11}, \\
\mathrm{E}(T_{12}^{(1)}C_{21}T_{11}^{(1)}) &= \Sigma_{11}C_{21}'\Sigma_{21} + \mathrm{tr}(C_{21}\Sigma_{12})\Sigma_{11}, \\
\mathrm{E}(T_{21}^{(1)}C_{11}T_{11}^{(1)}) &= \Sigma_{21}C_{11}'\Sigma_{11} + \mathrm{tr}(C_{11}\Sigma_{11})\Sigma_{21}, \\
\mathrm{E}(T_{22}^{(1)}C_{21}T_{11}^{(1)}) &= \Sigma_{21}C_{21}'\Sigma_{21} + \mathrm{tr}(\Sigma_{21}C_{21})\Sigma_{21},
\end{aligned}
$$

and

$$
\mathrm{E}(\boldsymbol{\delta}'\Sigma^{-1}T^{(1)}\Sigma^{-1}\boldsymbol{\delta}\boldsymbol{\delta}_1'\Sigma_{11}^{-1}T_{11}^{(1)}\Sigma_{11}^{-1}\boldsymbol{\delta}_1) = 2\delta_{11}^4,
$$

where C is $p \times p$ constant matrix, $\boldsymbol{c}$ is p dimensional constant vector, $C_{\ell m}$ is $p_\ell \times p_m$ constant matrix and $\boldsymbol{c}_\ell$ is p_ℓ dimensional constant vector.

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., 2000. Error bounds for asymptotic approximations of the linear discriminant function when the sample sizes and dimensionality are large. J. Multivariate Anal. 73, 1–17.

Fujikoshi, Y. and Kanazawa, M., 1976. The ML classification statistic in covariate discriminant analysis and its asymptotic expansions. In: Essays in probability and statistics. Shinko Tsusho, Tokyo, pp. 305–320.

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.

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.

Nel, H. M., 1977. On distributions and moments associated with matrix normal distributions. Department of Mathematical Statistics, University of the Orange Free State, Technical Report No.24.

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 Group.