Constrained sample discrimination with the Studentized linear discriminant function based on monotone missing training data

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 an asymptotic expansion for the distribution of the Studentized linear discriminant function with k-step monotone missing training data. It turns out to be a certain generalization of the result derived by Shutoh and Seo (2010). Furthermore, we also derive the cut-off point that controls a conditional probability of misdiscrimination using the result and the idea of McLachlan (1977). Finally, we perform Monte Carlo simulation to evaluate our result.

Keywords: linear discriminant analysis, asymptotic expansion, probabilities of misdiscrimination, monotone missing data.

2010 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

Discriminant analysis is well known as one of the statistical procedures for discriminating p–dimensional sample vector $\boldsymbol{x}$ which arises from one of the considered groups. In this paper, we primarily discuss the linear discrimination for two considered groups $\Pi^{(1)}: N_p(\boldsymbol{\mu}^{(1)}, \Sigma)$ and $\Pi^{(2)}: N_p(\boldsymbol{\mu}^{(2)}, \Sigma)$.

In this case, we usually estimate $\boldsymbol{\mu}^{(g)}$ and Σ using the sample vectors $\boldsymbol{x}_j^{(g)}$ $(j = 1, \ldots, N_1^{(g)},\ g = 1, 2)$ from $\Pi^{(g)}$ and construct the linear discriminant function:

$$W_1 = (\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)})' S^{-1} \left[\boldsymbol{x} - \frac{1}{2} (\overline{\boldsymbol{x}}^{(1)} + \overline{\boldsymbol{x}}^{(2)}) \right],$$

where

$$n_1 = N_1^{(1)} + N_1^{(2)} - 2, \quad \overline{\boldsymbol{x}}^{(g)} = \frac{1}{N_1^{(g)}} \sum_{j=1}^{N_1^{(g)}} \boldsymbol{x}_j^{(g)},$$

$$S = \frac{1}{n_1} \sum_{g=1}^{2} \sum_{j=1}^{N_1^{(g)}} (\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})(\boldsymbol{x}_j^{(g)} - \overline{\boldsymbol{x}}^{(g)})'.$$

On the basis of the cut-off point c, $\boldsymbol{x}$ may be assigned to $\Pi^{(1)}$ if $W_1 > c$.

Several authors have been interested in the following misdiscrimination probabilities:

$$e_1(2|1) = \Pr\left[W_1 \le c \middle| \boldsymbol{x} \in \Pi^{(1)} \right], \quad e_1(1|2) = \Pr\left[W_1 > c \middle| \boldsymbol{x} \in \Pi^{(2)} \right].$$

In order to handle these probabilities, it has been desired to study the distribution of W_1. Under an asymptotic framework $N_1^{(g)} \to \infty$ $(g = 1, 2)$ and $N_1^{(2)}/N_1^{(1)} \to k_1$, the limiting distribution of W_1 is $N((-1)^{g-1}(1/2)\Delta^2, \Delta^2)$ if $\boldsymbol{x}$ arises from $\Pi^{(g)}$ for $g = 1, 2$, where k_1 is a positive const., $\boldsymbol{\delta} = \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$, and $\Delta^2 = \boldsymbol{\delta}' \Sigma^{-1} \boldsymbol{\delta}$.

Okamoto (1963) derived an asymptotic expansion for the distribution of $[W_1 - (-1)^{g-1}(1/2)\Delta^2]/\Delta$. We can consider the asymptotic approximation for the conditional probabilities of misdiscrimination using the result derived by Okamoto (1963). The similar results have been also derived in Memon and Okamoto (1971), Wakaki (1994), and Shutoh (2010). Memon and Okamoto (1971) provided an asymptotic expansion for distribution of the discriminant function based on the maximum likelihood. Wakaki (1994) derived the same and several useful results under elliptical populations. Shutoh (2010) considered the same in the case of k-step monotone missing training data under multivariate normality using the results presented in Kanda and Fujikoshi (1998).

Some authors also considered the relations between the probabilities of misdiscrimination and the cut-off point c, as discussed in Anderson (1973) and McLachlan (1977). Anderson (1973) also derived an asymptotic expansion for the distribution of the Studentized W_1 by perturbation method:

$$\Pr\left[\frac{W_1 - \frac{1}{2}D_1^2}{D_1} \le v \middle| \boldsymbol{x} \in \Pi^{(1)}\right] = \Phi(v) + \frac{\phi(v)}{n_1} b_1(v) + \mathrm{O}(n_1^{-2}), \qquad (1)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution, $\phi(\cdot)$ denotes the probability distribution function of the same,

$$\begin{aligned} D_1^2 &= (\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)})' S^{-1} (\overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}), \\ v &= \frac{c - \frac{1}{2}D_1^2}{D_1}, \quad b_1(v) = \frac{p-1}{\Delta}(1 + k_1) - \left(p - \frac{1}{4} + \frac{1}{2}k_1\right)v - \frac{1}{4}v^3. \end{aligned}$$

Under $\boldsymbol{x} \in \Pi^{(2)}$, the cumulative distribution function of $[W_1 + (1/2)D_1^2]/D_1$ can be also expanded by using the result inverting k_1 in (1). Further, under $\boldsymbol{x} \in \Pi^{(2)}$, McLachlan (1977) derived the cut-off point c which controls the conditional probability of misdiscrimination expressed as $\Phi((vD_1+F_1)V_1^{-\frac{1}{2}})$, where

$$\begin{aligned} F_1 &= (\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)})'S^{-1}(\overline{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)}), \\ V_1 &= (\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)})'S^{-1}\Sigma S^{-1}(\overline{\boldsymbol{x}}^{(1)}-\overline{\boldsymbol{x}}^{(2)}). \end{aligned}$$

In other words, McLachlan (1977) derived the cut-off point c such that

$$\Pr\left[\Phi((vD_1+F_1)V_1^{-\frac{1}{2}}) < M\right] = 1-\alpha+\mathrm{O}(n_1^{-2}),$$

where $1-\alpha$ is the desired level of confidence and M is a specified upper bound. It could be obtained as $c=(1/2)D_1^2+vD_1$, where

$$v = m-\frac{h_1}{\sqrt{n_1}}-\frac{\widehat{h}_2}{n_1}-\frac{\widehat{h}_3}{n_1\sqrt{n_1}}, \tag{2}$$

where m satisfies that $M=\Phi(m)$, z_α is the upper 100α percentage point of the standard normal distribution, $b_i=b_i(m)$ $(i=1,2)$, $\widehat{h}_i$ $(i=2,3)$ denotes h_i with the estimates of Δ,

$$\begin{aligned} b_2(v) &= 1+k_1+\frac{1}{2}v^2, \\ h_1 &= z_\alpha b_2^{\frac{1}{2}}, \\ h_2 &= b_1+\frac{1}{2}z_\alpha^2 m(b_2-1), \\ h_3 &= z_\alpha b_2^{\frac{1}{2}}\Big[mb_1+\frac{m}{8b_2}(z_\alpha^2 m-4b_1)+\frac{z_\alpha^2 b_2}{3}(m^2-1) \\ &\qquad +\frac{3m^2}{4}(1-z_\alpha^2)+\frac{1}{4}z_\alpha^2+p-\frac{1}{4}+\frac{1}{2}k_1\Big]. \end{aligned}$$

Also, the similar result for $\boldsymbol{x}\in\Pi^{(2)}$ can be obtained by inverting k_1 in (2).

This paper primarily provides the results stated in (1) and (2) in the case of W_k, where W_k denotes the linear discriminant function with k-step monotone missing training data. In other words, this paper provides certain generalizations of Anderson's (1973) and McLachlan's (1977) results. For a special case $k = 2$, the result similar to (1) could be obtained in Shutoh and Seo (2010).

Besides, these asymptotic results in this asymptotic framework turn out to be poorer in high dimensional case. Fujikoshi and Seo (1998) proposed the asymptotic approximation for the probabilities of misdiscrimination under another asymptotic framework $N_1^{(g)} \to \infty$, $p \to \infty$, $N_1^{(2)}/N_1^{(1)} \to k_1$, $n_1 - p \to \infty$, and $\Delta^2 = \mathrm{O}(1)$. Matsumoto (2004) obtained an asymptotic expansion for the distributions of the discriminant functions in this asymptotic framework using the results derived by Fujikoshi and Seo (1998).

This paper is organized as follows. Section 2 reviews the statistics that construct the linear discriminant function W_k. Section 3 derives main results in this paper, i.e., an asymptotic expansion for the distribution of the Studentized W_k and a method for determining the cut-off point which controls the conditional probability of misdiscrimination. Section 4 conducts Monte Carlo simulation and compares the proposed procedure with the existing procedure derived by McLachlan (1977). Finally, Section 5 concludes our paper.

2. The conditional probabilities of misdiscrimination in W_k

At first, we express the conditional probabilities of misdiscrimination in W_k to derive main results shown in Section 3. In this paper, we assume

k-step monotone missing data in training data, i.e., the sample vectors

$$\boldsymbol{x}_{(k-t+1)j}^{(g)} = \begin{pmatrix} \boldsymbol{x}_{1j}^{(g)} \\ \vdots \\ \boldsymbol{x}_{k-t+1,j}^{(g)} \end{pmatrix} \sim N_{p_{[k-t+1]}}(\boldsymbol{\mu}_{(k-t+1)}^{(g)}, \Sigma_{(k-t+1)})$$
$$(g = 1, 2,\ t = 1, \ldots, k,\ j = N_{[t-1]}^{(g)} + 1, \ldots, N_{[t]}^{(g)})$$

can be obtained from $\Pi^{(g)}$ $(g = 1, 2)$, where $\boldsymbol{x}_{ij}^{(g)}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, N_{[k-i+1]}^{(g)}$ is p_i–dimensional partitioned sample vector, $\boldsymbol{\mu}_{\alpha}^{(g)}$ is p_α–dimensional partitioned vector of $\boldsymbol{\mu}^{(g)}$, $\Sigma_{\alpha\beta}$ is $p_\alpha \times p_\beta$ partitioned matrix of Σ for $\alpha = 1, \ldots, k,\ \beta = 1, \ldots, k,$

$$\boldsymbol{\mu}_{(k-t+1)}^{(g)} = \begin{pmatrix} \boldsymbol{\mu}_{1}^{(g)} \\ \vdots \\ \boldsymbol{\mu}_{k-t+1}^{(g)} \end{pmatrix}, \Sigma_{(k-t+1)} = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1,k-t+1} \\ \vdots & \ddots & \vdots \\ \Sigma_{k-t+1,1} & \cdots & \Sigma_{k-t+1,k-t+1} \end{pmatrix},$$
$$p_{[k-t+1]} = p_1 + \cdots + p_{k-t+1},\ N_{[0]}^{(g)} \equiv 0,$$
$$N_{[t]}^{(g)} = N_1^{(g)} + \cdots + N_t^{(g)}\ (t = 1, \ldots, k),$$

and it should be noted that $p \equiv p_{[k]}$ and $N^{(g)} \equiv N_{[k]}^{(g)}$.

Then, we can obtain the estimates of $\boldsymbol{\mu}^{(g)}$ and Σ and construct the linear discriminant function W_k:

$$W_k = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}\Big[\boldsymbol{x} - \frac{1}{2}(\widehat{\boldsymbol{\mu}}^{(1)} + \widehat{\boldsymbol{\mu}}^{(2)})\Big],$$

where the estimates of $\boldsymbol{\mu}^{(g)}$ is denoted by $\widehat{\boldsymbol{\mu}}^{(g)}$ and the same of Σ is denoted by $\widehat{\Sigma}$. The discriminant rule is as follows: if $W_k > c$, then $\boldsymbol{x}$ may be assigned to $\Pi^{(1)}$, otherwise it may be assigned to $\Pi^{(2)}$ on the basis of the cut-off point c. Thus, the conditional probabilities of misdiscrimination can be expressed as

$$e_k(2|1) = \Pr\Big[W_k \le c \Big| \boldsymbol{x} \in \Pi^{(1)}\Big] = \Pr\Big[\frac{W_k - \frac{1}{2}D_k^2}{D_k} \le u \Big| \boldsymbol{x} \in \Pi^{(1)}\Big], \quad (3)$$

$$e_k(1|2) = \Pr\left[W_k > c \middle| \boldsymbol{x} \in \Pi^{(2)}\right] = \Pr\left[\frac{W_k + \frac{1}{2}D_k^2}{D_k} > -u^* \middle| \boldsymbol{x} \in \Pi^{(2)}\right], \quad (4)$$

where $u = (c - (1/2)D_k^2)/D_k$, $-u^* = (c + (1/2)D_k^2)/D_k$, and

$$D_k^2 = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}). \quad (5)$$

By (3), as considered in Shutoh and Seo (2010) in the case of $k = 2$, we expand the following distribution function

$$\Phi((uD_k + F_k)V_k^{-\frac{1}{2}}), \quad (6)$$

where

$$F_k = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \boldsymbol{\mu}^{(1)}), \quad (7)$$

$$V_k = (\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}). \quad (8)$$

It follows from

$$Z_k = V_k^{-\frac{1}{2}}(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)})'\widehat{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}^{(1)})$$

is distributed as the standard normal distribution given $\widehat{\boldsymbol{\mu}}^{(1)}$, $\widehat{\boldsymbol{\mu}}^{(2)}$, $\widehat{\Sigma}$ and $\boldsymbol{x} \in \Pi^{(1)}$. We primarily consider the case of (3) since the result for (4) can be also derived using the same for (3). Indeed, (4) can be described as

$$\Phi((u^*D_k^* + F_k^*)\{V_k^*\}^{-\frac{1}{2}}), \quad (9)$$

where

$$F_k^* = (\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(2)} - \boldsymbol{\mu}^{(2)}),$$

$$D_k^* = (\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)})'\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)}) = D_k^2,$$

$$V_k^* = (\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)})'\widehat{\Sigma}^{-1}\Sigma\widehat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}^{(2)} - \widehat{\boldsymbol{\mu}}^{(1)}) = V_k.$$

Using the idea stated in Shutoh (2010), (5), (7), and (8) can be rewritten as

$$D_k^2 = D_1^2 + \sum_{\ell=2}^{k}\Big[\boldsymbol{d}_{(k-\ell+1)}^{[\ell]\prime}\widehat{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{d}_{(k-\ell+1)}^{[\ell]} - \boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]\prime}\widetilde{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]}\Big], \quad (10)$$

$$F_k = F_1 + \sum_{\ell=2}^{k}\Big[\boldsymbol{d}_{(k-\ell+1)}^{[\ell]\prime}\widehat{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{e}_{(k-\ell+1)}^{[\ell]} - \boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]\prime}\widetilde{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{e}_{(k-\ell+1)}^{[\ell-1]}\Big], \quad (11)$$

$$\begin{aligned} V_k = {} & V_1 + \sum_{\ell=2}^{k}\Big[\boldsymbol{d}_{(k-\ell+1)}^{[\ell]\prime}\widehat{\Psi}_{(k-\ell+1)}^{-1}\Sigma_{(k-\ell+1)}\widehat{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{d}_{(k-\ell+1)}^{[\ell]} \qquad (12) \\ & - \boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]\prime}\widetilde{\Psi}_{(k-\ell+1)}^{-1}\Sigma_{(k-\ell+1)}\widetilde{\Psi}_{(k-\ell+1)}^{-1}\boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]}\Big] \\ & +2\sum_{\ell=2}^{k}\sum_{i=k-\ell+2}^{k}\Bigg[\Bigg\{\boldsymbol{d}_{(k-\ell+1)}^{[\ell]\prime}\widehat{\Psi}_{(k-\ell+1)}^{-1} \\ & \times(\Sigma_{(k-\ell+1)i} - \Sigma_{(k-\ell+1)(i-1)}\widehat{\Psi}_{(i-1)i})\widehat{\Psi}_{ii}^{-1}\boldsymbol{\alpha}_{i,k}\Bigg\} - \Bigg\{\boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]\prime}\widetilde{\Psi}_{(k-\ell+1)}^{-1} \\ & \times(\Sigma_{(k-\ell+1)i} - \Sigma_{(k-\ell+1)(i-1)}\widetilde{\Psi}_{(i-1)i})\widetilde{\Psi}_{ii}^{-1}\boldsymbol{\alpha}_{i,k}\Bigg\}\Bigg], \end{aligned}$$

where $n_s = N_s^{(1)}+N_s^{(2)}-2$, $n_{[s]} = \sum_{t=1}^{s} n_t = N_{[s]}-2s$, $N_{[s]} = N_{[s]}^{(1)}+N_{[s]}^{(2)}$, $\Gamma_{\alpha\beta}^{(q)}$ is $p_\alpha \times p_\beta$ partitioned matrix of $\Gamma^{(q)}$ $(q = 1,\ldots,k,\ \ \alpha = 1,\ldots,k-q+1,\ \ \beta = 1,\ldots,k-q+1)$,

$$\begin{aligned} \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[g,s]} &= \frac{1}{N_{[s]}^{(g)}}\sum_{j=1}^{N_{[s]}^{(g)}}\boldsymbol{x}_{(k-\ell+1)j}^{(g)}\ (g=1,2,\ \ell=1,\ldots,k,\ s=1,\ldots,\ell), \\ \overline{\boldsymbol{x}}_{(k-\ell+1)}^{(g,s)} &= \frac{1}{N_s^{(g)}}\sum_{j=N_{[s-1]}^{(g)}+1}^{N_{[s]}^{(g)}}\boldsymbol{x}_{(k-\ell+1)j}^{(g)}\ (g=1,2,\ \ell=1,\ldots,k,\ s=1,\ldots,\ell), \\ \Gamma^{(q)} &= n_q S^{(q)} \\ &= \sum_{g=1}^{2}\sum_{j=N_{[q-1]}^{(g)}+1}^{N_{[q]}^{(g)}}(\boldsymbol{x}_{(k-q+1)j}^{(g)} - \overline{\boldsymbol{x}}_{(k-q+1)}^{(g,q)})(\boldsymbol{x}_{(k-q+1)j}^{(g)} - \overline{\boldsymbol{x}}_{(k-q+1)}^{(g,q)})' \end{aligned}$$

$$
\begin{aligned}
&= \begin{pmatrix} \Gamma_{11}^{(q)} & \cdots & \Gamma_{1,k-q+1}^{(q)} \\ \vdots & \ddots & \vdots \\ \Gamma_{k-q+1,1}^{(q)} & \cdots & \Gamma_{k-q+1,k-q+1}^{(q)} \end{pmatrix} \quad (q = 1, \ldots, k), \\
\Gamma_{\alpha\beta}^{[s]} &= \sum_{q=1}^{s} \Gamma_{\alpha\beta}^{(q)} \ \ (\alpha = 1, \ldots, k-s+1, \ \ \beta = 1, \ldots, k-s+1), \\
\Gamma_{(k-\ell+1)}^{[s]} &= \begin{pmatrix} \Gamma_{11}^{[s]} & \cdots & \Gamma_{1,k-\ell+1}^{[s]} \\ \vdots & \ddots & \vdots \\ \Gamma_{k-\ell+1,1}^{[s]} & \cdots & \Gamma_{k-\ell+1,k-\ell+1}^{[s]} \end{pmatrix} (\ell = 1, \ldots, k, \ s = 1, \ldots, \ell), \\
\Gamma_{(i-1)}^{[k-i+1]} &= \begin{pmatrix} \Gamma_{11}^{[k-i+1]} & \cdots & \Gamma_{1,i-1}^{[k-i+1]} \\ \vdots & \ddots & \vdots \\ \Gamma_{i-1,1}^{[k-i+1]} & \cdots & \Gamma_{i-1,i-1}^{[k-i+1]} \end{pmatrix} \quad (i = k-\ell+2, \ldots, k), \\
\Gamma_{i(i-1)}^{[k-i+1]} &= \begin{pmatrix} \Gamma_{i1}^{[k-i+1]} & \cdots & \Gamma_{i,i-1}^{[k-i+1]} \end{pmatrix} \quad (i = k-\ell+2, \ldots, k), \\
\Gamma_{ii\cdot(1\ldots i-1)}^{[k-i+1]} &= \Gamma_{ii}^{[k-i+1]} - \Gamma_{i(i-1)}^{[k-i+1]} \left\{ \Gamma_{(i-1)}^{[k-i+1]} \right\}^{-1} \Gamma_{(i-1)i}^{[k-i+1]} \ (i = k-\ell+2, \ldots, k),
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{d}_{(k-\ell+1)}^{[\ell]} &= \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[1,\ell]} - \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[2,\ell]}, \quad \boldsymbol{d}_{(k-\ell+1)}^{[\ell-1]} = \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[1,\ell-1]} - \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[2,\ell-1]}, \\
\boldsymbol{d}_{(i-1)}^{[k-i+1]} &= \overline{\boldsymbol{x}}_{(i-1)}^{[1,k-i+1]} - \overline{\boldsymbol{x}}_{(i-1)}^{[2,k-i+1]}, \quad \boldsymbol{d}_{i}^{[k-i+1]} = \overline{\boldsymbol{x}}_{i}^{[1,k-i+1]} - \overline{\boldsymbol{x}}_{i}^{[2,k-i+1]}, \\
\boldsymbol{e}_{(k-\ell+1)}^{[\ell]} &= \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[1,\ell]} - \boldsymbol{\mu}_{(k-\ell+1)}^{(1)}, \quad \boldsymbol{e}_{(k-\ell+1)}^{[\ell-1]} = \overline{\boldsymbol{x}}_{(k-\ell+1)}^{[1,\ell-1]} - \boldsymbol{\mu}_{(k-\ell+1)}^{(1)}, \\
\boldsymbol{\alpha}_{i,k} &= \boldsymbol{d}_{i}^{[k-i+1]} - \widehat{\Psi}_{i(i-1)} \boldsymbol{d}_{(i-1)}^{[k-i+1]}, \\
\widehat{\Psi}_{(k-\ell+1)} &= \frac{1}{n_{[\ell]}} \Gamma_{(k-\ell+1)}^{[\ell]}, \quad \widetilde{\Psi}_{(k-\ell+1)} = \frac{1}{n_{[\ell-1]}} \Gamma_{(k-\ell+1)}^{[\ell-1]}, \\
\widehat{\Psi}_{i(i-1)} &= \widetilde{\Psi}_{i(i-1)} = \Gamma_{i(i-1)}^{[k-i+1]} \left\{ \Gamma_{(i-1)}^{[k-i+1]} \right\}^{-1}, \\
\widehat{\Psi}_{ii} &= \widetilde{\Psi}_{ii} = \frac{1}{n_{[k-i+1]}} \Gamma_{ii\cdot(1\ldots i-1)}^{[k-i+1]},
\end{aligned}
$$

respectively. It should be noted that $S = S^{(1)}$. Then, we have the following lemmas for the distributions.

Lemma 1. *The statistics based on the sample mean vectors have the follow-*

ing distributions:

$$\begin{aligned}
\boldsymbol{d}^{[\ell]}_{(k-\ell+1)} &\sim N_{p_{[k-\ell+1]}}(\boldsymbol{\delta}_{(k-\ell+1)}, \frac{N_{[\ell]}}{N^{(1)}_{[\ell]}N^{(2)}_{[\ell]}}\Sigma_{(k-\ell+1)}),\\
\boldsymbol{d}^{[\ell-1]}_{(k-\ell+1)} &\sim N_{p_{[k-\ell+1]}}(\boldsymbol{\delta}_{(k-\ell+1)}, \frac{N_{[\ell-1]}}{N^{(1)}_{[\ell-1]}N^{(2)}_{[\ell-1]}}\Sigma_{(k-\ell+1)}),\\
\boldsymbol{e}^{[\ell]}_{(k-\ell+1)} &\sim N_{p_{[k-\ell+1]}}(\mathbf{0}, \frac{1}{N^{(1)}_{[\ell]}}\Sigma_{(k-\ell+1)}),\\
\boldsymbol{e}^{[\ell-1]}_{(k-\ell+1)} &\sim N_{p_{[k-\ell+1]}}(\mathbf{0}, \frac{1}{N^{(1)}_{[\ell-1]}}\Sigma_{(k-\ell+1)}),
\end{aligned}$$

where $\boldsymbol{\delta}_{(k-\ell+1)} = \boldsymbol{\mu}^{(1)}_{(k-\ell+1)} - \boldsymbol{\mu}^{(2)}_{(k-\ell+1)}$, *for* $\ell = 1, \ldots, k$.

Lemma 2. *The statistics based on the sample covariance matrices have the following distributions:*

$$\begin{aligned}
\Gamma^{[\ell]}_{(k-\ell+1)} &\sim W_{p_{[k-\ell+1]}}(n_{[\ell]}, \Sigma_{(k-\ell+1)}),\\
\Gamma^{[\ell-1]}_{(k-\ell+1)} &\sim W_{p_{[k-\ell+1]}}(n_{[\ell-1]}, \Sigma_{(k-\ell+1)}),\\
\Gamma^{[k-i+1]}_{(i)} &= \begin{pmatrix} \Gamma^{[k-i+1]}_{(i-1)} & \Gamma^{[k-i+1]}_{(i-1)i} \\ \Gamma^{[k-i+1]}_{i(i-1)} & \Gamma^{[k-i+1]}_{ii} \end{pmatrix} \sim W_{p_{[i]}}(n_{[k-i+1]}, \Sigma_{(i)}),\\
\Gamma^{[k-i+1]}_{ii\cdot(1\ldots i-1)} &\sim W_{p_i}(n_{[k-i+1]} - p_{[i-1]}, \Sigma_{ii\cdot(1\ldots i-1)}),
\end{aligned}$$

where $\Gamma^{[k-i+1]}_{(i-1)i}$ *is the transposed matrix of* $\Gamma^{[k-i+1]}_{i(i-1)}$,

$$\begin{aligned}
\Sigma_{ii\cdot(1\ldots i-1)} &= \Sigma_{ii} - \Sigma_{i(i-1)}\left\{\Sigma_{(i-1)}\right\}^{-1}\Sigma_{(i-1)i},\\
\Sigma_{i(i-1)} &= \Sigma'_{(i-1)i} = \begin{pmatrix} \Sigma_{i1} & \cdots & \Sigma_{i,i-1} \end{pmatrix},
\end{aligned}$$

for $\ell = 2, \ldots, k$ *and* $i = k - \ell + 2, \ldots, k$.

Lemma 3. $\Gamma^{[k-i+1]}_{ii\cdot(1\ldots i-1)}$ *is independent of* $(\Gamma^{[k-i+1]}_{(i-1)}, \Gamma^{[k-i+1]}_{i(i-1)})$.

3. Main results

3.1. Asymptotic expansion for the distribution of Studentized W_k

We prepare the following expressions of the random vectors and the random matrices for the perturbation method:

$$
\begin{aligned}
\boldsymbol{d}^{[\ell]}_{(k-\ell+1)} &= \boldsymbol{\delta}_{(k-\ell+1)} + \frac{1}{\sqrt{n\rho_{[\ell]}}}\boldsymbol{z}^{[\ell]}_{(k-\ell+1)},\\
\boldsymbol{d}^{[\ell-1]}_{(k-\ell+1)} &= \boldsymbol{\delta}_{(k-\ell+1)} + \frac{1}{\sqrt{n\rho_{[\ell-1]}}}\boldsymbol{z}^{[\ell-1]}_{(k-\ell+1)},\\
\boldsymbol{d}^{[k-i+1]}_{(i-1)} &= \boldsymbol{\delta}_{(i-1)} + \frac{1}{\sqrt{n\rho_{[k-i+1]}}}\boldsymbol{z}^{[k-i+1]}_{(i-1)},\ \ \boldsymbol{\delta}_{(i-1)} \ = \ \boldsymbol{\mu}^{(1)}_{(i-1)} - \boldsymbol{\mu}^{(2)}_{(i-1)},\\
\boldsymbol{d}^{[k-i+1]}_{i} &= \boldsymbol{\delta}_{i} + \frac{1}{\sqrt{n\rho_{[k-i+1]}}}\boldsymbol{z}^{[k-i+1]}_{i},\ \ \boldsymbol{\delta}_{i} \ = \ \boldsymbol{\mu}^{(1)}_{i} - \boldsymbol{\mu}^{(2)}_{i},\\
\boldsymbol{e}^{[\ell]}_{(k-\ell+1)} &= \frac{1}{\sqrt{n\rho_{[\ell]}}}\boldsymbol{y}^{[\ell]}_{(k-\ell+1)},\quad \boldsymbol{e}^{[\ell-1]}_{(k-\ell+1)} \ = \ \frac{1}{\sqrt{n\rho_{[\ell-1]}}}\boldsymbol{y}^{[\ell-1]}_{(k-\ell+1)},\\
S &= \Sigma + \frac{1}{\sqrt{n\rho_1}}T^{(1)},\quad S^{(q)}_{(k-\ell+1)} \ = \ \Sigma_{(k-\ell+1)} + \frac{1}{\sqrt{n\rho_q}}T^{(q)}_{(k-\ell+1)},\\
S_{ii\cdot(1\ldots i-1)} &= S_{ii} - S_{i(i-1)}S^{-1}_{(i-1)}S_{(i-1)i} \ = \ \Sigma_{ii\cdot(1\ldots i-1)} + \frac{1}{\sqrt{n\rho_1}}U^{(1)}_{ii\cdot(1\ldots i-1)},\\
S^{(q)}_{ii\cdot(1\ldots i-1)} &= S^{(q)}_{ii} - S^{(q)}_{i(i-1)}\left\{S^{(q)}_{(i-1)}\right\}^{-1}S^{(q)}_{(i-1)i}\\
&= \Sigma_{ii\cdot(1\ldots i-1)} + \frac{1}{\sqrt{n\rho_q}}U^{(q)}_{ii\cdot(1\ldots i-1)},\\
\rho_{[\ell]} &= \rho_1 + \cdots + \rho_\ell,\quad \rho_i \ = \ \frac{n_i}{n},\quad n \ = \ n_{[k]}.
\end{aligned}
$$

for $\ell = 2, \ldots, k$, $q = 2, \ldots, k$, and $i = k - \ell + 2, \ldots, k$. It should be noted that Lemma 3 holds.

Thus, using the above expressions, we can also write

$$
\widehat{\Psi}_{(k-\ell+1)} \ = \ \Sigma_{(k-\ell+1)} + \frac{1}{\sqrt{n}\rho_{[\ell]}}\sum_{q=1}^{\ell}\sqrt{\rho_q}T^{(q)}_{(k-\ell+1)},
$$

$$
\begin{aligned}
\widetilde{\Psi}_{(k-\ell+1)} &= \Sigma_{(k-\ell+1)} + \frac{1}{\sqrt{n}\rho_{[\ell-1]}}\sum_{q=1}^{\ell-1}\sqrt{\rho_q}T^{(q)}_{(k-\ell+1)}, \\
\widehat{\Psi}_{i(i-1)} = \widetilde{\Psi}_{i(i-1)} &= \left[\Sigma_{i(i-1)} + \frac{1}{\sqrt{n}\rho_{[k-i+1]}}\sum_{q=1}^{k-i+1}\sqrt{\rho_q}T^{(q)}_{i(i-1)}\right] \\
&\quad \times \left[\Sigma_{(i-1)} + \frac{1}{\sqrt{n}\rho_{[k-i+1]}}\sum_{q=1}^{k-i+1}\sqrt{\rho_q}T^{(q)}_{(i-1)}\right]^{-1}, \\
\widehat{\Psi}_{ii} = \widetilde{\Psi}_{ii} &= \Sigma_{ii\cdot(1\ldots i-1)} + \frac{1}{\sqrt{n}\rho_{[k-i+1]}}\sum_{q=1}^{k-i+1}\sqrt{\rho_q}U^{(q)}_{ii\cdot(1\ldots i-1)}.
\end{aligned}
$$

Furthermore, for (10), (11), and (12), we can determine the terms of

$$
\begin{aligned}
D_k^2 &= \Delta^2 + \frac{1}{\sqrt{n}}D_{k1} + \frac{1}{n}D_{k2} + \mathrm{o}_p(n^{-1}), \\
F_k &= \frac{1}{\sqrt{n}}F_{k1} + \frac{1}{n}F_{k2} + \mathrm{o}_p(n^{-1}), \\
V_k &= \Delta^2 + \frac{1}{\sqrt{n}}V_{k1} + \frac{1}{n}V_{k2} + \mathrm{o}_p(n^{-1}),
\end{aligned}
$$

using

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

where A is a matrix. Therefore, it holds that

$$
\Phi((uD_k + F_k)V_k^{-\frac{1}{2}}) = \Phi(u) + \phi(u)\left[\frac{1}{\sqrt{n}}w_{k1} + \frac{1}{n}w_{k2}\right] + \mathrm{o}_p(n^{-1}), \quad (13)
$$

where

$$
\begin{aligned}
w_{k1} &= -\frac{1}{2\Delta^2}uV_{k1} + \frac{1}{2\Delta^2}uD_{k1} + \frac{1}{\Delta}F_{k1}, \\
w_{k2} &= \frac{1}{4\Delta^4}u(u^2-1)D_{k1}V_{k1} + \frac{1}{2\Delta^3}(u^2-1)F_{k1}V_{k1} - \frac{1}{8\Delta^4}u(u^2+1)D_{k1}^2
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{8\Delta^4}u(u^2-3)V_{k1}^2-\frac{1}{2\Delta^3}u^2D_{k1}F_{k1}-\frac{1}{2\Delta^2}uF_{k1}^2+\frac{1}{2\Delta^2}uD_{k2}\\
&+\frac{1}{\Delta}F_{k2}-\frac{1}{2\Delta^2}uV_{k2}.
\end{aligned}
$$

For the outline of the derivation, refer Shutoh and Seo (2010).

Although we require too much calculation for our purpose, we use the following results in order to reduce the calculation.

Lemma 4. *For $i=k-\ell+2,\ldots,k$, it holds that*

$$
\Sigma_{(k-\ell+1)i} = \Sigma_{(k-\ell+1)(i-1)}\Sigma_{(i-1)}^{-1}\Sigma_{(i-1)i},
$$

where

$$
\begin{aligned}
\Sigma_{(k-\ell+1)i} &= \begin{pmatrix} \Sigma_{1i} \\ \vdots \\ \Sigma_{k-\ell+1,i} \end{pmatrix},\\
\Sigma_{(k-\ell+1)(i-1)} &= \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1,i-1} \\ \vdots & \ddots & \vdots \\ \Sigma_{k-\ell+1,1} & \cdots & \Sigma_{k-\ell+1,i-1} \end{pmatrix},\\
\Sigma_{(i-1)} &= \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1,i-1} \\ \vdots & \ddots & \vdots \\ \Sigma_{i-1,1} & \cdots & \Sigma_{i-1,i-1} \end{pmatrix},\quad \Sigma_{(i-1)i} = \begin{pmatrix} \Sigma_{1i} \\ \vdots \\ \Sigma_{i-1,i} \end{pmatrix}.
\end{aligned}
$$

Proof. For $i=k-\ell+2$, it clearly holds. For $i=k-\ell+3,\ldots,k$, consider the inverse matrix of $\Sigma_{(i-1)}$ partitioned by $\Sigma_{(k-\ell+1)}$ and the other three blocks. □

Lemma 5. *For $\ell=2,\ldots,k$, $i=k-\ell+2,\ldots,k$, and $q=1,\ldots,k-i+1$, it holds that*

$$
\mathrm{E}(T_{(i-1)i}^{(q)}|T_{(i-1)}^{(q)}) = T_{(i-1)}^{(q)}\Sigma_{(i-1)}^{-1}\Sigma_{(i-1)i}.
$$

Proof. It should be noted that

$$\Gamma^{(q)}_{(i-1)i}|\Gamma^{(q)}_{(i-1)} \sim N_{p_{[i-1]}\times p_i}(\Gamma^{(q)}_{(i-1)}\Sigma_{(i-1)}\Sigma_{(i-1)i},\ \Sigma^{-1}_{ii\cdot(1\ldots i-1)}\otimes\Gamma^{(q)}_{(i-1)}).$$

□

Thus, it follows from (13) and the lemmas shown in Section 2 that the following theorem holds. It should be noted that the expectation of the sum of the terms with $\mathrm{O}_p(n^{-\frac{3}{2}})$ turns out to be $\mathrm{O}(n^{-2})$.

Theorem 6. *The cumulative distribution function of the Studentized linear discriminant function, i.e.,* $[W_k-(1/2)D_k^2]/D_k$ *under* $\boldsymbol{x}\in\Pi^{(1)}$ *stated in (3) is expanded as*

$$\Phi(u)+\frac{\phi(u)}{n}b_{k1}(u)+\mathrm{O}(n^{-2}), \tag{14}$$

where

$$\begin{aligned}
b_{k1}(u) &= c_0+c_1u+c_3u^3,\\
c_0 &= \frac{p-1}{r_1\Delta}(1+k_1)+\sum_{\ell=2}^{k}\frac{p_{[k-\ell+1]}-\Delta^2_{k-\ell+1}}{\Delta}\Big(\frac{1+k_{[\ell]}}{r_{[\ell]}}-\frac{1+k_{[\ell-1]}}{r_{[\ell-1]}}\Big),\\
c_1 &= -\frac{1}{r_1}\Big(p-\frac{1}{4}+\frac{1}{2}k_1\Big)\\
&\quad -\sum_{\ell=2}^{k}\Delta^2_{k-\ell+1}\Big\{\frac{1}{r_{[\ell]}}\Big(p_{[k-\ell+1]}+\frac{3}{2}+\frac{1}{2}k_{[\ell]}-\frac{7}{4}\Delta^2_{k-\ell+1}\Big)\\
&\qquad -\frac{1}{r_{[\ell-1]}}\Big(p_{[k-\ell+1]}+\frac{3}{2}+\frac{1}{2}k_{[\ell-1]}-\frac{7}{4}\Delta^2_{k-\ell+1}\Big)\Big\},\\
c_3 &= -\frac{1}{4r_1}-\sum_{\ell=2}^{k}\frac{\Delta^4_{k-\ell+1}}{4}\Big(\frac{1}{r_{[\ell]}}-\frac{1}{r_{[\ell-1]}}\Big),\\
\delta^2_{k-\ell+1} &= \boldsymbol{\delta}'_{(k-\ell+1)}\Sigma^{-1}_{(k-\ell+1)}\boldsymbol{\delta}_{(k-\ell+1)},\quad \Delta_{k-\ell+1} = \delta_{k-\ell+1}/\Delta,
\end{aligned}$$

r_1 *denotes the limit of* n_1/n, $r_{[\ell]}$ *denotes the limit of* $n_{[\ell]}/n$, $k_1=k_{[1]}$, *and* $k_{[\ell]}$ *denotes the limit of* $N^{(2)}_{[\ell]}/N^{(1)}_{[\ell]}$, *for* $\ell=2,\ldots,k$.

Proof. It follows from the result derived by the lemmas that $\mathrm{E}(w_{k1}) = 0$ and $\mathrm{E}(w_{k2}) = b_{k1}(u)$ hold. □

Corollary 7. *The cumulative distribution function of the Studentized linear discriminant function* $[W_k + (1/2)D_k^2]/D_k$ *under* $\boldsymbol{x} \in \Pi^{(2)}$ *is also expanded as*

$$\Phi(u') - \frac{\phi(u')}{n} b'_{k1}(-u') + \mathrm{O}(n^{-2}),$$

where $b'_{k1}(u)$ *can be obtained by inverting* $k_1 = k_{[1]}$ *and* $k_{[\ell]}$ *for* $\ell = 2, \ldots, k$ *in* $b_{k1}(u)$ *and* $u' = [c + (1/2)D_k^2]/D_k$*.*

Proof. It follows from (9) that this corollary holds. It should be noted that the cumulative distribution function can be expressed as

$$1 - \Phi((u^* D_k^* + F_k^*)\{V_k^*\}^{-\frac{1}{2}})$$

and $u^* = -u'$ holds. □

Corollary 8. *For* $k = 2$*, Theorem 6 and Corollary 7 coincide with the result derived by Shutoh and Seo (2010).*

3.2. Constrained discriminant rule with Studentized W_k

The another main result is derived in this subsection. For this purpose, we are interested in the conditional distribution of the probabilities of misdiscrimination given $\boldsymbol{x} \in \Pi^{(1)}$. Therefore, we consider the characteristic function of (6):

$$c(t) \quad = \quad \mathrm{E}\left[\exp\left\{\mathrm{i}t\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\right\}\right],$$

where $\mathrm{i} = \sqrt{-1}$. Furthermore, it holds that

$$\begin{aligned}
c(t) &= \exp\Big[\mathrm{i}t\mathrm{E}\Big\{\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big\}\Big] \\
&\quad \times \Big[1 + \frac{(\mathrm{i}t)^2}{2}\mathrm{Var}\Big\{\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big\} + \mathrm{o}(t^2)\Big] \\
&= \exp\Big[\mathrm{i}t\mathrm{E}\Big\{\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big\} \\
&\qquad - \frac{t^2}{2}\mathrm{Var}\Big\{\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big\}\Big](1 + R),
\end{aligned}$$

where R is the remainder terms starting with $(\mathrm{i}t)^3$. Using the result stated in (13), the required results have the following forms:

$$\begin{aligned}
\mathrm{E}\Big[\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big] &= \Phi(u) + \phi(u)\Big[\frac{1}{\sqrt{n}}\mathrm{E}(w_{k1}) + \frac{1}{n}\mathrm{E}(w_{k2})\Big] \\
&\quad + \mathrm{O}(n^{-2}), \\
\mathrm{Var}\Big[\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})\Big] &= \frac{\{\phi(u)\}^2}{n}\mathrm{Var}(w_{k1}) + \mathrm{O}(n^{-2}),
\end{aligned}$$

and they have been obtained via the derivation of the result shown in Theorem 6. It should be noted that the expectation of the sum of the terms with $\mathrm{O}_p(n^{-\frac{3}{2}})$ turns out to be $\mathrm{O}(n^{-2})$. Thus, we can obtain the following lemma by considering up to the terms of the first order with respect to n^{-1}.

Lemma 9. *$\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})$ given $\boldsymbol{x} \in \Pi^{(1)}$ is asymptotically distributed as $N(\xi_k, \sigma_k^2)$, where*

$$\begin{aligned}
\xi_k &= \Phi(u) + \frac{\phi(u)}{n}b_{k1}(u), \quad \sigma_k^2 = \frac{\{\phi(u)\}^2}{n}b_{k2}(u), \\
b_{k1}(u) &= c_0 + c_1 u + c_3 u^3, \quad b_{k2}(u) = d_0 + d_2 u^2, \\
d_0 &= \frac{1 + k_1}{r_1} + \sum_{\ell=2}^{k}\Delta_{k-\ell+1}^2\Big\{\frac{1 + k_{[\ell]}}{r_{[\ell]}} - \frac{1 + k_{[\ell-1]}}{r_{[\ell-1]}}\Big\}, \\
d_2 &= \frac{1}{2}\Big\{\frac{1}{r_1} + \sum_{\ell=2}^{k}\Delta_{k-\ell+1}^4\Big(\frac{1}{r_{[\ell]}} - \frac{1}{r_{[\ell-1]}}\Big)\Big\}.
\end{aligned}$$

Corollary 10. $\Phi((u^* D_k^* + F_k^*)\{V_k^*\}^{-\frac{1}{2}})$ *given* $\boldsymbol{x} \in \Pi^{(2)}$ *is asymptotically distributed as the normal distribution whose mean and variance can be obtained inverting* $k_1 = k_{[1]}$ *and* $k_{[\ell]}$ *for* $\ell = 2, \ldots, k$ *in* ξ_k *and* σ_k^2*, respectively.*

Proof. It clearly holds by (9) and Lemma 9. □

In this subsection, we consider the cut-off point such that

$$\Pr\Big[\Phi((uD_k + F_k)V_k^{-\frac{1}{2}}) < M\Big] \;=\; 1 - \alpha.$$

However, $\Phi((uD_k + F_k)V_k^{-\frac{1}{2}})$ is asymptotically distributed as $N(\xi_k, \sigma_k^2)$, as shown in Lemma 9. Therefore, we determine $u = (c - (1/2)D_k^2)/D_k$ that satisfies the following equation

$$\Pr\Big[\Phi((uD_k + F_k)V_k^{-\frac{1}{2}}) < M\Big] \;=\; 1 - \alpha + \mathrm{O}(n^{-2}). \tag{15}$$

It follows from Lemma 9 that the following theorem holds.

Theorem 11. *The cut-off point* $c = (1/2)D_k^2 + uD_k$ *which satisfies (15) can be obtained by*

$$u \;=\; m - \frac{\widehat{h}_{k1}}{\sqrt{n}} - \frac{\widehat{h}_{k2}}{n} - \frac{\widehat{h}_{k3}}{n\sqrt{n}},$$

where

$$\begin{aligned}
h_{k1} &= z_\alpha b_{k2}^{\frac{1}{2}}, \\
h_{k2} &= b_{k1} + \frac{1}{2} z_\alpha^2 m (b_{k2} - 2d_2), \\
h_{k3} &= z_\alpha b_{k2}^{\frac{1}{2}} \Big[m b_{k1} + \frac{m d_2}{2 b_{k2}} (z_\alpha^2 m d_2 - 2 b_{k1}) + \frac{z_\alpha^2 b_{k2}}{3} (m^2 - 1) \\
&\qquad - \frac{3m^2}{2} (z_\alpha^2 d_2 + 2c_3) + \frac{z_\alpha^2 d_2}{2} - c_1 \Big],
\end{aligned}$$

$b_{k1} = b_{k1}(m)$*,* $b_{k2} = b_{k2}(m)$*, and* $\widehat{h}_{ki}$*'s* $(i = 1, 2, 3)$ *are* h_{ki}*'s with the estimates of* $\delta_{k-\ell+1}$ *and* Δ*.*

Proof. Using Lemma 9, we express (15) as follows:

$$\Phi\Big(\frac{M-\xi_k}{\sigma_k}\Big) \quad = \quad 1-\alpha+\mathrm{O}(n^{-2}).$$

Thus, it is sufficient to derive the cut-off point satisfies

$$M=\xi_k+z_\alpha\sigma_k \tag{16}$$

in order to achieve our purpose. To determine the point u, we put the solution of (16) as

$$u \quad = \quad m-\frac{h_{k1}}{\sqrt{n}}-\frac{h_{k2}}{n}-\frac{h_{k3}}{n\sqrt{n}},$$

where h_i's ($i=1,2,3$) are the unknown finite constants. Using

$$\begin{aligned}
\Phi(u) &= M+\frac{\phi(m)}{\sqrt{n}}\Big\{-h_{k1}\Big\}+\frac{\phi(m)}{n}\Big\{-h_{k2}-\frac{1}{2}mh_{k1}^2\Big\}\\
&\quad +\frac{\phi(m)}{n\sqrt{n}}\Big\{-h_{k3}-mh_{k1}h_{k2}-\frac{1}{6}(m^2-1)h_{k1}^3\Big\}+\mathrm{o}(n^{-\frac{3}{2}}),\\
\phi(u) &= \phi(m)+\frac{\phi(m)}{\sqrt{n}}\Big\{mh_{k1}\Big\}\\
&\quad +\frac{\phi(m)}{n}\Big\{mh_{k2}+\frac{1}{2}(m^2-1)h_{k1}^2\Big\}+\mathrm{o}(n^{-1}),\\
b_{k1}(u) &= b_{k1}+\frac{1}{\sqrt{n}}\Big\{-(c_1+3c_3m^2)h_{k1}\Big\}+\mathrm{o}(n^{-\frac{1}{2}}),\\
\{b_{k2}(u)\}^{\frac{1}{2}} &= b_{k2}^{\frac{1}{2}}+\frac{1}{\sqrt{n}}\Big\{-md_2b_{k2}^{-\frac{1}{2}}h_{k1}\Big\}\\
&\quad +\frac{1}{n}\Big\{-md_2b_{k2}^{-\frac{1}{2}}h_{k2}+\frac{1}{2}d_2(1-m^2d_2b_{k2}^{-1})b_{k2}^{-\frac{1}{2}}h_{k1}^2\Big\}+\mathrm{o}(n^{-1}),
\end{aligned}$$

we equate the terms of order $n^{-\frac{1}{2}}$, n^{-1}, and $n^{-\frac{3}{2}}$ in (16) and determine the unknown h_{ki}'s ($i=1,2,3$), which proves Theorem 11. It should be noted that determined h_{ki}'s depend on the unknown parameters for $i=1,2,3$. □

Corollary 12. *The cut-off point* $c = -(1/2)D_k^2 - u^* D_k$ *which satisfies*

$$\Pr\Big[\Phi((u^* D_k^* + F_k^*)\{V_k^*\}^{-\frac{1}{2}}) < M\Big] \;=\; 1 - \alpha + \mathrm{O}(n^{-2})$$

can be obtained by

$$u^* = m - \frac{\widehat{h}_{k1}^*}{\sqrt{n}} - \frac{\widehat{h}_{k2}^*}{n} - \frac{\widehat{h}_{k3}^*}{n\sqrt{n}},$$

where h_{ki}^**'s are defined as the constants inverting* $k_1 = k_{[1]}$ *and* $k_{[\ell]}$ $(\ell = 2, \ldots, k)$ *in* h_{ki}*'s.* $\widehat{h}_{ki}^*$*'s are* h_{ki}^**'s with the estimates of* $\delta_{k-\ell+1}$ *and* Δ*, for* $i = 1, 2, 3$*.*

4. Simulation studies

In this section, we perform Monte Carlo simulation in order to evaluate our result derived in Theorem 11. Especially, we compare our result with McLachlan's (1977) result. We select the eight cases, as listed in Table 1. For each case, we set $k = 3$ for our result. The sample size is set as $N_{(3)} \equiv N_1 = N_2 = N_3 = 15, 20, 40, 100$, where $N_\ell \equiv N_\ell^{(1)} = N_\ell^{(2)}$ $(\ell = 1, 2, 3)$. For the result derived by McLachlan (1977), the sample size is set as $N_1 = 15, 20, 40, 100$.

In Tables 2–3, it can be observed that the larger sample sizes result in large values of $e_k(2|1)$ for $k = 1, 3$. We also can observe that $e_k(1|2)$ has the opposite asymptotic behavior. Further, for the most cases, we can observed that the values of $e_k(1|2)$ are lower when the same of $e_k(2|1)$ are larger. In this section, we primarily evaluate $e_k(2|1)$ for $k = 1, 3$ since we have derived the cut-off point which focuses on $e_k(2|1)$.

First, we compare the results of Cases 1–3. The desired level of confidence $1 - \alpha$ for Case 2 is set as 0.90 and the same for Case 3 is set as 0.99, respectively. It can be also observed that the larger desired levels $1-\alpha$ result in lower $e_k(2|1)$ and larger $e_k(1|2)$.

Next, we compare the results of Cases 1, 4, and 5. In Case 4, the Mahalanobis distance $\Delta = \delta_1$ is set as 0.50. In Case 5, the same is set as 1.36. When the parameters are known, the exact probabilities of misdiscrimination $\Phi(-(1/2)\Delta)$ is nearly equal to 0.25. It can be observed that the exact probabilities $\Phi(-(1/2)\Delta)$ closer to M result in lower $e_k(2|1)$ and $e_k(1|2)$ for $k = 1$, i.e., the result derived by McLachlan (1977). However, $e_k(2|1)$ for $k = 3$, i.e., our result does not have this property.

By the simulation results conducted in Shutoh (2011), it may depend on the value of $\delta_{k-\ell+1}$ for $\ell = 2, \ldots, k$. So, we also conduct the additional simulation listed in Table 5. It should be noted that Case 5(a) denotes the result for Case 5 in Tables 1–4. The results in Tables 6–7 indicate that lower Δ_1 and lower Δ_2 result in lower $e_k(2|1)$ and larger $e_k(1|2)$ for $k = 3$. On the other hand, for $k = 1$, we cannot observed that $e_k(2|1)$ depends on Δ_1 and Δ_2. $e_k(1|2)$ for $k = 1$ has the property similar to $e_k(2|1)$ for $k = 1$.

Furthermore, we compare the results for Cases 1, 6, and 7 listed in Tables 2–3. The specified upper bound for Case 6 is set as 0.05 and the same for Case 7 is set as 0.25. As a matter of course, it can be observed the larger M results in larger $e_k(2|1)$ and lower $e_k(1|2)$. The result for Case 8 indicates that our and McLachlan's (1977) results are poorer procedures for large dimensionality.

Finally, we compare the values of $1-\alpha$ listed in Table 4 and Table 8. In Cases 1, 3, 4, 5(b), 5(c), and 6, our result also provides the level $1-\alpha$ which is closer to the specified value than the result for $k = 1$.

5. Conclusion

This paper provided the asymptotic expansion for the Studentized linear discriminant function based on monotone missing training data in subsection 3.1. Moreover, subsection 3.2 provided a certain method for determining the cut-off point which could be obtained using the result derived by subsection 3.1. As it turns out, we derive a certain extension of the results derived by Anderson (1973), Shutoh and Seo (2010), and McLachlan (1977). Section 4 evaluated our result by Monte Carlo simulation for the selected parameters. Especially, our result provided the better value of $1 - \alpha$, i.e., the value is closer to the specified level than the result for McLachlan (1977) when α, Δ, M, and Δ_i $(i = 1, 2)$ are lower.

As a future problem, if we can construct the discriminant rule based on maximum likelihood with k-step monotone missing training data, we may obtain the similar result.

Acknowledgements

The author is greatly indebted to Professor Takashi Seo of Tokyo University of Science for his valuable comments and suggestions.

References

Anderson, T. W., 1973. An asymptotic expansion of the distribution of the Studentized classification statistic W. The Annals of Statistics 1, 964–972.

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

Kanda, T., Fujikoshi, Y., 1998. Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data. American Journal of Mathematical and Management Sciences 18, 161–190.

Matsumoto, C., 2004. An optimal discriminant rule in the class of linear and quadratic discriminant functions for large dimension and samples. Hiroshima Mathematical Journal 34, 231–250.

McLachlan, G. J., 1977. Constrained sample discrimination with the Studentized classification statistic W. Communications in Statistics – Theory and Methods 6, 575–583.

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

Okamoto, M., 1963. An asymptotic expansion for the distribution of the linear discriminant function. Annals of Mathematical Statistics 34, 1286–1301.

Shutoh, N., 2010. An asymptotic expansion for the distribution of the linear discriminant function based on monotone missing training data. Submitted for publication.

Shutoh, N., 2011. Asymptotic expansions relating to discrimination based on two-step monotone missing samples. Journal of Statistical Planning and Inference 141, 1297–1306.

Shutoh, N., Seo, T., 2010. Asymptotic expansion of the distribution of the Studentized linear discriminant function based on two-step monotone missing samples. Communications in Statistics – Simulation and Computation 39, 1365–1383.

Wakaki, H., 1994. Discriminant analysis under elliptical populations. Hiroshima Mathematical Journal 24, 257–298.

Table 1 The selected parameters in simulation studies for Cases 1–8.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
p	3	3	3	3	3	3	3	7
p_1	1	1	1	1	1	1	1	3
p_2	1	1	1	1	1	1	1	2
p_3	1	1	1	1	1	1	1	2
Δ	1.05	1.05	1.05	0.50	1.36	1.05	1.05	1.05
M	0.20	0.20	0.20	0.20	0.20	0.10	0.25	0.20
α	0.05	0.10	0.01	0.05	0.05	0.05	0.05	0.05

Table 2 The values of $e_1(2|1)$ and $e_3(2|1)$ for Cases 1–8.

	$e_1(2\|1)$ for the result of McLachlan (1977)							
N_1	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.1212	0.1337	0.1050	0.1279	0.1194	0.0636	0.1525	0.1507
20	0.1256	0.1380	0.1079	0.1307	0.1243	0.0621	0.1599	0.1444
40	0.1384	0.1499	0.1206	0.1408	0.1379	0.0645	0.1784	0.1461
100	0.1571	0.1656	0.1427	0.1575	0.1571	0.0734	0.2011	0.1594
	$e_3(2\|1)$ for the proposed result							
$N_{(3)}$	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.1428	0.1536	0.1258	0.1384	0.1441	0.0698	0.1818	0.1644
20	0.1482	0.1583	0.1315	0.1438	0.1490	0.0707	0.1892	0.1595
40	0.1596	0.1677	0.1457	0.1566	0.1601	0.0749	0.2038	0.1635
100	0.1737	0.1794	0.1636	0.1725	0.1738	0.0828	0.2204	0.1745

Table 3 The values of $e_1(1|2)$ and $e_3(1|2)$ for Cases 1–8.

	$e_1(1\|2)$ for the result of McLachlan (1977)							
N_1	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.6260	0.6029	0.6575	0.8017	0.5086	0.7535	0.5696	0.6477
20	0.5983	0.5752	0.6327	0.7870	0.4780	0.7399	0.5376	0.6291
40	0.5446	0.5247	0.5783	0.7512	0.4214	0.7067	0.4786	0.5720
100	0.4945	0.4807	0.5196	0.7075	0.3731	0.6674	0.4279	0.5093
	$e_3(1\|2)$ for the proposed result							
$N_{(3)}$	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.5601	0.5418	0.5913	0.7710	0.4336	0.7158	0.4966	0.5881
20	0.5374	0.5204	0.5672	0.7533	0.4119	0.7003	0.4271	0.5678
40	0.4982	0.4850	0.5219	0.7174	0.3742	0.6692	0.4309	0.5180
100	0.4642	0.4556	0.4799	0.6820	0.3438	0.6391	0.3977	0.4737

Table 4 The values of the desired level $1-\alpha$ of confidence for Cases 1–8.

	The values $1-\alpha$ for the result of McLachlan (1977)							
N_1	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.8799	0.8376	0.9242	0.8507	0.8875	0.8256	0.8954	0.7604
20	0.8959	0.8518	0.9430	0.8701	0.9017	0.8562	0.9075	0.8125
40	0.9196	0.8723	0.9671	0.9034	0.9225	0.8997	0.9256	0.8821
100	0.9331	0.8834	0.9791	0.9262	0.9342	0.9247	0.9358	0.9195
	The values $1-\alpha$ for the proposed result							
$N_{(3)}$	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
15	0.8816	0.8331	0.9380	0.8693	0.8842	0.8441	0.8933	0.7539
20	0.8975	0.8474	0.9528	0.8861	0.9001	0.8700	0.9062	0.8123
40	0.9195	0.8676	0.9716	0.9115	0.9212	0.9062	0.9239	0.8833
100	0.9320	0.8792	0.9808	0.9273	0.9331	0.9272	0.9338	0.9188

Table 5 The selected parameters in simulation studies for Case 5(a), Case 5(b), and Case 5(c).

	Case 5(a)	Case 5(b)	Case 5(c)
Δ_1	1	$1/\sqrt{2}$	$1/\sqrt{3}$
Δ_2	1	1	$\sqrt{2}/\sqrt{3}$

Table 6 The values of $e_1(2|1)$ and $e_3(2|1)$ for Case 5(a), Case 5(b), and Case 5(c).

	$e_1(2\|1)$ for the result of McLachlan (1977)		
N_1	Case 5(a)	Case 5(b)	Case 5(c)
15	0.1194	0.1193	0.1192
20	0.1243	0.1244	0.1244
40	0.1379	0.1382	0.1385
100	0.1571	0.1576	0.1571
	$e_3(2\|1)$ for the proposed result		
$N_{(3)}$	Case 5(a)	Case 5(b)	Case 5(c)
15	0.1441	0.1384	0.1302
20	0.1490	0.1437	0.1357
40	0.1601	0.1558	0.1496
100	0.1738	0.1714	0.1662

Table 7 The values of $e_1(1|2)$ and $e_3(1|2)$ for Case 5(a), Case 5(b), and Case 5(c).

	$e_1(1\|2)$ for the result of McLachlan (1977)		
N_1	Case 5(a)	Case 5(b)	Case 5(c)
15	0.5086	0.6348	0.6859
20	0.4780	0.6136	0.6688
40	0.4214	0.5699	0.6322
100	0.3731	0.5264	0.5936
	$e_3(1\|2)$ for the proposed result		
$N_{(3)}$	Case 5(a)	Case 5(b)	Case 5(c)
15	0.4336	0.5848	0.6116
20	0.4119	0.5670	0.5924
40	0.3742	0.5319	0.5571
100	0.3438	0.5011	0.5219

Table 8 The values of the desired level $1-\alpha$ of confidence for Case 5(a), Case 5(b), and Case 5(c).

	The values $1-\alpha$ for the result of McLachlan (1977)		
N_1	Case 5(a)	Case 5(b)	Case 5(c)
15	0.8875	0.8872	0.8877
20	0.9017	0.9016	0.9014
40	0.9225	0.9220	0.9219
100	0.9342	0.9338	0.9342
	The values $1-\alpha$ for the proposed result		
$N_{(3)}$	Case 5(a)	Case 5(b)	Case 5(c)
15	0.8842	0.8917	0.9059
20	0.9001	0.9054	0.9176
40	0.9212	0.9249	0.9326
100	0.9331	0.9351	0.9405