

## Computable error bounds for asymptotic approximations of the quadratic discriminant function

Yasunori FUJIKOSHI

(Received June 6, 2019)

(Revised April 2, 2020)

**ABSTRACT.** This paper is concerned with computable error bounds for asymptotic approximations of the expected probabilities of misclassification (EPMC) of the quadratic discriminant function  $Q$ . A location and scale mixture expression for  $Q$  is given as a special case of a general discriminant function including the linear and quadratic discriminant functions. Using the result, we provide computable error bounds for asymptotic approximations of the EPMC of  $Q$  when both the sample size and the dimensionality are large. The bounds are numerically explored. Similar results are given for a quadratic discriminant function  $Q_0$  when the covariance matrix is known.

### 1. Introduction

An important concern in discriminant analysis is the classification of a  $p \times 1$  observation vector  $x$  as coming from one of two populations  $\Pi_1$  and  $\Pi_2$ . Let  $\Pi_i$  be  $p$  variate normal populations  $N_p(\mu_i, \Sigma)$ , where  $\mu_1 \neq \mu_2$  and  $\Sigma$  is positive definite. Suppose that all the parameters are unknown. However,  $N_i$ -samples are available from  $\Pi_i$ ,  $i = 1, 2$ . It is assumed that  $n = N - 2 > 0$ , where  $N = N_1 + N_2$ . Then, there are two well-known discriminant procedures. One is based on the linear discriminant function  $W$ , and the other is based on the quadratic discriminant function  $Q$ . The usual linear discriminant rule is to classify  $x$  as  $\Pi_1$  or  $\Pi_2$  according to  $W \geq 0$  or  $W < 0$ . Similarly the quadratic discriminant rule is defined by using  $Q$ .

These expected probabilities of misclassification (EPMC) have been obtained under two asymptotic frameworks; one is a large-sample asymptotic framework, and the other is a high-dimensional and large-sample asymptotic framework. Asymptotic results under a large-sample asymptotic framework

---

The author is supported by Grant-in-aid for Science Research (C), 16K00047, 2016–2018.

2010 *Mathematics Subject Classification.* Primary 62H30; Secondary 62E12.

*Key words and phrases.* Asymptotic approximations, Error bounds, Expected probability of misclassification, High-dimension, Large-sample, Linear discriminant function, Quadratic discriminant function.

were reviewed by Siotani [9] and by McLachlan [8]. Fujikoshi and Seo [3] derived asymptotic approximations of EPMC of a general discriminant function  $T_g$  including  $W$  and  $Z$  under a high-dimensional and large-sample asymptotic framework. Their extensions to asymptotic expansions were given in Fujikoshi [1] for  $W$ . Matsumoto [7] extended the result due to Fujikoshi and Seo [3] to an asymptotic expansion. For further results, see Hyodo and Kubokawa [6], Tonda et al. [10], and Yamada et al. [12].

This paper is concerned with computable error bounds for asymptotic approximations. These are based on a location and scale mixture of  $W$ , and use a general result on error bounds by Fujikoshi [1] and Fujikoshi and Ulyanov [4]. We note that a location and scale mixture can be obtained for a general discriminant function. These results will be useful for approximating  $T_g$  and its error bound. However, such general problems will be discussed in a future paper. Herein, we focus on the quadratic discriminant function  $Q$ .

The remainder of the paper is organized as follows. In Section 2, we provide preliminary results on a location and scale mixture of a normal distribution and its error bound, taking the linear discriminant function  $W$  as an example. In Section 3, we derive a location and scale mixture expression of a general discriminant function  $T_g$  including  $W$  and  $Q$ . It is noted that this may be applied for approximations of a general discriminant function  $T_g$  and its error bounds. However, in Section 4, details are discussed with respect for the quadratic discriminant function  $Q$ . We provide computable error bounds for high-dimensional and large-sample approximations for EPMC of  $Q$ , including details of their numerical accuracy. As a special case, we provide similar results for a quadratic discriminant function  $Q_0$  when the covariance matrix is known.

## 2. Preliminaries

**2.1. Discriminant functions.** Suppose that we are interested in classifying a  $p \times 1$  observation vector  $\mathbf{x}$  as coming from one of two populations  $\Pi_1$  and  $\Pi_2$ . Let  $\Pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$  ( $i = 1, 2$ ) be the two  $p$  variate normal populations, where  $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$  and  $\boldsymbol{\Sigma}$  is positive definite. When the parameters are unknown, we assume that random samples of sizes  $N_1$  and  $N_2$  are available from  $\Pi_1$  and  $\Pi_2$ , respectively. Let  $\bar{\mathbf{x}}_1$ ,  $\bar{\mathbf{x}}_2$  and  $\mathbf{S}$  be the sample mean vectors and the sample covariance matrix. It is assumed that  $n = N - 2 > p$ , where  $N = N_1 + N_2$ . Then, a well-known linear discriminant function is defined by

$$W = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \left\{ \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right\}. \quad (1)$$

The observation  $\mathbf{x}$  may be classified as  $\Pi_1$  or  $\Pi_2$  according to  $W \geq 0$  or  $W < 0$ .

In this paper, we consider classification of  $\mathbf{x}$  using a quadratic discriminant function  $Q$  defined by

$$Q = \frac{1}{2} \{ (1 + N_2^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) - (1 + N_1^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \}. \quad (2)$$

The observation  $\mathbf{x}$  may be classified to  $\Pi_1$  or  $\Pi_2$  according to  $Q \geq 0$  or  $Q < 0$ . The discriminant functions  $W$  and  $Q$  may be considered as special cases of a general discriminant function defined by

$$T_g = \frac{1}{2} \{ (\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) - g (\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \}, \quad (3)$$

where  $g$  is a positive number. The observation  $\mathbf{x}$  may be classified to  $\Pi_1$  or  $\Pi_2$  according to  $T_g \geq 0$  or  $T_g < 0$ . Then, it holds that

$$T_1 = W, \quad T_a = (1 + N_2^{-1})Q, \quad (4)$$

where  $a = (1 + N_2^{-1}) / (1 + N_1^{-1})$ .

**2.2. Error bounds for location and scale mixture variable.** Error estimates for asymptotic approximations of  $W$  have been studied by using its location and scale mixture of the standardized normal distribution. In general, a random variable  $Y$  is called a location and scale mixture of the standardized normal distribution, if  $Y$  is expressed as

$$Y = V^{1/2}Z - U, \quad (5)$$

where  $Z \sim N(0, 1)$ ,  $Z$  and  $(U, V)$  are independent, and  $V > 0$ . It is known (see Fujikoshi [1]) that the linear discriminant function  $W$  can be expressed as a location and scale mixture of the standardized normal distribution. In fact, when  $\mathbf{x}$  comes from  $\Pi_1$ , the variables  $(Z, U, V)$  may be defined as

$$\begin{aligned} V &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \Sigma \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \\ Z &= V^{-1/2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1), \\ U &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2} D^2, \end{aligned} \quad (6)$$

where  $D = \{ (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \}^{1/2}$  is the sample Mahalanobis distance between two populations. Since  $Z \sim N(0, 1)$  and is independent of  $(U, V)$ ,

we have

$$\Pr\{Y \leq y\} = E_{(U, V)}[\Phi\{V^{-1/2}(y + U)\}]. \quad (7)$$

From (7), we have an approximation  $\Phi\{v_0^{-1/2}(y + u_0)\}$  for the distribution function of  $Y$ , where  $(u_0, v_0)$  is a given point in the range space of  $(U, V)$ . Then the following bound was given by Fujikoshi [1].

**THEOREM 1.** *Let  $Y$  be a location and scale mixture of  $Z$  in (5). Let  $(u_0, v_0)$  be any given point in the range space of  $(U, V)$ . Assume that  $E(U^2) < \infty$  and  $E(V^2) < \infty$ . Then*

$$|\Pr\{Y \leq y\} - \Phi(\tilde{y})| \leq B_0 + B_1, \quad (8)$$

where  $\tilde{y} = v_0^{-1/2}(y + u_0)$ , and

$$\begin{aligned} B_0 &= \frac{1}{2\sqrt{2\pi e}} v_0^{-1} E[(U - u_0)^2] + \frac{1}{2} v_0^{-2} E[(V - v_0)^2] \\ &\quad + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{E[(U - u_0)^2] E[(V - v_0)^2]\}^{1/2}, \\ B_1 &= \frac{1}{\sqrt{2\pi}} v_0^{-1/2} |E(U - u_0)| + \frac{1}{2\sqrt{2\pi e}} v_0^{-1} |E(V - v_0)|. \end{aligned}$$

**COROLLARY 1.** *Under Theorem 1, assume that  $u_0 = E(U)$ , and  $v_0 = E(V)$ . Then*

$$|\Pr\{Y \leq y\} - \Phi(\tilde{y})| \leq B_0, \quad (9)$$

where  $\tilde{y} = v_0^{-1/2}(y + u_0)$ , and

$$\begin{aligned} B_0 &= \frac{1}{2\sqrt{2\pi e}} v_0^{-1} \text{Var}(U) + \frac{1}{2} v_0^{-2} \text{Var}(V) \\ &\quad + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{\text{Var}(U) \text{Var}(V)\}^{1/2}. \end{aligned}$$

### 3. Location and scale mixture for a general discriminant function

In this section we express a general discriminant function  $T_g$  as a location and scale mixture. Note that  $T_g$  can be expressed as

$$\begin{aligned} T_g &= \frac{1}{2} \{-\sqrt{g}(\mathbf{x} - \bar{\mathbf{x}}_1) + \mathbf{x} - \bar{\mathbf{x}}_2\}' \mathbf{S}^{-1} \{\sqrt{g}(\mathbf{x} - \bar{\mathbf{x}}_1) + \mathbf{x} - \bar{\mathbf{x}}_2\} \\ &= \frac{1}{2} b_1 b_2 \mathbf{t}'_1 \mathbf{B}^{-1} \mathbf{t}_2. \end{aligned} \quad (10)$$

Here

$$\begin{aligned}\mathbf{t}_1 &= b_1^{-1} \boldsymbol{\Sigma}^{-1/2} \{(1 - \sqrt{g})\mathbf{x} + \sqrt{g}\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\}, \\ \mathbf{t}_2 &= b_2^{-1} \boldsymbol{\Sigma}^{-1/2} \{(1 + \sqrt{g})\mathbf{x} - \sqrt{g}\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\}, \\ \mathbf{B} &= \boldsymbol{\Sigma}^{-1/2} \mathbf{S} \boldsymbol{\Sigma}^{-1/2},\end{aligned}\tag{11}$$

and

$$\begin{aligned}b_1 &= \{1 + N_2^{-1} - 2\sqrt{g} + g(1 + N_1^{-1})\}^{1/2}, \\ b_2 &= \{1 + N_2^{-1} + 2\sqrt{g} + g(1 + N_1^{-1})\}^{1/2}.\end{aligned}$$

Note that  $\mathbf{B}$  obeys the Wishart distribution  $W_p(n, \mathbf{I}_p)$ , and is independent of  $\mathbf{t}_1$  and  $\mathbf{t}_2$ . Suppose that  $\mathbf{x}$  belongs to  $\Pi_1$ . Then, it holds that

$$\mathbf{t}_i \sim N_p(b_i^{-1} \boldsymbol{\delta}, \mathbf{I}_p), \quad i = 1, 2,\tag{12}$$

where  $\boldsymbol{\delta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ . In general,  $\mathbf{t}_1$  and  $\mathbf{t}_2$  are not independent and their covariance matrix is computed as

$$\text{Cov}(\mathbf{t}_1, \mathbf{t}_2) = b_0(b_1 b_2)^{-1} \mathbf{I}_p,$$

where  $b_0 = 1 + N_2^{-1} - g(1 + N_1^{-1})$ . Therefor,  $\mathbf{t}_1$  and  $\mathbf{t}_2$  are independent if and only if

$$g = (1 + N_1^{-1})^{-1}(1 + N_2^{-1}) \equiv a,\tag{13}$$

i.e.,  $T_a = (1 + N_2^{-1})Q$ .

To express  $T_g$  as a location and scale mixture, let us consider a transformed variate  $\tilde{\mathbf{t}}_2$  of  $\mathbf{t}_2$  defined by

$$\tilde{\mathbf{t}}_2 = b_3^{-1/2} \left\{ \mathbf{t}_2 - \frac{1}{b_2} \boldsymbol{\delta} - \frac{b_0}{b_1 b_2} \left( \mathbf{t}_1 - \frac{1}{b_1} \boldsymbol{\delta} \right) \right\},\tag{14}$$

where  $b_3 = [1 - \{b_0/(b_1 b_2)\}^2]^{1/2}$ . Then,  $\tilde{\mathbf{t}}_2$  is independent of  $\mathbf{t}_1$ , since  $\mathbf{t}_1$  and  $\tilde{\mathbf{t}}_2$  are normal and  $\text{Cov}(\mathbf{t}_1, \tilde{\mathbf{t}}_2) = \mathbf{O}$ . We can write  $T_g$  in terms of  $\mathbf{t}_1$ ,  $\tilde{\mathbf{t}}_2$  and  $\mathbf{B}$  as

$$\begin{aligned}T_g &= \frac{1}{2} b_1 b_2 \mathbf{t}'_1 \mathbf{B}^{-1} \mathbf{t}_2 \\ &= \frac{1}{2} b_1 b_2 \mathbf{t}'_1 \mathbf{B}^{-1} \left\{ b_3 \tilde{\mathbf{t}}_2 + \frac{b_0}{b_1 b_2} \left( \mathbf{t}_1 - \frac{1}{b_1} \boldsymbol{\delta} \right) \right\} \\ &= \frac{1}{2} b_1 b_2 b_3 \{V^{1/2} Z - U\},\end{aligned}\tag{15}$$

where

$$\begin{aligned} Z &= (\mathbf{t}'_1 \mathbf{B}^{-2} \mathbf{t}_1)^{-1/2} \mathbf{t}'_1 \mathbf{B}^{-1} (\tilde{\mathbf{t}}_2 - b_2^{-1} \boldsymbol{\delta}), \\ U &= b_3^{-1} \left\{ -\frac{b_0}{b_1 b_2} \mathbf{t}'_1 \mathbf{B}^{-1} \mathbf{t}_1 + \frac{1}{b_2} \left( \frac{b_0}{b_1^2} - 1 \right) \mathbf{t}'_1 \mathbf{B}^{-1} \boldsymbol{\delta} \right\}, \\ V &= \mathbf{t}'_1 \mathbf{B}^{-2} \mathbf{t}_1. \end{aligned} \quad (16)$$

It is observed that  $Z \sim \mathbf{N}(0, 1)$ , and is independent of  $(U, V)$ . These imply the following Theorem.

**THEOREM 2.** *Let  $T_g$  be a general discriminant function defined by (3) based on  $N_i$  samples from  $\Pi_i : \mathbf{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ ,  $i = 1, 2$ . Then,  $T_g$  can be expressed as a location and scale mixture. More precisely, when  $\mathbf{x}$  belongs to  $\Pi_1$ , we can express as*

$$T_g = \frac{1}{2} b_1 b_2 b_3 \{ V^{1/2} Z - U \}, \quad (17)$$

where  $Z$ ,  $U$  and  $V$  are given by (16).

As a special case of Lemma 2, we have a location and scale expression of  $W$ . Note that the expression is different from that in (6). Similarly, we have a location and scale expression of  $Q$  as a special case of  $g = (1 + N_1^{-1})^{-1} (1 + N_2^{-1})$  whose result is essentially the same as that obtained by Yamada et al. [12].

Using Theorem 1 and Theorem 2, approximations for a general discriminant function  $T_g$  and its error bound can be obtained. It is interesting to study how the error bound depends on  $g$ . However, such results are beyond the scope of the current paper. In the next section, we focus on results for the quadratic discriminant function  $Q$ .

#### 4. Approximations for EPMC of $Q$ and error bounds

In this section we discuss approximations for the quadratic discriminant function  $Q$  which is given as a general discriminant function with  $g = a = (1 + N_1^{-1})^{-1} (1 + N_2^{-1})$ . Noting that  $b_3 = 1$ , from Theorem 2 we have

$$Q = (1 + N_2^{-1})^{-1} T_a = \frac{1}{2} (1 + N_2^{-1})^{-1} b_1 b_2 \mathbf{t}'_1 \mathbf{B}^{-1} \mathbf{t}_2, \quad (18)$$

where

$$\begin{aligned}
\mathbf{t}_1 &= b_1^{-1} \boldsymbol{\Sigma}^{-1/2} \{(-\sqrt{a} + 1)\mathbf{x} + \sqrt{a}\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\}, \\
\mathbf{t}_2 &= b_2^{-1} \boldsymbol{\Sigma}^{-1/2} \{(\sqrt{a} + 1)\mathbf{x} - \sqrt{a}\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\}, \\
\mathbf{B} &= \boldsymbol{\Sigma}^{-1/2} \mathbf{S} \boldsymbol{\Sigma}^{-1/2},
\end{aligned} \tag{19}$$

and

$$b_1 = \sqrt{2}\{1 + N_2^{-1} - \sqrt{a}\}^{1/2}, \quad b_2 = \sqrt{2}\{1 + N_2^{-1} + \sqrt{a}\}^{1/2}. \tag{20}$$

Suppose that  $\mathbf{x}$  belongs to  $\Pi_1$ , i.e.,  $\mathbf{x} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ . Then,  $\mathbf{t}_i \sim N_p(b_i^{-1}\boldsymbol{\delta}, \mathbf{I}_p)$ ,  $i = 1, 2$ ,  $n\mathbf{B} \sim W_p(n, \mathbf{I}_p)$ , and  $\mathbf{t}_1$ ,  $\mathbf{t}_2$  and  $\mathbf{B}$  are independent. Further, using Theorem 2, we have

$$Q = b\{V^{1/2}Z - U\}, \tag{21}$$

where

$$\begin{aligned}
Z &= (\mathbf{t}_1 \mathbf{B}^{-2} \mathbf{t}_1)^{-1/2} \mathbf{t}_1' \mathbf{B}^{-1} (\mathbf{t}_2 - b_2^{-1} \boldsymbol{\delta}), \\
U &= c_1 \boldsymbol{\gamma}' \mathbf{B}^{-1} \mathbf{t}_1, \quad V = \mathbf{t}_1' \mathbf{B}^{-2} \mathbf{t}_1.
\end{aligned} \tag{22}$$

Here,

$$\begin{aligned}
b &= [(1 + N^{-1}) / \{(1 + N_1^{-1})(1 + N_2^{-1})\}]^{1/2} c_2, \\
c_1 &= -b_1 b_2^{-1}, \quad c_2 = \{N / (N_1 N_2)\}^{1/2}, \\
\boldsymbol{\gamma} &= b_1^{-1} \boldsymbol{\delta}, \quad \tau^2 = \boldsymbol{\gamma}' \boldsymbol{\gamma} = b_1^{-2} \Delta^2.
\end{aligned} \tag{23}$$

Note that  $(U, V)$ 's in (22) and in (17) with  $g = a$  are the same.

In general, the Q-rule with a cutoff point 0 classifies  $\mathbf{x}$  as  $\Pi_1$  if  $Q > 0$  and  $\Pi_2$  if  $Q < 0$ . Then, there are two types of probability of misclassification. One is the probability of allocating  $\mathbf{x}$  into  $\Pi_2$  even though it actually belongs to  $\Pi_1$ . The other is the probability that  $\mathbf{x}$  is classified as  $\Pi_1$  although it actually belongs to  $\Pi_2$ . These two types of expected probabilities of misclassification (EPMC) for the Q-rule are expressed as

$$e_Q(2|1) = \Pr(Q < 0 | \mathbf{x} \in \Pi_1) \quad \text{and} \quad e_Q(1|2) = \Pr(Q > 0 | \mathbf{x} \in \Pi_2).$$

As is well known, the distribution of  $Q$  when  $\mathbf{x} \in \Pi_1$  is the same as that of  $-Q$  when  $\mathbf{x} \in \Pi_2$  by interchanging  $N_1$  and  $N_2$ . This indicates that  $e_Q(1|2)$  (or  $e_Q(1|2)$ ) is obtained from  $e_Q(2|1)$  (or  $e_Q(2|1)$ ) by replacing  $(N_1, N_2)$  with  $(N_2, N_1)$ . Thus, in this paper, we only deal with  $e_Q(2|1)$ . Then, we have the following expression:

$$\begin{aligned}
e_{\mathcal{Q}}(2|1) &= \Pr\{b(V^{1/2}Z - U) < 0\} \\
&= E_{(U,V)}\{\Phi(V^{-1/2}U)\}.
\end{aligned} \tag{24}$$

Next in the following we choose the range point  $(u_0, v_0)$  of  $(U, V)$  as

$$u_0 = E(U), \quad v_0 = E(V). \tag{25}$$

Consider approximating  $e_{\mathcal{Q}}(2|1)$  by  $\Phi(v_0^{-1/2}u_0)$ . For use of Theorem 1, the means and variances of  $U$  and  $V$  in (27) are required, and are given in the following Lemma:

LEMMA 1. *Let  $U$  and  $V$  be the random variables defined by (22). Then their means and variances are given as follows:*

$$\begin{aligned}
E(U) &= \frac{nc_1\tau^2}{m-1}, \quad m > 1, \\
\text{Var}(U) &= \frac{(nc_1)^2\tau^2}{(m-1)(m-3)} \left( \frac{n-1}{m} + \frac{2\tau^2}{m-1} \right), \quad m > 3, \\
E(V) &= \frac{n^2(n-1)(p+\tau^2)}{m(m-1)(m-3)}, \quad m > 3, \\
\text{Var}(V) &= \frac{n^4(n-1)}{m(m-1)(m-3)} \left[ \frac{2(n-3)(p+2\tau^2)}{(m-2)(m-5)(m-7)} \right. \\
&\quad \left. + (p+\tau^2)^2 \left\{ \frac{n-3}{(m-2)(m-5)(m-7)} - \frac{n-1}{m(m-1)(m-3)} \right\} \right], \\
&\quad m > 7,
\end{aligned} \tag{26}$$

where  $c_1$  is given by (23),  $m = n - p$ , and  $\tau^2 = b_1^{-2}A^2$ .

PROOF. The random variables  $U$  and  $V$  are expressed as

$$U = nc_1\gamma'\mathbf{A}^{-1}\mathbf{t}_1, \quad V = n^2\mathbf{t}_1'\mathbf{A}^{-2}\mathbf{t}_1, \tag{27}$$

where  $\mathbf{A} = n\mathbf{B}$ . Note that  $\mathbf{t}_1 \sim N_p(\gamma, \mathbf{I}_p)$ ,  $\mathbf{A} \sim W_p(n, \mathbf{I}_p)$ , and  $\mathbf{t}_1$  and  $\mathbf{A}$  are independent. The results are obtained by using the following distributional expressions (see, e.g., Fujikoshi [2], Yamada et al. [11]):

$$\begin{aligned}
\gamma'\mathbf{A}^{-1}\mathbf{t}_1 &= \tau Y_1^{-1} \{Z_1 + \tau - (Y_2/Y_3)^{1/2}Z_2\}, \\
\mathbf{t}_1'\mathbf{A}^{-2}\mathbf{t}_1 &= Y_1^{-2}(1 + Y_2 Y_3^{-1}) \{(Z_1 + \tau)^2 + Z_2^2 + Y_4\}.
\end{aligned}$$



Here,  $Y_i \sim \chi_{f_i}^2$ ,  $i = 1, \dots, 4$ ;  $Z_i \sim N(0, 1)$ ,  $i = 1, 2$ ; and

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

Further, all the variables  $Y_1, Y_2, Y_3, Y_4, Z_1$  and  $Z_2$  are independent.

Let us consider an approximation

$$e_Q(2|1) \sim \Phi(y_0), \quad y_0 = v_0^{-1/2} u_0, \quad (28)$$

where  $u_0 = E(U)$  and  $v_0 = E(V)$ . Applying Corollary 1 to this approximation, we have the following result.

**THEOREM 3.** *Let  $u_0$  and  $v_0$  be defined as  $u_0 = E(U)$  and  $v_0 = E(V)$ , which are given in (26), and  $y_0 = v_0^{-1/2} u_0$ . Then, if  $m = N_1 + N_2 - p - 2 > 7$ ,*

$$|e_Q(2|1) - \Phi(y_0)| \leq B_0, \quad (29)$$

where

$$B_0 = \frac{1}{2\sqrt{2\pi e}} v_0^{-1} V_U + \frac{1}{2} v_0^{-2} V_V + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{V_U V_V\}^{1/2}, \quad (30)$$

where  $V_U = \text{Var}(U)$  and  $V_V = \text{Var}(V)$  are given by (26).

Now, let us consider a high-dimensional and large-sample asymptotic framework given by

$$(\text{AF}): \quad p/N_i \rightarrow h_i > 0, \quad i = 1, 2, \quad \Delta^2 = O(1). \quad (31)$$

Then, under (AF), from Theorem 3 we have

$$B_0 = O_1, \quad \text{and} \quad e_Q(2|1) = \Phi(y_0) + O_1, \quad (32)$$

where  $O_j$  denotes the term of the  $j$ th order with respect to  $(N_1^{-1}, N_2^{-1}, p^{-1})$ . Hitherto, various approximation errors have been formally stated without rigorous proofs. However, by virtue of Theorem 3, our result (32) is based on a rigorous proof.

When  $\Sigma$  is known, we use the quadratic discriminant function  $Q_0$  defined by

$$Q_0 = \frac{1}{2} \left\{ (1 + N_2^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2)' \Sigma^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right. \\ \left. - (1 + N_1^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1)' \Sigma^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \right\}. \quad (33)$$

Assume that  $\mathbf{x}$  belongs to  $\Pi_1$ , i.e.,  $\mathbf{x} \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ . Then, we can write  $Q_0$  as

$$Q_0 = b \{ V_0^{1/2} Z_0 - U_0 \}. \quad (34)$$

Here,  $(Z_0, U_0, V_0)$  is defined from  $(Z, U, V)$  by putting  $\mathbf{B} = \mathbf{I}_p$ , that is,

$$Z_0 = (\mathbf{t}_1 \mathbf{t}_1)^{-1/2} \mathbf{t}'_1 (\mathbf{t}_2 - b_2^{-1} \boldsymbol{\delta}), \quad U_0 = c_1 \mathbf{y}' \mathbf{t}_1, \quad V_0 = \mathbf{t}'_1 \mathbf{t}_1, \quad (35)$$

and the constants  $b$ ,  $c_1$  and  $c_2$  are the same ones as in (23). The conditional distribution of  $Z_0$  given  $\mathbf{t}_1$  is  $N(0, 1)$ . Therefore,  $Z_0 \sim N(0, 1)$ , and  $Z_0$  is independent of  $\mathbf{t}_1$ . This implies that  $Q_0/b$  is a location and scale mixture of  $N(0, 1)$ . Note that the marginal distributions of  $(U_0, V_0)$  may be expressed as

$$U_0 = c_1(\tau X + \tau^2), \quad V_0 = \chi_p^2(\tau^2),$$

where  $X$  is the  $N(0, 1)$  variable. Using these distributional results, the means and variances of  $U_0$  and  $V_0$  are obtained as follows:

$$\begin{aligned} E(U_0) &= c_1 \tau^2, & \text{Var}(U_0) &= c_1^2 \tau^2, \\ E(V_0) &= p + \tau^2, & \text{Var}(V_0) &= (p + 2\tau^2). \end{aligned} \quad (36)$$

**THEOREM 4.** *Let  $\tilde{u}_0$  and  $\tilde{v}_0$  be defined as  $\tilde{u}_0 = E(U_0)$  and  $\tilde{v}_0 = E(V_0)$ , which are given in (36). Consider the error probability  $e_{Q_0}(2|1) = \Pr(Q_0 < 0 | \mathbf{x} \in \Pi_1)$ . Then, we have*

$$|e_{Q_0}(2|1) - \Phi(\tilde{y}_0)| \leq \tilde{B}_0, \quad (37)$$

where  $\tilde{y}_0 = \tilde{v}_0^{-1/2} \tilde{u}_0$ , and

$$\tilde{B}_0 = \frac{1}{2\sqrt{2\pi e}} \tilde{v}_0^{-1} V_{U_0} + \frac{1}{2} \tilde{v}_0^{-2} V_{V_0} + \frac{1}{2\sqrt{2\pi}} \tilde{v}_0^{-3/2} \{V_{U_0} V_{V_0}\}^{1/2}. \quad (38)$$

Here,  $V_{U_0} = \text{Var}(U_0)$ ,  $V_{V_0} = \text{Var}(V_0)$ , and they are given by (36).

We provide numerical values for the upper bounds  $B_0$  in (30) and  $\tilde{B}_0$  in (38) in Tables 4.1 and 4.2. Table 4.1 pertains to the case where  $\Delta = 1.68$ , and Table 4.2 to the case where  $\Delta = 2.56$ . As a matter of course, the bounds will be smaller as  $\Delta$  becomes larger. Similarly, the bounds when the covariance matrix is known are smaller in comparison to those when the covariance matrix is unknown. The bounds will be useful for moderate values as well as large values of  $p$  and for large values of  $N_1$  and  $N_2$  except for the case where  $m = N_1 + N_2 - p - 2$  is small, though their accuracy depends on whether the covariance matrix is known or unknown.

### Acknowledgement

The author is grateful to Dr. T. Yamada, Shimane University for many helpful comments. This research was partially supported by the Ministry of Education, Science, Sports, and Culture through a Grant-in-Aid for Scientific Research (C), 16K00047, 2016–2018.

Table 4.1. Values of  $B_0$  in (30) and  $\tilde{B}_0$  in (38);  $\Delta = 1.68$ 

$p$	$N_1$	$N_2$	$B_0$	$\tilde{B}_0$
5	10	10	1.1430	0.1112
	20	20	0.2762	0.0678
	30	10	0.2978	0.0855
	75	75	0.0581	0.0214
10	10	10	7.4916	0.0812
	20	20	0.3143	0.0558
	30	10	0.3280	0.0669
	75	75	0.0582	0.0201
30	30	30	0.2833	0.0272
	60	60	0.0809	0.0186
	90	60	0.0616	0.0165
	100	100	0.0438	0.0130

Table 4.2. Values of  $B_0$  in (30) and  $\tilde{B}_0$  in (38);  $\Delta = 2.56$ 

$p$	$N_1$	$N_2$	$B_0$	$\tilde{B}_0$
5	10	10	1.0846	0.0672
	20	20	0.2541	0.0371
	30	10	0.2671	0.0486
	75	75	0.0509	0.0107
10	10	10	7.2841	0.0567
	20	20	0.3032	0.0338
	30	10	0.3133	0.0429
	75	75	0.0521	0.0104
30	30	30	0.2867	0.0190
	60	60	0.0786	0.0113
	90	60	0.0587	0.0097
	100	100	0.0410	0.0073

## References

- [1] Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample size and dimensionality are large. *J. Multivariate Anal.*, **73**, 1–17.
- [2] Fujikoshi, Y. (2002). Selection of variables for discriminant analysis in a high-dimensional case. *Sankhyā Ser. A*, **64**, 256–257.
- [3] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant functions when the samples sizes and the dimension are large. *Statist. Anal. Random Arrays*, **6**, 269–280.
- [4] Fujikoshi, Y. and Ulyanov, V. V. (2006). On accuracy of approximations for location and scale mixture. *J. Math. Sci.*, **138**, 5390–5395.
- [5] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Analysis: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, New Jersey.

- [6] Hyodo, M. and Kubokawa, T. (2014). A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data. *J. Multivariate Anal.*, **123**, 364–379.
- [7] Matsumoto, C. (2004). An optimal discriminant rule in the class of linear and quadratic discriminant functions for large dimension and samples. *Hiroshima Math. J.*, **34**, 231–250.
- [8] McLachlan, G. J. (1991). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York.
- [9] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistic. *Handbook of Statistics 2* (P. R. Krishnaiah and L. N. Kanal, Eds.), North-Holland Publishing Company, 47–60.
- [10] Tonda, T., Nakagawa, T. and Wakaki, H. (2017). EPMC estimation in discriminant analysis when the dimension and sample are large. *Hiroshima Math. J.*, **47**, 43–62.
- [11] Yamada, T., Himeno, T. and Sakurai, T. (2017). Asymptotic cut-off point in linear discriminant rule to adjust the misclassification probability for large dimensions. *Hiroshima Math. J.*, **47**, 319–334.
- [12] Yamada, T., Sakurai, T. and Fujikoshi, Y. (2017). High-dimensional asymptotic results for EPMCs of W- and Z-rules. Hiroshima Statistical Research Group, TR: **17-12**.

*Yasunori Fujikoshi*  
*Department of Mathematics*  
*Graduate School of Science*  
*Hiroshima University*  
*Higashi-Hiroshima 739-8526, Japan*  
*E-mail: fujikoshi\_y@yahoo.co.jp*