Computable Error Bounds for Asymptotic Approximations of the Quadratic Discriminant Function

Yasunori Fujikoshi

Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8626, Japan

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.

AMS 2000 Subject Classification: primary 62H30; secondary 62H12

Key Words and Phrases: Asymptotic approximations, Error bounds, Expected probability of misclassification, High-dimension, Large-sample, Linear discriminant function, Quadratic discriminant function.

Abbreviated title: Error Bounds for the Quadratic Discriminant Function.

1 Introduction

An important concern in discriminant analysis is the classification of a $p \times 1$ observation vector $\boldsymbol{x}$ as coming from one of two populations Π_1 and Π_2. Let Π_i be p-variate normal populations $\mathrm{N}_p(\boldsymbol{\mu}_i, \Sigma)$, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and Σ 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 $\boldsymbol{x}$ as Π_1 or Π_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 framework. Asymptotic results under a large-sample framework were reviewed by Siotani (1982) and by McLachlan (1992). Fujikoshi and Seo (1998) derived asymptotic approximations of EPMC of a general discriminant function T_g including W and Z under a high-dimension and large-sample framework. Their extensions to asymptotic expansions were given in Fujikoshi (2000) for W. Matsumoto (2004) extended Fujikoshi and Seo's (1998) result to an asymptotic expansion. For further results, see Hyodo and Kubokawa (2014), Tonda et al. (2017), and Yamada et al. (2017).

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 (2000) and Fujikoshi and Ulyanov (2006). We note that a location and scale mixture can be obtained for a general discriminant function. These result 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 providegive 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 $\boldsymbol{x}$ as coming from one of two populations Π_1 and Π_2. Let $\Pi_i : N_p(\boldsymbol{\mu}_i, \Sigma)$ be the two p- variate normal, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ and Σ is positive definite. When the values of the parameters are unknown, we assume that random samples of sizes N_1 and N_2 are available from Π_1 and Π_2, respectively. Let $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$ 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

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}\left\{\boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\right\}. \tag{2.1}$$

The observation $\boldsymbol{x}$ may be classified as Π_1 or Π_2 according to $W \geq 0$ or $W < 0$.

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

$$\begin{aligned} Q =& \frac{1}{2}\left\{(1+N_2^{-1})^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_2)\right. \\ & \left.-(1+N_1^{-1})^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)\right\}. \end{aligned} \tag{2.2}$$

The observation $\boldsymbol{x}$ may be classified to Π_1 or Π_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

$$T_g = \frac{1}{2}\left\{(\boldsymbol{x}-\bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_2) - g(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)\right\}, \tag{2.3}$$

where g is a positive number. The observation $\boldsymbol{x}$ may be classified to Π_1 or Π_2 according to $T_g \geq 0$ or $T_g < 0$. Then, it holds that

$$T_1 = W, \quad T_a = (1+N_2^{-1})Q, \tag{2.4}$$

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

$$Y = V^{1/2}Z - U, \tag{2.5}$$

where $Z \sim \mathrm{N}(0,1)$, Z and (U, V) are independent, and $V > 0$. It is known (see Fujikoshi (2000)) that the linear discriminant function W can be expressed as a location and scal mixture of the standardized normal distribution. In fact, when $\boldsymbol{x}$ comes from Π_1, the variables (Z, U, V) may be defined as

$$
\begin{aligned}
V &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}\boldsymbol{\Sigma}\mathbf{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\
Z &= V^{-\frac{1}{2}}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1), \\
U &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}(\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2}D^2,
\end{aligned}
\tag{2.6}
$$

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

$$
\Pr\{Y \le y\} = \mathrm{E}_{(U,V)}\left[\Phi\left\{V^{-\frac{1}{2}}(y + U)\right\}\right]. \tag{2.7}
$$

From (2.7), we have an approximation $\Phi\left\{v_0^{-\frac{1}{2}}(y + u_0)\right\}$ 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 (2000),

Theorem 2.1. *Let Y be a location and scale mixture of Z in* (2.5). *Let (u_0, v_0) be any given point in the range space of (U, V). Assume that $\mathrm{E}(U^2) < \infty$ and $\mathrm{E}(V^2) < \infty$. Then*

$$
\left|\Pr\{Y \le y\} - \Phi(\tilde{y})\right| \le B_0 + B_1, \tag{2.8}
$$

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

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

Corollary 2.1. *Under Theorem* 2.1, *assume that $u_0 = \mathrm{E}(U)$, and $v_0 = \mathrm{E}(V)$. Then*

$$
\left|\Pr\{Y \le y\} - \Phi(\tilde{y})\right| \le B_0, \tag{2.9}
$$

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

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

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

$$\begin{aligned} T_g &= \frac{1}{2}\left\{-\sqrt{g}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)+\boldsymbol{x}-\bar{\boldsymbol{x}}_2\right\}'\mathbf{S}^{-1}\left\{\sqrt{g}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)+\boldsymbol{x}-\bar{\boldsymbol{x}}_2\right\} \\ &= \frac{1}{2}b_1b_2\boldsymbol{t}_1'\mathbf{B}^{-1}\boldsymbol{t}_2. \end{aligned} \tag{3.1}$$

Here

$$\begin{aligned} \boldsymbol{t}_1 &= b_1^{-1}\boldsymbol{\Sigma}^{-1/2}\left\{(1-\sqrt{g})\boldsymbol{x}+\sqrt{g}\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2\right\}, \\ \boldsymbol{t}_2 &= b_2^{-1}\boldsymbol{\Sigma}^{-1/2}\left\{(1+\sqrt{g})\boldsymbol{x}-\sqrt{g}\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2\right\}, \\ \mathbf{B} &= \boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2}, \end{aligned} \tag{3.2}$$

and

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

Note that $\mathbf{B}$ obeys a Wishart distribution $\mathrm{W}_p(n, \mathbf{I}_p)$, and is independent of $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$. Suppose that $\boldsymbol{x}$ belongs to Π_1. Then, it holds that

$$\boldsymbol{t}_i \sim \mathrm{N}_p(b_i^{-1}\boldsymbol{\delta}, \mathbf{I}_p), \quad i = 1, 2, \tag{3.3}$$

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

$$\mathrm{Cov}(\boldsymbol{t}_1, \boldsymbol{t}_2) = b_0(b_1b_2)^{-1}\mathbf{I}_p,$$

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

$$g = (1 + N_1^{-1})^{-1}(1 + N_2^{-1}) \equiv a, \tag{3.4}$$

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 $\tilde{\boldsymbol{t}}_2$ of $\boldsymbol{t}_2$,

$$\tilde{\boldsymbol{t}}_2 = \boldsymbol{t}_2 - \frac{b_0}{b_1 b_2}\left(\boldsymbol{t}_1 - \frac{1}{b_1}\boldsymbol{\delta}\right). \tag{3.5}$$

Then, $\tilde{\boldsymbol{t}}_2$ is independent of $\boldsymbol{t}_1$, since $\boldsymbol{t}_1$ and $\tilde{\boldsymbol{t}}_2$ are normal and $\mathrm{Cov}(\boldsymbol{t}_1, \tilde{\boldsymbol{t}}_2) = \mathbf{O}$. We can write T_g in terms of $\boldsymbol{t}_1$, $\tilde{\boldsymbol{t}}_2$ and $\mathbf{B}$ as

$$\begin{aligned} T_g &= \frac{1}{2} b_1 b_2 \boldsymbol{t}_1' \mathbf{B}^{-1} \boldsymbol{t}_2 \\ &= \frac{1}{2} b_1 b_2 \boldsymbol{t}_1' \mathbf{B}^{-1} \left\{ \tilde{\boldsymbol{t}}_2 + \frac{b_0}{b_1 b_2}\left(\boldsymbol{t}_1 - \frac{1}{b_1}\boldsymbol{\delta}\right)\right\} \\ &= \frac{1}{2} b_1 b_2 \left\{ V^{1/2} Z - U \right\}, \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} Z &= (\boldsymbol{t}_1' \mathbf{B}^{-2} \boldsymbol{t}_1)^{-1/2} (\tilde{\boldsymbol{t}}_2 - b_2^{-1}\boldsymbol{\delta}), \\ U &= -\frac{b_0}{b_1 b_2} \boldsymbol{t}_1' \mathbf{B}^{-1} \boldsymbol{t}_1 + \frac{1}{b_2}\left(\frac{b_0}{b_1^2} - 1\right) \boldsymbol{t}_1' \mathbf{B}^{-1} \boldsymbol{\delta}, \\ V &= \boldsymbol{t}_1' \mathbf{B}^{-2} \boldsymbol{t}_1. \end{aligned} \tag{3.7}$$

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

Lemma 3.1. *Let T_g be a general discriminant function defined by* (2.3) *based on N_i samples from $\Pi_i : \mathrm{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 1, 2$. Then, T_g can be expressed as a location and scale mixture. More precisely, when $\boldsymbol{x}$ belongs to Π_1, we can express as*

$$T_g = \frac{1}{2} b_1 b_2 \left\{ V^{1/2} Z - U \right\}, \tag{3.8}$$

where Z, U and V are given by (3.7).

As a special case of Lemma 3.1, we have a location and scale expression of W. Note that the expression is different from that in (2.6). 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 the same as that obtained by Yamada et al. (2017).

Using Theorem 2.1 and Lemma 3.1, 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 $g = a = (1 + N_1^{-1})^{-1}(1 + N_2^{-1})$. From Section 3, we have

$$Q = (1 + N_2^{-1})^{-1} T_a = \frac{1}{2}(1 + N_2^{-1})^{-1} b_1 b_2 \boldsymbol{t}_1' \mathbf{B}^{-1} \boldsymbol{t}_2, \tag{4.1}$$

where

$$\begin{aligned} \boldsymbol{t}_1 &= b_1^{-1} \boldsymbol{\Sigma}^{-1/2} \left\{ (-\sqrt{a} + 1)\boldsymbol{x} + \sqrt{a}\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 \right\}, \\ \boldsymbol{t}_2 &= b_2^{-1} \boldsymbol{\Sigma}^{-1/2} \left\{ (\sqrt{a} + 1)\boldsymbol{x} - \sqrt{a}\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 \right\}, \\ \mathbf{B} &= \boldsymbol{\Sigma}^{-1/2} \mathbf{S} \boldsymbol{\Sigma}^{-1/2}, \end{aligned} \tag{4.2}$$

and

$$b_1 = \sqrt{2} \left\{ 1 + N_2^{-1} - \sqrt{a} \right\}^{1/2}, \quad b_2 = \sqrt{2} \left\{ 1 + N_2^{-1} + \sqrt{a} \right\}^{1/2}. \tag{4.3}$$

Suppose that $\boldsymbol{x}$ belongs to Π_1, i.e., $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Then, $\boldsymbol{t}_i \sim \mathrm{N}(b_i^{-1}\boldsymbol{\delta},\ \mathbf{I}_p)$, $i = 1, 2$, $n\mathbf{B} \sim \mathrm{W}_p(n, \mathbf{I}_p)$, and $\boldsymbol{t}_1, \boldsymbol{t}_2$ and $\mathbf{B}$ are independent. Further, using Lemma 3.1, we have

$$Q = b \left\{ V^{1/2} Z - U \right\}, \tag{4.4}$$

where

$$Z = (\boldsymbol{t}_1\mathbf{B}^{-2}\boldsymbol{t}_1)^{-1/2}\boldsymbol{t}_1'\mathbf{B}^{-1}(\boldsymbol{t}_2 - b_2^{-1}\boldsymbol{\delta}),$$
$$U = c_1\boldsymbol{\gamma}'\mathbf{B}^{-1}\boldsymbol{t}_1, \quad V = c_2\boldsymbol{t}_1'\mathbf{B}^{-2}\boldsymbol{t}_1. \tag{4.5}$$

Here,

$$\begin{aligned} c_1 &= -\{N/(N_1N_2)\}^{1/2}b_1b_2^{-1}, \quad c_2 = \{N/(N_1N_2)\}^{1/2}, \\ b &= \left[(1+N^{-1})/\{(1+N_1^{-1})(1+N_2^{-1})\}\right]^{1/2}, \\ \boldsymbol{\gamma} &= b_1^{-1}\boldsymbol{\delta}, \quad \tau^2 = \boldsymbol{\gamma}'\boldsymbol{\gamma} = b_1^{-2}\Delta^2. \end{aligned} \tag{4.6}$$

Note that (U, V)'s in (4.5) and in (3.8) with $g = a$ are the same, except for the constant multiple.

In general, the Q-rule with a cutoff point 0 classifies $\boldsymbol{x}$ as Π_1 if $Q > 0$ and Π_2 if $Q < 0$. Then, there are two types of probability of misclassification. One is the probability of allocating $\boldsymbol{x}$ into Π_2 even though it actually belongs to Π_1. The other is the probability that $\boldsymbol{x}$ is classified as Π_1 although it actually belongs to Π_2. These two types of expected probabilities of misclassification (EPMC) for the Q- rule are expressed as

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

As is well known, the distribution of Q when $\boldsymbol{x} \in \Pi_1$ is the same as that of $-Q$ when $\boldsymbol{x} \in \Pi_2$ 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:

$$\begin{aligned} e_Q(2|1) &= \Pr\left\{b(V^{1/2}Z - U) < 0\right\} \\ &= \mathrm{E}_{(U,V)}\left\{\Phi(V^{-1/2}U)\right\}. \end{aligned} \tag{4.7}$$

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

$$u_0 = \mathrm{E}(U), \quad v_0 = \mathrm{E}(V). \tag{4.8}$$

Consider approximating $e_Q(2|1)$ by $\Phi(v_0^{-1/2}u_0)$. For use of Theorem 2.1, the means and variances of U and V in (4.10) are required, and are given in the following Lemma:

Lemma 4.1. *Let U and V be random variables defined by* (4.5)*. Then their means and variances are given as follows:*

$$\begin{aligned}
\mathrm{E}(U) &= \frac{nc_1\tau^2}{m-1}, \quad m>1, \\
\mathrm{Var}(U) &= \frac{(nc_1)^2\tau^2}{(m-1)(m-3)}\left(\frac{n-1}{m}+\frac{2\tau^2}{m-1}\right), \quad m>3, \\
\mathrm{E}(V) &= \frac{n^2c_2(n-1)(p+\tau^2)}{m(m-1)(m-3)}, \quad m>3, \\
\mathrm{Var}(V) &= \frac{(n^2c_2)^2(n-1)}{m(m-1)(m-3)}\left[\frac{2(n-3)(p+2\tau^2)}{(m-2)(m-5)(m-7)}\right. \\
&\quad \left.+(p+\tau^2)^2\left\{\frac{n-3}{(m-2)(m-5)(m-7)}-\frac{n-1}{m(m-1)(m-3)}\right.\right], \\
&\quad m>7,
\end{aligned} \tag{4.9}$$

where c_1 and c_2 are given by (4.6)*, $m=n-p$, and $\tau^2=b_1^{-2}\Delta^2$.*

Proof. The random variables U and V are expressed as

$$U=nc_1\boldsymbol{\gamma}'\mathbf{A}^{-1}\boldsymbol{t}_1, \quad V=n^2c_2\boldsymbol{t}_1'\mathbf{A}^{-2}\boldsymbol{t}_1, \tag{4.10}$$

where $\mathbf{A}=n\mathbf{B}$, Note that $\boldsymbol{t}_1\sim \mathrm{N}_p(\boldsymbol{\gamma},\mathbf{I}_p)$, $\mathbf{A}\sim \mathrm{W}_p(n,\mathbf{I}_p)$, and $\boldsymbol{t}_1$ and $\mathbf{A}$ are independent. The results are obtained by using the following distributional expressions (see, e.g., Fujikoshi (2002)):

$$\begin{aligned}
\boldsymbol{\gamma}'\mathbf{A}^{-1}\boldsymbol{t}_1 &= \tau Y_1^{-1}\left\{Z_1+\tau-(Y_2/Y_3)^{1/2}Z_2\right\}, \\
\boldsymbol{t}_1'\mathbf{A}^{-2}\boldsymbol{t}_1 &= Y_1^{-2}\left(1+Y_2Y_3^{-1}\right)\left\{(Z_1+\tau)^2+Z_2^2+Y_4\right\}.
\end{aligned}$$

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

$$e_Q(2|1) \sim \Phi(y_0), \quad y_0 = v_0^{-1/2} u_0, \tag{4.11}$$

where $u_0 = \mathrm{E}(U)$ and $v_0 = \mathrm{E}(V)$. Applying Corroraly 2.1 to this approximation, we have the following result.

Theorem 4.1. *Let u_0 and v_0 be defined as $u_0 = \mathrm{E}(U)$ and $v_0 = \mathrm{E}(V)$, which are given in* (4.9)*, and $y_0 = v_0^{-1/2} u_0$. Then, if $m = N_1 + N_2 - p - 2 > 7$,*

$$|e_Q(2|1) - \Phi(y_0)| \leq B_0, \tag{4.12}$$

where

$$B_0 = \frac{1}{2\sqrt{2\pi \mathrm{e}}} v_0^{-1} V_U + \frac{1}{2} v_0^{-2} V_V + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{V_U V_V\}^{1/2}. \tag{4.13}$$

where $V_U = \mathrm{Var}(U)$ and $V_V = \mathrm{Var}(V)$ are given by (4.9).

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

$$\text{(AF)}: \quad p/N_i \ \to h_i > 0, \quad i = 1, 2, \quad \Delta^2 = \mathrm{O}(1). \tag{4.14}$$

Then, under (AF), from Theorem 4.1 we have

$$B_0 = \mathrm{O}_1, \quad \text{and} \quad e_Q(2|1) = \Phi(y_0) + \mathrm{O}_1, \tag{4.15}$$

where O_j denotes the term of the jth 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 virture of Theorem 4.1, our result (4.15) is based on a rigorous proof.

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

$$\begin{aligned} Q_0 = & \frac{1}{2} \left\{ (1 + N_2^{-1})^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_2)' \mathbf{\Sigma}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_2) \right. \\ & \left. - (1 + N_1^{-1})^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_1)' \mathbf{\Sigma}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_1) \right\}. \end{aligned} \tag{4.16}$$

Assume that $\boldsymbol{x}$ belongs to Π_1, i.e., $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Then, we can write Q_0 as

$$Q_0 = b\left\{V_0^{1/2} Z_0 - U_0\right\}. \tag{4.17}$$

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

$$\begin{aligned} Z_0 &= (\boldsymbol{t}_1 \boldsymbol{t}_1)^{-1/2} \boldsymbol{t}_1'(\boldsymbol{t}_2 - b_2^{-1}\boldsymbol{\delta}), \\ U_0 &= c_1 \boldsymbol{\gamma}' \boldsymbol{t}_1, \quad V_0 = c_2 \boldsymbol{t}_1' \boldsymbol{t}_1, \\ c_1 &= -db_1 b_2^{-1} n^{-1/2}, \quad c_2 = d^2 n^{-1}. \end{aligned}$$

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

$$U_0 = c_1(\tau X + \tau^2), \quad V_0 = c_2 \chi_p^2(\tau^2),$$

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

$$\begin{aligned} &\mathrm{E}(U_0) = c_1 \tau^2 = u_0, \quad \mathrm{Var}(U_0) = c_1^2 \tau^2, \\ &\mathrm{E}(V_0) = c_2(p + \tau^2) = v_0, \quad \mathrm{Var}(V_0) = 2c_2(p + 2\tau^2). \end{aligned} \tag{4.18}$$

Theorem 4.2. *Let u_0 and v_0 be defined as $u_0 = \mathrm{E}(U_0)$ and $v_0 = \mathrm{E}(V_0)$, which are given in* (4.18). *Consider the error probability $e_{Q_0}(2|1) = \Pr(Q_0 < 0 | \boldsymbol{x} \in \Pi_1)$. Then, we have*

$$|e_{Q_0}(2|1) - \Phi(\tilde{y}_0)| \le \tilde{B}_0, \tag{4.19}$$

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

$$\tilde{B}_0 = \frac{1}{2\sqrt{2\pi \mathrm{e}}} \tilde{v}_0^{-1} V_{U_0} + \frac{1}{2} \tilde{v}_0^{-2} V_{V_0} + \frac{1}{2\sqrt{2\pi}} \tilde{v}_0^{-3/2} \{V_{U_0} V_{V_0}\}^{1/2}. \tag{4.20}$$

Here, $V_{U_0} = \mathrm{Var}(U_0)$, $V_{V_0} = \mathrm{Var}(V_0)$, and they are given by (4.18).

We provide numerical values for the upper bounds B_0 in (4.13) and $\tilde{B}_0$ in (4.20) 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 Δ 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 and the covariance matrix is unknown.

Table 4.1. Values of B_0 in (4.13) and $\tilde{B}_0$ in (4.20); $\Delta = 1.68$

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 (4.13) and $\tilde{B}_0$ in (4.20); $\Delta = 2.56$

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

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.

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., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Analysis: High-Dimensional and Large-Sample Approximations.* Wiley, Hoboken, New Jersey.

[5] 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.

[6] 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.

[7] McLachlan, G. J. (1991). *Discriminant Analysis and Statistical Pattern Recognition.* Wiley, New York.

[8] 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.

[9] Tonda, T., Nakagawa, T. and Wakaki, H. (2017). EPMC estimation in discriminant analysi when the dimension and sample are large. *Hiroshima Math. J.*, **47**, 43–62.

[10] Yamada, T., Sakurai, T. and Fujikoshi, Y. (2017). High-dimensional asymptotic results for EPMCs of W- and Z- rules. *Hiroshima Statistical Research Group*, **17-12**.