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 $\mathbf{x}$ as coming from one of two populations $\Pi_1$ and $\Pi_2$. Let $\Pi_i$ be $p$-variate normal populations $N_p(\boldsymbol{\mu}_i, \Sigma)$, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ 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 $\mathbf{x}$ 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 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 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 $\Pi_i : N_p(\mu_i, \Sigma)$ be the two $p$-variate normal, where $\mu_1 \neq \mu_2$ and $\Sigma$ 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 $\Pi_1$ 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

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

(2.1)

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

3
In this paper, we consider classification of $\mathbf{x}$ using a quadratic discriminant function $Q$ defined by

$$Q = \frac{1}{2} \left\{ (1 + N_2^{-1})^{-1}(\mathbf{x} - \bar{x}_2)'S^{-1}(\mathbf{x} - \bar{x}_2) 
- (1 + N_1^{-1})^{-1}(\mathbf{x} - \bar{x}_1)'S^{-1}(\mathbf{x} - \bar{x}_1) \right\}. \quad (2.2)$$

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

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

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

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

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

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

where $Z \sim N(0, 1)$, $Z$ 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 scale mixture of the standardized normal distribution. In fact, when $\mathbf{x}$ comes from $\Pi_1$, the variables $(Z, U, V)$ may be defined...
\[ V = (\bar{x}_1 - \bar{x}_2)' S^{-1} \Sigma S^{-1} (\bar{x}_1 - \bar{x}_2), \]
\[ Z = V^{-\frac{1}{2}} (\bar{x}_1 - \bar{x}_2)' S^{-1} (x - \mu_1), \]  
\[ U = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \mu_1) - \frac{1}{2} D^2. \]  

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

\[ \Pr \{ Y \leq y \} = \mathbb{E}_{(U, V)} \left[ \Phi \left\{ V^{-\frac{1}{2}} (y + U) \right\} \right]. \] 

(2.7)

From (2.7), we have an approximation \( \{ v_0^{-\frac{1}{2}} (y + u_0) \} \) 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 (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 \( \mathbb{E}(U^2) < \infty \) and \( \mathbb{E}(V^2) < \infty \). Then

\[ \left| \Pr \{ Y \leq y \} - \Phi(\tilde{y}) \right| \leq B_0 + B_1, \] 

(2.8)

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

\[ B_0 = \frac{1}{2\sqrt{2\pi}} v_0^{-1} \mathbb{E}[(U - u_0)^2] + \frac{1}{2} v_0^{-2} \mathbb{E}[(V - v_0)^2] \]
\[ + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{ \mathbb{E}[(U - u_0)^2] \mathbb{E}[(V - v_0)^2] \}^{1/2}, \]
\[ B_1 = \frac{1}{\sqrt{2\pi}} v_0^{-1/2} |\mathbb{E}(U - u_0)| + \frac{1}{2\sqrt{2\pi}} v_0^{-1} \mathbb{E}(V - v_0)|. \]

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

\[ \left| \Pr \{ Y \leq y \} - \Phi(\tilde{y}) \right| \leq B_0, \] 

(2.9)

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

\[ B_0 = \frac{1}{2\sqrt{2\pi}} v_0^{-1} \text{Var}(U) + \frac{1}{2} v_0^{-2} \text{Var}(V) \]
\[ + \frac{1}{2\sqrt{2\pi}} v_0^{-3/2} \{ \text{Var}(U) \text{Var}(V) \}^{1/2}. \]
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

$$T_g = \frac{1}{2} \{-\sqrt{g}(x - \bar{x}_1) + x - \bar{x}_2\}'S^{-1}\{\sqrt{g}(x - \bar{x}_1) + x - \bar{x}_2\} = \frac{1}{2}b_1b_2t_1'B^{-1}t_2.$$ 

Here

$$t_1 = b_1^{-1}\Sigma^{-1/2}\{(1 - \sqrt{g})x + \sqrt{g}\bar{x}_1 - \bar{x}_2\},$$

$$t_2 = b_2^{-1}\Sigma^{-1/2}\{(1 + \sqrt{g})x - \sqrt{g}\bar{x}_1 - \bar{x}_2\},$$

$$B = \Sigma^{-1/2}S\Sigma^{-1/2},$$

and

$$b_1 = \{1 + N_2^{-1} - 2\sqrt{g} + g(1 + N_1^{-1})\}^{1/2},$$

$$b_2 = \{1 + N_2^{-1} + 2\sqrt{g} + g(1 + N_1^{-1})\}^{1/2}.$$ 

Note that $B$ obeys a Wishart distribution $W_p(n, I_p)$, and is independent of $t_1$ and $t_2$. Suppose that $x$ belongs to $\Pi_1$. Then, it holds that

$$t_i \sim N_p(b_i^{-1}\delta, I_p), \quad i = 1, 2,$$ 

where $\delta = \Sigma^{-1/2}(\mu_1 - \mu_2)$. In general, $t_1$ and $t_2$ are not independent and their covariance matrix is computed as

$$\text{Cov}(t_1, t_2) = b_0(b_1b_2)^{-1}I_p,$$

where $b_0 = 1 + N_2^{-1} - g(1 + N_1^{-1})$. Therefore, $t_1$ and $t_2$ are independent if and only if

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

\(3.4\)
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 $\tilde{t}_2$ of $t_2$,

$$\tilde{t}_2 = t_2 - \frac{b_0}{b_1b_2} \left( t_1 - \frac{1}{b_1} \delta \right).$$

(3.5)

Then, $\tilde{t}_2$ is independent of $t_1$, since $t_1$ and $\tilde{t}_2$ are normal and $\text{Cov}(t_1, \tilde{t}_2) = 0$.

We can write $T_g$ in terms of $t_1$, $\tilde{t}_2$ and $B$ as

$$T_g = \frac{1}{2} b_1 b_2 t_1' B^{-1} t_2$$

$$= \frac{1}{2} b_1 b_2 t_1' B^{-1} \left\{ \tilde{t}_2 + \frac{b_0}{b_1b_2} \left( t_1 - \frac{1}{b_1} \delta \right) \right\}$$

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

(3.6)

where

$$Z = (t_1' B^{-2} t_1)^{-1/2} (\tilde{t}_2 - b_2^{-1} \delta),$$

$$U = -\frac{b_0}{b_1 b_2} t_1' B^{-1} t_1 + \frac{1}{b_2} \left( \frac{b_0}{b_1^2} - 1 \right) t_1' B^{-1} \delta,$$

$$V = t_1' B^{-2} t_1.$$

(3.7)

It is observed that $Z \sim N(0, 1)$, 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 $\Pi_i : N_p(\mu_i, \Sigma)$, $i = 1, 2$. Then, $T_g$ can be expressed as a location and scale mixture. More precisely, when $x$ belongs to $\Pi_1$, we can express as

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

(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

$$Q = (1 + N_2^{-1})^{-1} T_a = \frac{1}{2}(1 + N_2^{-1})^{-1} b_1 b_2 t_1' B^{-1} t_2,$$

(4.1)

where

$$t_1 = b_1^{-1} \Sigma^{-1/2} \{(-\sqrt{a} + 1)x + \sqrt{a} x_1 - x_2\},$$
$$t_2 = b_2^{-1} \Sigma^{-1/2} \{(\sqrt{a} + 1)x - \sqrt{a} x_1 - x_2\},$$

(4.2)

$$B = \Sigma^{-1/2} S \Sigma^{-1/2},$$

and

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

(4.3)

Suppose that $\mathbf{x}$ belongs to $\Pi_1$, i.e., $\mathbf{x} \sim N(\mu_1, \Sigma)$. Then, $t_i \sim N(b_i^{-1} \delta, I_p)$, $i = 1, 2$, $nB \sim W_p(n, I_p)$, and $t_1, t_2$ and $B$ are independent. Further, using Lemma 3.1, we have

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

(4.4)
where
\[
Z = (t_1 B^{-2} t_1)^{-1/2} t_1' B^{-1} (t_2 - b_2^{-1} \delta),
U = c_1 \gamma' B^{-1} t_1, \quad V = c_2 t_1' B^{-2} t_1.
\] (4.5)

Here,
\[
c_1 = -\{N/(N_1 N_2)\}^{1/2} b_1^{-1}, \quad c_2 = \{N/(N_1 N_2)\}^{1/2},
\]
\[
b = \left[\left(1 + N^{-1}\right)/\left\{\left(1 + N_1^{-1}\right)\left(1 + N_2^{-1}\right)\right\}\right]^{1/2},
\]
\[
\gamma = b_1^{-1} \delta, \quad \tau^2 = \gamma' \gamma = b_1^{-2} \Delta^2.
\] (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 \(x\) 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 \(x\) into \(\Pi_2\) even though it actually belongs to \(\Pi_1\). The other is the probability that \(x\) 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

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

As is well known, the distribution of \(Q\) when \(x \in \Pi_1\) is the same as that of \(-Q\) when \(x \in \Pi_2\) 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:

\[
e_Q(2|1) = \Pr \{b(V^{1/2} Z - U) < 0\}
= \mathbb{E}_{(U, V)} \{\Phi(V^{-1/2} U)\}.
\] (4.7)

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

\[
u_0 = \mathbb{E}(U), \quad v_0 = \mathbb{E}(V).
\] (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:

\[
\begin{align*}
E(U) &= \frac{nc_1 \tau^2}{m - 1}, \quad m > 1, \\
\text{Var}(U) &= \frac{(nc_1)^2 \tau^2}{(m - 1)(m - 3)} \left( \frac{n - 1}{m} + \frac{2\tau^2}{m - 1} \right), \quad m > 3, \\
E(V) &= \frac{n^2 c_2 (n - 1)(p + \tau^2)}{m(m - 1)(m - 3)}, \quad m > 3, \\
\text{Var}(V) &= \frac{(n^2 c_2)^2(n - 1)}{m(m - 1)(m - 3)} \left[ \frac{2(n - 3)(p + 2\tau^2)}{(m - 2)(m - 3)(m - 7)} - \frac{n - 1}{m(m - 1)(m - 3)} \right], \quad m > 7,
\end{align*}
\]

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

\[ U = nc_1 \gamma' A^{-1} t_1, \quad V = n^2 c_2 t_1' A^{-2} t_1, \tag{4.10} \]

where \( A = nB \). Note that \( t_1 \sim N_p(\gamma, I_p) \), \( A \sim W_p(n, I_p) \), and \( t_1 \) and \( A \) are independent. The results are obtained by using the following distributional expressions (see, e.g., Fujikoshi (2002)):

\[
\begin{align*}
\gamma' A^{-1} t_1 &= \tau Y_1^{-1} \left\{ Z_1 + \tau - (Y_2/Y_3)^{1/2} Z_2 \right\}, \\
t_1' A^{-2} t_1 &= Y_1^{-2} \left( 1 + Y_2 Y_3^{-1} \right) \left\{ (Z_1 + \tau)^2 + Z_2^2 + Y_4 \right\}.
\end{align*}
\]

Here, \( Y_i \sim \chi_i^2, i = 1, \ldots, 4; Z_i \sim N(0,1), i = 1, 2 \); 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. \( \square \)
Let us consider an approximation
\[ e_Q(2|1) \sim \Phi(y_0), \quad y_0 = v_0^{-1/2}u_0, \]  
(4.11)
where \( u_0 = E(U) \) and \( v_0 = E(V) \). Applying Corollary 2.1 to this approximation, we have the following result.

**Theorem 4.1.** Let \( u_0 \) and \( v_0 \) be defined as \( u_0 = E(U) \) and \( v_0 = E(V) \), which are given in (4.9), and \( y_0 = v_0^{-1/2}u_0 \). Then, if \( m = N_1 + N_2 - p - 2 > 7 \),
\[ |e_Q(2|1) - \Phi(y_0)| \leq B_0, \]  
(4.12)
where
\[ B_0 = \frac{1}{2\sqrt{2\pi e}}v_0^{-1}V_U + \frac{1}{2}v_0^{-2}V_V + \frac{1}{2\sqrt{2\pi e}}v_0^{-3/2}\{V_UV_V\}^{1/2}. \]  
(4.13)
where \( V_U = \text{Var}(U) \) and \( V_V = \text{Var}(V) \) are given by (4.9).

Now, let us consider a high-dimensional and large-sample asymptotic framework given by
\[ (AF) : \quad p/N_i \to h_i > 0, \quad i = 1, 2, \quad \Delta^2 = O(1). \]  
(4.14)
Then, under (AF), from Theorem 4.1 we have
\[ B_0 = O_1, \quad \text{and} \quad e_Q(2|1) = \Phi(y_0) + O_1, \]  
(4.15)
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 4.1, our result (4.15) is based on a rigorous proof.

When \( \Sigma \) is known, we use the quadratic discriminant function \( Q_0 \) defined by
\[ Q_0 = \frac{1}{2} \left\{ (1 + N_2^{-1})^{-1}(x - \bar{x}_2)'\Sigma^{-1}(x - \bar{x}_2) ight. \\
- (1 + N_1^{-1})^{-1}(x - \bar{x}_1)'\Sigma^{-1}(x - \bar{x}_1) \right\}. \]  
(4.16)
Assume that \(x\) belongs to \(\Pi_1\), i.e., \(x \sim N(\mu_1, \Sigma)\). Then, we can write \(Q_0\) as

\[
Q_0 = b \left\{ V_0^{1/2} Z_0 - U_0 \right\}.
\] (4.17)

Here, \((Z_0, U_0, V_0)\) is defined from \((Z, U, V)\) by putting \(B = I_p\), that is,

\[
\begin{align*}
Z_0 &= (t_1 t_1)^{-1/2} t_1 (t_2 - b_2^{-1} \delta), \\
U_0 &= c_1 \gamma' t_1, \\
V_0 &= c_2 t_1' t_1,
\end{align*}
\]

\[
\begin{align*}
c_1 &= -db_1 b_2^{-1} n^{-1/2}, \\
c_2 &= d^2 n^{-1}.
\end{align*}
\]

The conditional distribution of \(Z_0\) given \(t_1\) is \(N(0, 1)\). Therefore, \(Z_0 \sim N(0, 1)\), and \(Z_0\) is independent of \(t_1\). 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

\[
\begin{align*}
U_0 &= c_1 (\tau X + \tau^2), \\
V_0 &= c_2 \chi_p^2(\tau^2),
\end{align*}
\]

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:

\[
\begin{align*}
E(U_0) &= c_1 \tau^2 = u_0, \\
\text{Var}(U_0) &= c_1^2 \tau^2, \\
E(V_0) &= c_2 (p + \tau^2) = v_0, \\
\text{Var}(V_0) &= 2c_2 (p + 2 \tau^2). 
\end{align*}
\] (4.18)

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

\[
|e_{Q_0}(2|1) - \Phi(\tilde{y}_0)| \leq \tilde{B}_0,
\] (4.19)

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

\[
\tilde{B}_0 = \frac{1}{2\sqrt{2\pi}} \tilde{v}_0^{-1} V_{U_0} + \frac{1}{2} \tilde{v}_0^{-2} V_{V_0} + \frac{1}{2\sqrt{2\pi}} \tilde{v}_0^{-3/2} \{V_{U_0} V_{V_0}\}^{1/2}. 
\] (4.20)

Here, \(V_{U_0} = \text{Var}(U_0)\), \(V_{V_0} = \text{Var}(V_0)\), 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 $\Delta$ 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 and the covariance matrix is unknown.
Table 4.1. Values of $B_0$ in (4.13) and $\tilde{B}_0$ in (4.20); $\Delta = 1.68$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$N_1$</th>
<th>$N_2$</th>
<th>$B_0$</th>
<th>$\tilde{B}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>1.1430</td>
<td>0.1112</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0.2762</td>
<td>0.0678</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>10</td>
<td>0.2978</td>
<td>0.0855</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>75</td>
<td>0.0581</td>
<td>0.0214</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7.4916</td>
<td>0.0812</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0.3143</td>
<td>0.0558</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>10</td>
<td>0.3280</td>
<td>0.0669</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>75</td>
<td>0.0582</td>
<td>0.0201</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>30</td>
<td>0.2833</td>
<td>0.0272</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
<td>0.0809</td>
<td>0.0186</td>
</tr>
<tr>
<td>90</td>
<td>60</td>
<td>60</td>
<td>0.0616</td>
<td>0.0165</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0.0438</td>
<td>0.0130</td>
</tr>
</tbody>
</table>
Table 4.2. Values of $B_0$ in (4.13) and $\tilde{B}_0$ in (4.20); $\Delta = 2.56$

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

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


