Selection of the linear and the quadratic discriminant functions when the difference between two covariance matrices is small

Tomoyuki Nakagawa and Hirofumi Wakaki
Department of Mathematics, Graduate School of Science, Hiroshima University

(Last modified: April 19, 2016)

Abstract

We consider selecting of the linear and the quadratic discriminant functions in two normal populations. We do not know which of two discriminant functions lowers the expected probability of misclassification. When difference of the covariance matrices is large, it is known that the expected probability of misclassification of the quadratic discriminant functions is smaller than that of linear discriminant function. Therefore, we should consider only the selection when the difference between covariance matrices is small. In this paper we suggest a selection method using asymptotic expansion for the linear and the quadratic discriminant functions when the difference between the covariance matrices is small.

1 Introduction

We consider classifying an individual coming from one of two populations \( \Pi_1 \) and \( \Pi_2 \). We assume that \( \Pi_j \) is \( p \)-variate normal population with mean vector \( \mu_j \) and the covariance matrix \( \Sigma_j (j = 1, 2) \), that is,

\[
\Pi_1 : N_p(\mu_1, \Sigma_1), \quad \Pi_2 : N_p(\mu_2, \Sigma_2).
\]

When population parameters are known, the optimal Bayes discriminant function is given as follows (see Anderson, 1958):

\[
L(X; \mu_1, \mu_2, \Sigma) = (\mu_2 - \mu_1)\Sigma^{-1}\{X - (\mu_1 + \mu_2)/2\},
\]

where \( a' \) is the transpose of \( a \). In the case that the covariance matrices are equal, that is, \( \Sigma_1 = \Sigma_2 = \Sigma \), the optimal Bayes discriminant function is

\[
Q(X; \mu_1, \mu_2, \Sigma_1, \Sigma_2) = (X - \mu_1)'\Sigma_1^{-1}(X - \mu_1) - (X - \mu_2)'\Sigma_2^{-1}(X - \mu_2) + \log|\Sigma_2^{-1}\Sigma_1|,
\]

where \( |A| \) is the determinant of \( A \). \( L(X) \) and \( Q(X) \) are called the linear and the quadratic discriminant functions, respectively.

However, the population parameters are unknown in practice. Therefore, it is necessary for us to estimate the population parameters. A sample of size \( N_j \) coming from \( \Pi_j \), \( X_{j1}, \ldots, X_{jN_j} \) is available to estimate these parameters. Let \( \bar{X}_j \) and \( S_j \) be the sample mean and the sample covariance matrix of \( \Pi_j \), respectively, and let \( \bar{S} \) be the pooled sample covariance matrix, that is,

\[
\bar{X}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} X_{jk}, \quad \bar{S}_j = \frac{1}{n_j} \sum_{k=1}^{N_j} (X_{jk} - \bar{X}_j)(X_{jk} - \bar{X}_j)',
\]

\[
\bar{S} = k_1\bar{S}_1 + k_2\bar{S}_2,
\]

where \( n = n_1 + n_2 \) and \( k_j = n_j/n \) with \( n_j = N_j - 1 \) (\( j = 1, 2 \)). Replacing the unknown parameters with these estimators, we obtain \( \hat{L}(X) = L(X; \bar{X}_1, \bar{X}_2, \bar{S}) \) and \( \hat{Q}(X) = Q(X; \bar{X}_1, \bar{X}_2, \bar{S}_1, \bar{S}_2) \), respectively. \( \hat{L}(X) \) and \( \hat{Q}(X) \) are called the sample linear and the sample quadratic discriminant functions, respectively. An observation \( X \) is classified, for example into \( \Pi_1 \) for negative value of these functions.

If the covariance matrices are unequal, \( \hat{Q}(X) \) is better than \( \hat{L}(X) \) for large samples, since \( \hat{Q}(X) \) is a consistent estimator of the optimal Bayes discriminant function while \( \hat{L}(X) \) is not consistent. But even if the covariance matrices are unequal, \( \hat{Q}(X) \) is not always better than \( \hat{L}(X) \) for small samples. When the difference between the covariance matrices is large, \( \hat{Q}(X) \) is better than \( \hat{L}(X) \) (see Marks and Dunn, 1974; Wahl and Kronmal, 1977). These previous works are simulation study. Wakaki (1990) investigated performance of the two discriminant functions for moderately large samples using asymptotic expansions of the distributions of two discriminant functions under covariance matrices are proportional, that is, \( \Sigma_2 = c\Sigma_1 \) (\( c : \) constant).

We investigate performance of the discriminant functions using asymptotic expansions of the distributions of two discriminant functions under the difference between the covariance matrices is small, that is

\[
\Sigma_1 - \Sigma_2 = \frac{1}{\sqrt{n}}A \quad (A: \text{constant matrix}).
\]

1
We suggest a selection method for the discriminant functions using these asymptotic expansions.

The remainder of the present paper is organized as follows. In Section 2, we give asymptotic expansions for two discriminant functions when difference between two covariance matrices is small. In Section 3, we will suggest a selection method for the linear and the quadratic discriminant functions by using these asymptotic expansions. In Section 4, we will perform numerical study for investigating the performance of our selection method. In Section 5, we present a discussion and our conclusions.

2 Asymptotic expansion of the linear and the quadratic discriminant functions

We consider the following asymptotic framework:

\[ N_1 \to \infty, \quad N_2 \to \infty, \quad \frac{N_1}{N_2} = O(1), \quad \frac{N_2}{N_1} = O(1). \]

We can assume the following condition without loss of generality,

\[ \mu_1 + \mu_2 = 0, \quad k_1 \Sigma_1 + k_2 \Sigma_2 = I_p. \] (2.1)

2.1 Asymptotic expansion for the linear discriminant function

Wakaki (1990) proposed Theorem 2.1 which derived an asymptotic expansion of the sample linear discriminant function in a general case where covariance matrices are unequal.

**Theorem 2.1.** Let \( X \) be an observation vector coming from \( \Pi_j (j = 1, 2) \) and let

\[ L_j^* = (d' \Sigma_j d)^{-1/2} \{ \hat{L}(X) + \mu'_j d \}, \]

where \( d = \mu_1 - \mu_2 \). Then for large \( N_1 \) and \( N_2 \),

\[ P(L_j^* \leq x) = \Phi(x) + \sum_{i=1}^{2} \sum_{s=1}^{4} (d' \Sigma_i d)^{-s/2} a_{jis}^{*} H_{s-1}(x) \phi(x) + O_2, \] (2.2)

where \( \Phi(.) \) and \( \phi(.) \) are the distribution function and the density function of \( N(0,1) \), \( H_s(.) \)'s are the Hermite polynomials, and \( O_2 \) stands for the terms of the 2-th order with respect to \( N_1^{-1} \) and \( N_2^{-1} \). The coefficient \( a_{jis}^{*} \)'s are given by

\[
\begin{align*}
a_{j11} & = \frac{1}{2} (-1)^{i+1} \text{tr}(\Sigma_i) + k_i^2 \{ \mu'_j \Sigma_i^2 d + \text{tr}(\Sigma_i) \mu'_j \Sigma_i d \}, \\
a_{j12} & = \frac{1}{2} \{ \text{tr}(\Sigma_j \Sigma_i) + (\mu_j - \mu_i) \Sigma_i (\mu_j - \mu_i) \} \\\n & \quad - \frac{1}{2} k_i^2 \{ \text{tr}(\Sigma_j \Sigma_i) d' \Sigma_i d + 2 \text{tr}(\Sigma_i) d' \Sigma_j \Sigma_i d + d' \Sigma_j \Sigma_i \Sigma_i d + 2 d' \Sigma_j \Sigma_i^2 d + \frac{1}{2} (d' \Sigma_i d)^2 \}, \\
a_{j13} & = (\mu_j - \mu_i) \Sigma_j \Sigma_i d + 2 k_i^2 \mu'_j \Sigma_i d d' \Sigma_j \Sigma_i d, \\
a_{j14} & = \frac{1}{2} d' \Sigma_j \Sigma_i \Sigma_i d - \frac{1}{2} k_i^2 \{ d' \Sigma_j d d' \Sigma_i d + (d' \Sigma_j d)^2 \}.
\end{align*}
\]

We can easily obtain the following corollary under the assumption (1.2).

**Corollary 2.1.** Suppose that the condition of Theorem 2.1 and (1.2) hold. Let

\[ L_j = (d' d)^{-1/2} \{ \hat{L}(X) + \mu'_j d \}. \]

Then for large \( N_1 \) and \( N_2 \),

\[
\begin{align*}
P(L_j \leq x) & = \Phi(x) + \phi(x) \left[ - \frac{k_{s(j)}}{2 \sqrt{n}} D_0^{-1} D_1 H_1(x) - \frac{k_{s(j)}}{8n} D_0^{-2} D_1^2 H_3(x) + \sum_{i=1}^{2} \sum_{s=1}^{4} (D_0)^{-s/2} a_{jis}^{*} H_{s-1}(x) \right] + O_{3/2},
\end{align*}
\] (2.3)
where $D_k = d' A^k d$, $\sigma(1) = \sigma(2) + 1 = 2$. The coefficient $a_{ji}$'s are given by
\[
a_{j1} = -(-1)^{i+1}\frac{p}{2} + k^2_i \{ \mu_j' d + p \mu_j' d \},
\]
\[
a_{j2} = \frac{1}{2}(p + (\mu_j - \mu_i)'(\mu_j - \mu_i)) - \frac{1}{2}k^2_i \{ 3(p + 1)D_0 + D_0^2/2 \},
\]
\[
a_{j3} = (\mu_j - \mu_i)' d + 2k^2_i \mu_j' d D_0,
\]
\[
a_{j4} = -\frac{1}{2}D_0 - k^2_i D_0^2.
\]

2.2 Asymptotic expansion for the quadratic discriminant function

We derive an asymptotic expansion of the sample quadratic discriminant function under the assumption (1.2). The following lemmas are very important.

**Lemma 2.1.** If $W$ is a random matrix distributed as Wishart distribution $W_p(n, \Sigma)$, where $n$ is a positive integer and $Y$ is any $p$-variate random vector which is independent of $W$ with $P(Y = 0) = 0$ then
\[
Y'W^{-1}Y + \log |W| \sim V_{n-p+1}^{-1}Y'\Sigma^{-1}Y + \log |\Sigma| + \sum_{i=1}^p \log V_{n-i+1}, \quad V_{n-i+1} \sim \chi^2_{n-i+1},
\]
where $\chi^2_m$ is the chi-square distribution with $m$ degrees of freedom. Moreover, $Y'\Sigma^{-1}Y$ and $V_{n-i+1}(i = 1, \ldots, p)$ are mutually independent.

**Lemma 2.2** (characteristic function of noncentral chi-square distribution). Let $Z$ be a random variable distributed as normal distribution with mean $\mu$ and variance $1$. Then
\[
E[\exp(itZ^2)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} (z - \mu)^2 + itz^2 \right\} dz = (1 - 2it)^{-1/2} \exp \left\{ \frac{\mu^2it}{1 - 2it} \right\}.
\]

The proofs of the above two lemmas can be seen in Muirhead (1982) and Fujikoshi et al. (2010).

**Lemma 2.3.** Let $X$ be $p$-variate normal random vector with mean vector $0$ and covariance matrix $I_p$, and let
\[
g(t_1, t_2) = g(t_1, t_2; \eta_1, \eta_2, \Gamma_1, \Gamma_2) = \exp[t_1 (X - \eta_1)\Gamma_1 (X - \eta_1) + t_2 (X - \eta_2)\Gamma_2 (X - \eta_2)],
\]
\[
L(t_1, t_2) = L(t_1, t_2; \eta_1, \eta_2, \Gamma_1, \Gamma_2) = \tr[\{I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2\}^{-1}\Gamma_1]
\]
\[
+ \{\eta_1 - 2t_2 \Gamma_2(\eta_1 - \eta_2)\}'(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1}\{\eta_1 - 2t_2 \Gamma_2(\eta_1 - \eta_2)\}.
\]

Then
\[
g(t_1, t_2) = |I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2|^{-1/2} \exp \left\{ -\frac{1}{2} \eta_1' \eta_1 + t_2(\eta_1 - \eta_2)' \Gamma_2 (\eta_1 - \eta_2) \right\},
\]
\[
\frac{\partial g(t_1, t_2)}{\partial t_1} = g(t_1, t_2)L(t_1, t_2),
\]
\[
\frac{\partial L(t_1, t_2)}{\partial t_1} = L(t_1, t_2; \eta_1, \eta_2, \Gamma_1, \Gamma_2) = 2\tr[\{(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1}\Gamma_1\}^2]
\]
\[
+ 4\{\eta_1 - 2t_2 \Gamma_2(\eta_1 - \eta_2)\}'(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1}\{\Gamma_1(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1}\}^2(\eta_1 - 2t_2 \Gamma_2(\eta_1 - \eta_2)).
\]

The proof is given in Appendix. Using these lemmas, We obtain an asymptotic expansion of $Q(X)$ under the assumption (1.2) as in the following theorem.

**Theorem 2.2.** Let $X$ be an observation vector coming from $\Pi_j (j = 1, 2)$, and let $Q_j = (d'd)^{-1/2} \{Q(X)/2 + \mu_j' d\}$. Then for large $N_1$ and $N_2$,
\[
P(Q_j \leq x) = \Phi(x) + \phi(x)n^{-1/2} \sum_{s=1}^3 b_{js} H_{s-1}(x) + \sum_{i=0}^2 n_i^{-1} \sum_{s=1}^6 (d'd)^{-s/2} b_{js} H_{s-1}(x) + O_{3/2},
\]
where $n_0 = n$. The coefficient $b_{j_1}$’s and $b_{j_2}$’s are given by

\[ b_{j_1} = (-1)^{j_1} k_1 D_1/2, \quad b_{j_2} = -(1 + k_1) D_1/2, \quad b_{j_3} = (-1)^{j_1-1} D_1/2, \]

\[ b_{j_{01}} = (-1)^{j_1} \{ T_2/4 + k_2 D_2/2 \}, \quad b_{j_{02}} = -T_2/4 - (k_2^2 + 2k_1) D_2/2 - k_{v(j)} D_2^2/8, \]

\[ b_{j_{03}} = (-1)^{j_1} \{ (1 + 2k_1) D_2/2 + k_1(1 + k_1) D_1^2/4 \}, \quad b_{j_{04}} = D_2/2 - (1 + 4k_1 + k_1^2) D_1^2/8, \]

\[ b_{j_{05}} = (-1)^{j_1} (1 + k_2) D_1^2/4, \quad b_{j_{06}} = -D_1^2/8, \]

\[ b_{j_{11}} = (-1)^{j_1} (p^2 + 3p)/4 + (p + 1)(\mu_j - \mu_i)' d/2, \]

\[ b_{j_{12}} = -(p^2 + 3p)/4 - (4p + 5)((\mu_j - \mu_i)'(\mu_j - \mu_i))/2 \]

\[ b_{j_{13}} = -(p + 1)\mu_j' d D_0 + (p + 4)((\mu_j - \mu_i)' d D_0)/2 + (p + 4)(\mu_j - \mu_i)' d/2, \]

\[ b_{j_{14}} = -(p + 1)D_0/2 - 3((\mu_j - \mu_i)'(\mu_j - \mu_i))/2, \]

\[ b_{j_{15}} = (\mu_j - \mu_i)' d D_0, \]

\[ b_{j_{16}} = ((\mu_j - \mu_i)'(\mu_j - \mu_i))/2, \]

where $T_k = \text{tr}(A^k)$.

**Proof.**

Since $X_1, \ldots, X_{N_j}$ are normal random vectors with mean vector $\mu_j$ and covariance matrix $\Sigma_j$ and are mutually independent, $X_j$ is distributed as $p$-variate normal distribution with mean vector $\mu_j$ and covariance matrix $N_j^{-1}\Sigma_j$, and $n_j S_j$ is distributed as Wishart distribution $W_p(n_j, \Sigma_j)$. Suppose that $V_{jk}$ is the chi-square random variable with degrees of freedom $n_j - k + 1$. From Lemma 2.1 and 2.2,

\[
\mathbb{E} \left[ \exp \left( \frac{it}{2} \left( (X - X_j)' S_j^{-1}(X - X_j) + \log |S_j| \right) \right) \right] = \mathbb{E} \left[ \exp \left( \frac{it}{2} \left( n_j V_{jp}^{-1}(X - X_j)' S_j^{-1}(X - X_j) + \log |n_j^{-1} \Sigma_j| + \sum_{k=1}^{p} \log V_{jk} \right) \right) \right] = \mathbb{E} \left[ \exp \left( \frac{it}{2} \left( \log |n_j^{-1} \Sigma_j| + \sum_{k=1}^{p} \log V_{jk} \right) \right) \right].
\]

Let

\[
v_{jk} = \sqrt{\frac{n_j - k + 1}{2} \left( \frac{V_{jk}}{n_j - k + 1} - 1 \right)},
\]

then $v_{jk} = O_p(1)$ follows from the central limit theorem. The following result is given by using above formulae.

\[
\exp \left[ \frac{it n_j}{2 V_{jp}} \Omega_j \left( 1 - \frac{it n_j}{N_j V_{jp}} \right) \right] = e^{it n_j/2} \left( 1 + \frac{it}{2} \frac{n_j - v_{jk}}{\sqrt{\frac{2}{n_j} v_{jk}^2 + \frac{2}{n_j} v_{jp}^2}} + \frac{(it)^2}{4 \Omega_j} v_{jp}^2 \right) + O_p(n_j^{-3/2}),
\]

where $\Omega_j = (\mu_j - X)' S_j^{-1}(\mu_j - X)$.

\[
\exp \left( \frac{it}{2} \sum_{k=1}^{p} V_{jk} \right) = 1 + \frac{it}{2} \sum_{k=1}^{p} \left( \frac{v_{jk}}{\sqrt{n_j}} - \frac{v_{jk}^2}{n_j} \right) + \frac{(it)^2}{8} \left( \sum_{k=1}^{p} \frac{v_{jk}^2}{n_j} \right)^2 + O_p(n_j^{-3/2}).
\]

From (2.4), (2.5), $E[v_{jk}] = 0$, $E[v_{jk}^2] = 1$ and Lemma 2.1,

\[
\mathbb{E} \left[ \exp \left( \frac{it}{2} \left( (X - X_j)' S_j^{-1}(X - X_j) + \log |S_j| \right) \right) \right] = \exp(it \Omega_j) \left[ 1 + \frac{1}{n_j} \left( \frac{it}{2} \left( -\frac{(p - p - 1)}{2} + (p + 1) \Omega_j \right) + \frac{(it)^2}{4} (p + \Omega_j^2) \right) \right] + O_p(n_j^{-3/2}).
\]

Suppose that $X$ belongs to $\Pi_j$, and let

\[
\psi_j(t) = \mathbb{E}[e^{it Q(X)}/2].
\]
From Lemma 2.3 and (2.6),
\[
\psi_j(t) = g(it/2, -it/2)|\Sigma_1 \Sigma_2^{-1}|^{it/2} \left[ 1 + \frac{1}{n_1} \left\{ \frac{it}{2} \left( -\frac{p(p-1)}{2} + (p+1)L_1 \right) + \frac{(it)^2}{4} (p + (L_1^2 + L_{11})) \right\} 
+ \frac{1}{n_2} \left\{ -\frac{it}{2} \left( \frac{p(p-1)}{2} + (p+1)L_2 \right) + \frac{(it)^2}{4} (p + (L_2^2 + L_{22})) \right\} \right] + O_{3/2}, \quad (k = 1, 2),
\]
where
\[
L_k = \frac{1}{g(t_1, t_2)} \left( \frac{\partial g(t_1, t_2)}{\partial t_k} \right)_{(t_1, t_2) = (it/2, -it/2)}, \quad L_{kk} = \frac{\partial}{\partial t_k} \frac{1}{g(t_1, t_2)} \left( \frac{\partial g(t_1, t_2)}{\partial t_k} \right)_{(t_1, t_2) = (it/2, -it/2)}, \quad (k = 1, 2).
\]
From (1.2) and (2.1), we have
\[
\Sigma_1 = I_p + \frac{k_2}{\sqrt{n}} A, \quad \Sigma_2 = I_p - \frac{k_1}{\sqrt{n}} A. \tag{2.7}
\]
For \(j = 1\), the parameters of \(g\) and \(L\) are given as follows.
\[
\eta_1 = 0, \quad \eta_2 = \Sigma_1^{-1/2}(\mu_2 - \mu_1), \quad \Gamma_1 = I_p, \quad \Gamma_2 = \Sigma_1^{1/2} \Sigma_2^{-1} \Sigma_1^{1/2}. \tag{2.8}
\]
From (2.7) and (2.8), we obtain the following expansion.
\[
g(it/2, -it/2)|\Sigma_1 \Sigma_2^{-1}|^{it/2} = \exp \left[ -\frac{it}{2} (1 - it)D_0 + \frac{1}{\sqrt{n}} \left\{ \frac{k_2 (1 - it)}{2} D_1 - \frac{(k_2 + it)(1 - it)}{2} D_1 \right\} 
+ \frac{1}{n} \left\{ \frac{(it)^2 - it}{2} D_2 - \frac{k_2^2 (1 - it)}{2} D_2 + \frac{(1 - it)^2 (k_2^2 - k_1 it + k_2^2)}{2} D_2 \right\} \right] + O_{3/2},
\]
\[
L_1 = p + (it)^2 D_0 + O_{1/2}, \quad L_2 = p + (1 - it)^2 D_0 + O_{1/2}, \quad L_{11} = 2p + 4(it)^2 D_0 + O_{1/2}, \quad L_{22} = 2p + 4(1 - it)^2 D_0 + O_{1/2}.
\]
Hence, we obtain the following expansion of \(\psi_1\).
\[
\psi_1(t) = \exp \left\{ -\frac{it(1 - it)}{2} D_0 \right\} \left[ 1 + \frac{1}{n} \left\{ \frac{1}{2} (-k_1 it + (1 + k_1)(it)^2 - (it)^3) \right\} D_1 
+ \frac{1}{n} \left\{ \frac{(it)^2 - it}{4} T_2 + \frac{1}{2} \left\{ -k_2 it + (k_2^2 + 2k_1)(it)^2 - (1 + 2k_1)(it)^3 + (it)^4 \right\} D_2 \right. 
+ \frac{1}{8} \left\{ k_2 it + (1 + k_1)(it)^2 - (it)^3 \right\}^2 D_2 \right. 
+ \frac{1}{n_1} \left\{ \frac{it}{2} \left( p^2 + 3p \right) D_0 + \frac{(it)^2}{4} \left\{ p + (p + (it)^2 D_0)^2 + 2p + 4(it)^2 D_0 \right\} \right. 
+ \frac{1}{n_2} \left\{ -\frac{it}{2} \left( p^2 + 3p \right) D_0 + (p + 1)(1 - it)^2 D_0 \right. \right. 
+ \frac{(it)^2}{4} \left\{ p + (p + (1 - it)^2 D_0)^2 + 2p + 4(1 - it)^2 D_0 \right\} \right\} + O_{3/2}.
\]

The expansion of \(\psi_2\) is given by replacing the parameters \((it, k_1, k_2, n_1, n_2)\) of \(\psi_1\) with \((-it, k_2, k_1, n_2, n_1)\). Inverting \(\psi_j (j = 1, 2)\) formally, we obtain the desired result.

\[\square\]

3 The criterion for selecting between the linear and quadratic discriminant functions

3.1 Derivation of the criterion

In this section, we consider the expected misclassification probabilities \(P_L\) and \(P_Q\) of \(\hat{L}(X)\) and \(\hat{Q}(X)\), respectively. We assume that a priori probabilities are 1/2. Then \(P_L\) and \(P_Q\) are given by
\[
P_L = \frac{1}{2} \left\{ 1 - P \left( L_1 \leq \frac{1}{2} (d'd)^{1/2} \right) + P \left( L_2 \leq -\frac{1}{2} (d'd)^{1/2} \right) \right\},
\]
\[
P_Q = \frac{1}{2} \left\{ 1 - P \left( Q_1 \leq \frac{1}{2} (d'd)^{1/2} \right) + P \left( Q_2 \leq -\frac{1}{2} (d'd)^{1/2} \right) \right\},
\]
respectively.
Then $D(A, d)$ is given by following formula.

$$D(A, d) = \phi(D_0^{1/2}/2)\sum_{s=1}^{6} D_0^{-s/2} H_{s-1}(D_0^{1/2}/2)c_s,$$

where the coefficient $c_s$'s are given by

- $c_0 = -T_2/2 - (k_2^2 + k_3^2)D_2 + \{k_2^2(p+1)D_0 - (p+1)D_0^2/2\}/k_1 + \{k_3^2(p+1)D_0 - (p+1)D_0^2/2\}/k_2,$
- $c_1 = T_2/2 + (k_1^2 + k_2^2)D_2/2 + (k_1 + k_2)D_1^2/8$
  + \{(p^2 + p)/2 + (4p + 5)D_0^2/4 + (p + 1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2)/2\}/k_1
  + \{(p^2 + p)/2 + (4p + 5)D_2^2/4 + (p + 1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2)/2\}/k_2,$
- $c_2 = -2D_2 - (1 + k_2^2 + k_3^2)D_4^2/4 + \{-D_0^2 - (p + 1)D_0 + 2k_1D_0^2\}/k_1 + \{-D_0^2 - (p + 1)D_0 + 2k_2D_0^2\}/k_2,$
- $c_3 = D_2 + 3D_2^2/4 + \{(p + 1)D_0 + 3D_0^2/2 - 2k_1^2D_0^2\}/k_1 + \{(p + 1)D_0 + 3D_0^2/2 - 2k_2^2D_0^2\}/k_2,$
- $c_4 = -3D_0^2/4 - D_0^2/2k_1 - D_0^2/k_2,$
- $c_5 = D_0^2/4 + D_0^2/2k_1 + D_0^2/2k_2.$

Proof. The result is easily obtained from Theorem 2.1 and 2.2.

We obtain a criterion for selecting between the linear and the quadratic discriminant functions as $D(A, d)$. Then if $D(A, d)$ is negative, we can consider that $\hat{Q}(X)$ is better than $\hat{L}(X)$. Otherwise, we can consider that $\hat{L}(X)$ is better than $\hat{Q}(X)$. However, $A$ and $d$ are unknown parameters which should be estimated. We may consider to use simple estimators,

$$\hat{d} = S^{-1/2}(X_1 - X_2), \quad \hat{A} = \sqrt{n}S^{-1/2}(S_1 - S_2)S^{-1/2}. \quad (3.1)$$

But it is insufficient for a criterion for selecting between the linear and the quadratic discriminant functions only by replacing the unknown parameters with these estimators. Because, $\hat{A}$ is not consistent, $D(A, d)$ does not converge in probability to $D(A, d)$ for large samples. Moreover, $E[D(\hat{A}, \hat{d})]$ do not converge to $D(A, d)$. Therefore, we correct the bias in the next section.

### 3.2 Correcting the bias

We have the criterion for selecting between the linear and the quadratic discriminant functions in Theorem 3.1. But this include the unknown parameters $A$, $d$. When replacing these parameters with the estimators (3.1), $D(\hat{A}, \hat{d})$ have the asymptotic bias for $D(A, d)$. So we will correct the bias of the criterion. The criterion can be given as a linear combination of the following terms.

$$\text{tr}(A^2)(d'd')^{1/2}, \quad (d'd')^{1/2}d'Ad, \quad (d'd')^{1/2}(d'Ad)^2, \quad (d'd')^{1/2}. \quad (l \in \mathbb{Z}). \quad (3.2)$$

Theorem 3.2. We obtain the criterion $D^*$ which correct the bias as the following formula. Define

$$D^*(A, d) = \phi(D_0^{1/2}/2)\sum_{s=1}^{6} D_0^{-s/2} H_{s-1}(D_0^{1/2}/2)c_s^*,$$

where the coefficient $c_s^*$'s are given by

- $c_1^* = -T_2/2 - (k_2^2 + k_3^2)D_2 + \{k_2^2(p+1)D_0 - (p+1)D_0^2/2\}/k_1 + \{k_3^2(p+1)D_0 - (p+1)D_0^2/2\}/k_2,$
- $c_2^* = T_2/2 + (k_1^2 + k_2^2)D_2/2 + (k_1 + k_2)D_1^2/8$
  + \{(p^2 + p)/2 + (4p + 5)D_0^2/4 + (p + 1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2)/2\}/k_1
  + \{(p^2 + p)/2 + (4p + 5)D_2^2/4 + (p + 1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2)/2\}/k_2,$
- $c_3^* = -2D_2 - (1 + k_2^2 + k_3^2)D_4^2/4 + \{-D_0^2 - (p + 1)D_0 + 2k_1D_0^2\}/k_1 + \{-D_0^2 - (p + 1)D_0 + 2k_2D_0^2\}/k_2,$
- $c_4^* = D_2 + 3D_2^2/4 + \{(p + 1)D_0 + 3D_0^2/2 - 2k_1^2D_0^2\}/k_1 + \{(p + 1)D_0 + 3D_0^2/2 - 2k_2^2D_0^2\}/k_2,$
- $c_5^* = -3D_0^2/4 - D_0^2/2k_1 - D_0^2/k_2,$
- $c_6^* = D_0^2/4 + D_0^2/2k_1 + D_0^2/2k_2.$
where the coefficient $c^*_i$’s are given by

\[ c^*_i = -\frac{T_2}{2} - \frac{(p^2 + p)}{k_1} - \frac{(p^2 + p)}{k_2} - \frac{k_1^2}{D_2 - (p + 1)D_0/k_1} - \frac{(p - 1)D_0}{k_2} \]

+ \left\{ k_2^2 \frac{(p + 1)D_0 - (p + 1)D_0^2/k_1} + \frac{k_2^2}{(p + 1)D_0 - (p + 1)D_0^2/k_2} \right\}/k_2.

\[ c^*_2 = \frac{T_2}{2} - \frac{(p^2 + p)}{k_1} - \frac{(p^2 + p)}{k_2} \]

+ \left\{ \frac{k_1^2}{D_2 - (p + 1)D_0/k_1} - \frac{k_2^2}{D_2 - (p + 1)D_0/k_2} \right\}/k_2.

\[ c^*_3 = \frac{T_2}{2} - \frac{(p^2 + p)}{k_1} - \frac{(p^2 + p)}{k_2} \]

+ \left\{ \frac{k_1^2}{D_2 - (p + 1)D_0/k_1} - \frac{k_2^2}{D_2 - (p + 1)D_0/k_2} \right\}/k_2.

Then

\[ E[D^*(\hat{A}, \hat{d})] = D(A, d) + O(n^{-1/2}). \]

Proof.

From the central limit theorem, we can obtain the following,

\[ Z_j = \sqrt{n}(\hat{X}_j - \mu_j) = O_p(1), \quad W_j = \sqrt{n}(S_j - \Sigma_j) = O_p(1), \quad (j = 1, 2). \]

Then the following statistics are expanded as

\[ \text{tr}(\hat{A})(\hat{d} \hat{d}^T)^{1/2} = \text{tr}\left\{ \sqrt{n}S^{-1}(S_1 - S_2) \right\} \left\{ (\hat{X}_1 - \hat{X}_2)^T S^{-1}(\hat{X}_1 - \hat{X}_2) \right\}^{1/2} \]

= \text{tr}\left\{ (A + k_1^{1/2}W_1 - k_2^{1/2}W_2^2) \right\} \left( (\hat{d} \hat{d}^T)^{1/2} + O_p(n^{-1/2}) \right),

\[ (\hat{d} \hat{d}^T)^{1/2} \hat{A} \hat{d} = \left\{ (\hat{X}_1 - \hat{X}_2)^T S^{-1}(\hat{X}_1 - \hat{X}_2) \right\}^{1/2} \left\{ (\hat{X}_1 - \hat{X}_2)^T S^{-1}(\hat{X}_1 - \hat{X}_2) \right\}^{1/2} \]

= \left\{ (\hat{d} \hat{d}^T)^{1/2} \right\} \left\{ (A + k_1^{1/2}W_1 - k_2^{1/2}W_2^2) \right\} \hat{d} + O_p(n^{-1/2}).

Moreover, $\hat{W}_1$ is independent of $\hat{W}_2$ and the following formulae can be seen in Fujikoshi et al. (2010).

\[ E[\hat{W}_1] = O, \quad E[\hat{W}_1 \hat{B} \hat{W}_2] = \text{tr}(B \Sigma_1) \Sigma_j + \Sigma_j' B \Sigma_j, \]

where $B$ is an arbitrary constant matrix. Thus we obtain the following,

\[ E \left[ \text{tr}(\hat{A})(\hat{d} \hat{d}^T)^{1/2} \right] = (\hat{d} \hat{d}^T)^{1/2} \left[ \text{tr}(A^2) + (p^2 + p)/k_1 + (p^2 + p)/k_2 \right] + O(n^{-1/2}). \]

\[ E \left[ (\hat{d} \hat{d}^T)^{1/2} \hat{A} \hat{d} \right] = (\hat{d} \hat{d}^T)^{1/2} \left\{ (\hat{d} A \hat{d})^2 + 2(\hat{d} \hat{d}^T)^2/k_1 + 2(\hat{d} \hat{d}^T)^2/k_2 \right\} + O(n^{-1/2}). \]

Using these formulae, replacing \( \text{tr}(A^2)(d d')^{1/2} \), \( (d d')^{1/2} d A^2 d \), \( (d d')^{1/2} (d A d)^2 \) with

\[ (d d')^{1/2} \left[ \text{tr}(A^2) - (p^2 + p)/k_1 - (p^2 + p)/k_2 \right], \]

\[ (d d')^{1/2} \left\{ (d A d) - (p + 1)d d'/k_1 - (p + 1)d d'/k_2 \right\}, \]

\[ (d d')^{1/2} \left( (d A d)^2 - 2(d d')^2/k_1 - 2(d d')^2/k_2 \right), \]

respectively, we obtain the desired result.

\[ \square \]
4 Numerical study

In this section, we perform numerical study for investigating the performance of $D^*$ and other selection method. We assume that

$$d = (\sqrt{p}, \ldots, \sqrt{p})/p, \quad \Sigma_1 = I_p, \quad \Sigma_2 = I_p + C \cdot \text{diag}(1, \ldots, p)/p,$$

in numerical studies. The expected misclassification probabilities are calculated by the Monte Carlo simulation with 100,000 iterations.

4.1 Comparison of Cross-Validation method and $D^*$

We consider Cross-Validation (CV) as one of an estimation method of the expected misclassification probability. Let $X_{i1}, \ldots, X_{iN_i}$ be a training samples from $\Pi_i (i = 1, 2)$, and let $d(X, \mu_1, \mu_2, \Sigma_1, \Sigma_2)$ be arbitrary discriminant function which involves unknown parameters, then the expected misclassification probability $P_d$ is given as

$$P_d(X, \hat{X}_1, \hat{X}_2, S_1, S_2) = \frac{1}{2} \{ P(d(X, \hat{X}_1, \hat{X}_2, S_1, S_2) > 0 \mid X \in \Pi_1) + P(d(X, \hat{X}_1, \hat{X}_2, S_1, S_2) \leq 0 \mid X \in \Pi_2) \},$$

where estimators of these parameters are given by (1.1). Suppose that $\bar{X}_j^{(-k)}, S_j^{(-k)}$ are the sample mean and the sample covariance matrix by using training samples which are deleted $X_{jk}(k = 1, \ldots, N_j, \quad j = 1, 2)$, that is,

$$\bar{X}_j^{(-k)} = \frac{1}{N_j - 1} \sum_{i=1, i \neq k}^{N_j} X_{ji}, \quad S_j^{(-k)} = \frac{1}{N_j - 2} \sum_{i=1, i \neq k}^{N_j} (X_{ji} - \bar{X}_j)(X_{ji} - \bar{X}_j)' .$$

Then we obtain the CV estimation of $P_d$ as the following:

$$\hat{P}_d^{(CV)} = \frac{1}{2} \left\{ \hat{P}_d^{(CV)}(2|1) + \hat{P}_d^{(CV)}(1|2) \right\},$$

$$\hat{P}_d^{(CV)}(2|1) = \frac{1}{N_1} \sum_{k=1}^{N_1} \chi(d(X_{1k}, \bar{X}_1^{(-k)}, \bar{X}_2, S_1^{(-k)}, S_2) > 0),$$

$$\hat{P}_d^{(CV)}(1|2) = \frac{1}{N_2} \sum_{k=1}^{N_2} \chi(d(X_{2k}, \bar{X}_1, \bar{X}_2^{(-k)}, S_1, S_2^{(-k)}) \leq 0),$$

where $\chi(A)$ is defined as follows. If $A$ is true, then $\chi(A) = 1$. Otherwise, $\chi(A) = 0$. By using CV estimation, we obtain a criterion $D_{CV}$ for selecting between the linear and the quadratic discriminant functions, that is,

$$D_{CV} = \hat{P}_Q^{(CV)} - \hat{P}_L^{(CV)} .$$

Hence if $D_{CV}$ is negative, we can consider that $\tilde{Q}(X)$ is better than $\tilde{L}(X)$. Otherwise, we can consider that $\tilde{L}(X)$ is better than $\tilde{Q}(X)$. The asymptotic bias of CV estimation is 0, that is,

$$E[\hat{P}_d^{(CV)}] \to P_d \quad (n \to \infty).$$

However, CV estimation takes heavy costs of calculating. Hence it is hard to use it in practice.

Table 1 gives the expected misclassification probabilities for different $p$, $C$, $N_1$ and $N_2$. Here the columns $L$ and $Q$ are the expected misclassification probabilities by using only $\tilde{L}(X)$ and $\tilde{Q}(X)$, respectively. The columns $D^*$ and $D_{CV}$ are the expected misclassification probability by selecting between the linear and the quadratic discriminant functions with using $D^*$ and $D_{CV}$, respectively, and using the selected discriminant function. The numerical values on the double line and the line are the minimum and value without significance difference from the minimum in each row, respectively.

From Table 1, we can see that the performance of $D^*$ is better than $D_{CV}$ when $C$ is small, but the performance of $D^*$ is worse then $D_{CV}$ when $C$ is large.

4.2 Method of using hypothesis test and $D^*$

We performed the numerical study for investigating the performance of $D^*$ and $D_{CV}$ in the previous section. From the result of numerical study, performance of $D^*$ is worse than $D_{CV}$ when the difference between two covariance matrices is large. Because the setting of numerical study does not match the framework of asymptotic expansion.

8
The difference between two covariance matrices is large when \( P \) discriminant functions by using \( D \) difference between two covariance matrices is small. Hence we consider selecting between the linear and the quadratic discriminant functions. On the other hand, if \( 1 \) then if \( 2 \) quadratic discriminant functions as follows. Let \( f \) where \( f \) Moreover, we are given following result when the null hypothesis \( H_0 \) against \( H_1 : \Sigma_1 \neq \Sigma_2 \), the modified likelihood ratio test statistics is given as follows.

\[
T = -2 \rho \log \Lambda,
\]

where \( \rho = 1 - \frac{2p^2 + 3p - 1}{6(p+1)n} \left( \sum_{j=1}^{n} \frac{n_j}{n_j - 1} \right) \).

\[
\Lambda = \frac{|S_1|^{n_1}|S_2|^{n_2}}{|S|^{n}}.
\]

Moreover, we are given following result when the null hypothesis \( H_0 \) is true.

\[
P(T \leq x) = P(\chi^2_f \leq x) + O_2
\]

where \( f = p(p+1)/2 \). The proof of the above lemma can be seen in Muirhead (1982). We consider that the difference between two covariance matrices is large when \( P(\chi^2_f > T) < \alpha \), where \( \alpha \) is significance level and \( \chi^2_f \) is chi-square random variable with degree of freedom \( f \). Hence, we obtain a criterion for selecting between the linear and the quadratic discriminant functions as follows. Let

\[
D_H = P(\chi^2_f > T),
\]

then if \( D_H \) is lower than \( \alpha \), then we consider that the difference between two covariance matrices is large. Thus we select the quadratic discriminant function. On the other hand, if \( D_H \) is not lower than \( \alpha \), then we consider that the difference between two covariance matrices is small. Hence we consider selecting between the linear and the quadratic discriminant functions by using \( D^* \). So we suggest the following selection method \( D^*_H \).

**STEP 1.** We decide significance level \( \alpha \).
STEP 2. If $D_H < \alpha$, We select the quadratic discriminant function. Otherwise, we go next step.

STEP 3. We select between the linear and quadratic discriminant functions by $D^*$. We perform numerical study for comparison of performances of $D^*$, $D_{CV}$, $D_H$ and $D_H^*$ in the following. Here we obtain the selection method $D_H$ as follows. If $D_H$ is lower than $\alpha$, then we select the quadratic discriminant function. Otherwise, we select the linear discriminant function. Table 2 gives the expected misclassification probabilities for different $p$, $C$, $N_1$ and $N_2$. The column $D_H$ and $D_H^*$ are the expected misclassification probability for selecting between the linear and quadratic discriminant functions by using $D_H$ and $D_H^*$, respectively, and using the selected discriminant function. Here the number in the parentheses is the significance level.

| $p$ | $C$ | $N_1$ | $N_2$ | $L$ | $Q$ | $D^*$ | $D_{CV}$ | $D_H(0.01)$ | $D_H(0.05)$ | $D_H(0.1)$ | $D_H^*(0.01)$ | $D_H^*(0.05)$ | $D_H^*(0.1)$ |
|-----|-----|-------|-------|-----|-----|-------|----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2   | 0.0 | 25    | 25    | 0.3205 | 0.3328 | 0.3245 | 0.3251 | 0.3210 | 0.3221 | 0.3236 | 0.3246 | 0.3250 | 0.3258 |
| 0.1 | 25  | 25    | 0.3229 | 0.3364 | 0.3276 | 0.3235 | 0.3250 | 0.3265 | 0.3273 | 0.3280 | 0.3287 |
| 0.5 | 25  | 25    | 0.3316 | 0.3383 | 0.3348 | 0.3348 | 0.3326 | 0.3341 | 0.3351 | 0.3351 | 0.3357 | 0.3362 |
| 1.0 | 25  | 25    | 0.3428 | 0.3595 | 0.3408 | 0.3404 | 0.3432 | 0.3422 | 0.3408 | 0.3407 | 0.3401 | 0.3394 |
| 2.5 | 25  | 25    | 0.3747 | 0.4433 | 0.2717 | 0.2501 | 0.2455 | 0.2436 | 0.2435 | 0.2435 | 0.2434 | 0.2434 |

From Table 2, we can see that $D_H^*$ is better than $D^*$ and $D_{CV}$, and performance of $D_H^*$ is the same as performance of $D_H$ in the cases that $C$ is large. In most cases the difference between two covariance matrices is small, the performances of all selection method are about same. In the case of $C = 1$, $D_H^*$ better than $D_H$. Moreover, the performances $D_H$ and $D_H^*$ depend on signification level $\alpha$, but in the cases of $\alpha = 0.01, 0.05, 0.1$, the performances are very close.

5 Conclusion

As the first, We suggest the method $D^*$ for selecting of the linear and the quadratic discriminant functions by using the asymptotic expansion when the difference between two covariance matrices is small. In the case that the difference between two covariance matrices is small, $D^*$ is better then $D_{CV}$. However, $D^*$ is worse then $D_{CV}$ when the difference between two covariance matrices is large. So secondly, we suggest selection method $D_H^*$ of using the
hypothesis test in addition to $D^*$. We see performance of $D^*_H$ in the numerical studies. The performance of $D^*_H$ is better than performance other selection method, or equally. However from numerical studies, the performance of $D^*_H$ is worse than $D_{CV}$ when $p$ is large. Because the asymptotic expansions usually do not give good approximation formulae for large $p$. It may be possible to improve our selection method with using asymptotic expansions in high-dimensional and large samples framework, that is, both $N$, and $p$ become large, which is left for future work.

**Appendix**

**A.1 Proof of lemma 2.1**

*Proof.* Since $W \sim W_p(n, \Sigma)$,

$$Y'W^{-1}Y \sim Y'\Sigma^{-1/2}W_0^{-1}\Sigma^{-1/2}Y, \quad \log |W| \sim \log |\Sigma| \log |W_0|$$

where $W_0 \sim W_p(n, \Sigma)$. Let $W_0$ be partition as

$$W_0 = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix},$$

where $W_{11} : 1 \times 1$, and let

$$M = \begin{pmatrix} 1 & -W_{12}W_0^{-1} \\ 0 & I_{p-1} \end{pmatrix}.$$

Then we can obtain the following, (see e.g. Fujikoshi et al. (2010))

$$W_0^{-1} = W_0 = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & W_{22}^{-1} \end{pmatrix} + \begin{pmatrix} -W_{22}^{-1}W_{21} \\ I_{11} \end{pmatrix}W_{11,2}^{-1} \begin{pmatrix} 1 & -W_{12}W_2^{-1} \end{pmatrix},$$

$$|W_0| = |MW_0M'| = \begin{vmatrix} W_{11,2} & O \\ O & W_{22} \end{vmatrix} = W_{11,2}|W_{22}|,$$

where $W_{11,2} = W_{11} - W_{12}W_{22}W_{21}$. Moreover, since $W_0 \sim W_p(n, \Sigma)$, $W_{11,2}$ is distributed as $W_1(n - p + 1, 1)$, that is the chi-square distribution with degree of freedom $n - p + 1$, and $W_{22} \sim W_{p-1}(n, I_{p-1})$, $W_{11,2}$ is independent of $W_{22}$. Let $H = (H_1, H_2)$ be the orthogonal matrix such that $H_1 = \Sigma^{-1/2}Y$, then

$$Y'\Sigma^{-1/2}W_0^{-1}\Sigma^{-1/2}Y \sim Y'\Sigma^{-1/2}HW_0^{-1}H'\Sigma^{-1/2}Y$$

$$= W_{11,2}^{-1}Y'\Sigma^{-1/2}Y.$$

Therefore, we obtain the desired result. $\square$

**A.2 Proof of lemma 2.2**

*Proof.*

$$E[\exp\{itZ^2\}] = \int_{-\infty}^{\infty} \exp\{itz^2\} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(z - \mu)^2\right\} dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{itz^2 - \frac{1}{2}(z^2 - 2z\mu) + \mu^2\right\} dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}((1 - 2it)z^2 - 2z\mu + \mu^2)\right\} dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(1 - 2it)\left(z - \frac{\mu}{1 - 2it}\right)^2 - \frac{it\mu^2}{(1 - 2it)}\right\} dz$$

$$= (1 - 2it)^{-1/2} \exp\left\{-\frac{it\mu^2}{(1 - 2it)}\right\}.$$

$\square$
A.3 Proof of Lemma 2.3

Proof.

\[ g(t_1, t_2) = E[\exp \{ X'(t_1 \Gamma_1 + t_2 \Gamma_2) X + 2(t_1 \eta'_1 \Gamma_1 + t_2 \eta'_2 \Gamma_2) x + t_1 \eta'_1 \Gamma_1 \eta_1 + t_2 \eta'_2 \Gamma_2 \eta_2] \]

\[ = \int_{-\infty}^{\infty} (2\pi)^{-p/2} \exp \left\{ -x'x/2 \right\} \exp \{ x' (t_1 \eta'_1 \Gamma_1 + t_2 \eta'_2 \Gamma_2) x + t_1 \eta'_1 \Gamma_1 \eta_1 + t_2 \eta'_2 \Gamma_2 \eta_2 \} \, dx \]

\[ = \int_{-\infty}^{\infty} (2\pi)^{-p/2} \exp \left\{ \frac{1}{2} \left\{ x' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) x - 4(t_1 \eta'_1 \Gamma_1 + t_2 \eta'_2 \Gamma_2) x - 2t_1 \eta'_1 \Gamma_1 \eta_1 - 2t_2 \eta'_2 \Gamma_2 \eta_2 \right\} \right\} \, dx \]

\[ = \int_{-\infty}^{\infty} (2\pi)^{-p/2} \exp \left\{ \frac{1}{2} \left\{ x - 2(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2) \right\}' \right\} \, dx \]

\[ + 2(t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2)' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2) \]

\[ + t_1 \eta'_1 \Gamma_1 \eta_1 + t_2 \eta'_2 \Gamma_2 \eta_2] \exp \left\{ 2(t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2)' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2) \right\} \]

From the above result, we can easily obtain \( g(t_1, t_2; \eta_1, \eta_2, \Gamma_1, \Gamma_2) = g(t_2, t_1; \eta_2, \eta_1, \Gamma_2, \Gamma_1) \). By using \( t_1 \eta'_1 \Gamma_1 \eta_1 + t_2 \eta'_2 \Gamma_2 \eta_2 + 2(t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2)' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2) \),

\[ t_1 \eta'_1 \Gamma_1 \eta_1 + t_2 \eta'_2 \Gamma_2 \eta_2 + 2(t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2)' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (t_1 \Gamma_1 \eta_1 + t_2 \Gamma_2 \eta_2) \]

\[ = \frac{1}{2} \eta'_1 (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 + \frac{1}{2} \eta'_2 (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_2 \]

\[ + \frac{1}{2} \left\{ (\eta_1 - 2t_2 \Gamma_2 (\eta_1 - \eta_2)) - (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 \right\}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 \]

\[ + \frac{1}{2} \left\{ (\eta_1 - 2t_2 \Gamma_2 (\eta_1 - \eta_2)) - (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 \right\}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_2 \]

\[ = -\frac{1}{2} \eta'_1 \eta_1 + t_1 (\eta_1 - \eta_2)' \Gamma_2 (\eta_1 - \eta_2) \]

\[ + \frac{1}{2} \left\{ (\eta_1 - 2t_2 \Gamma_2 (\eta_1 - \eta_2)) - (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 \right\}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2) \eta_1 \]

Let \( A, B, C \) and \( M \) be arbitrary \( p \times p \) matrix, then we obtain the following formulae.

\[ |I_p + hC| = E[|I_p + hC|] = E[|I_p + h(C + O(h^2))|] = 1 + htr(C) + O(h^2), \]

\[ |A + (t + h)B| = |A + tB + hB| = |A + tB| + h(A + B)^{-1}B \]

\[ = |A + tB| \left[ 1 + htr((A + B)^{-1}B) + O(h^2) \right], \]

\[ \text{tr} \left\{ M(A + (t + h)B)^{-1} \right\} = \text{tr} \left\{ M(A + tB)^{-1} \right\} \left( I_p + hB(A + tB)^{-1} \right)^{-1} \]

\[ = \text{tr} \left\{ M(A + tB)^{-1} \right\} - htr \left\{ M(A + tB)^{-1}B(A + tB)^{-1} \right\} + O(h^2). \]

From (A.3.1) and (A.3.2),

\[ \frac{d}{dt} |A + tB| = \lim_{h \to 0} \frac{1}{h} \left[ |A + (t + h)B| - |A + tB| \right] = |A + tB| \text{tr} \left\{ (A + tB)^{-1}B \right\}, \]

\[ \frac{d}{dt} \text{tr} \left\{ M(A + tB)^{-1} \right\} = \lim_{h \to 0} \frac{1}{h} \left[ \text{tr} \left\{ M(A + (t + h)B)^{-1} \right\} - \text{tr} \left\{ M(A + tB)^{-1} \right\} \right] \]

\[ = -\text{tr} \left\{ M(A + tB)^{-1}B(A + tB)^{-1} \right\}. \]
By using these results,

\[
\frac{\partial g(t_1, t_2)}{\partial t_1} = \left[ \frac{\partial}{\partial t_1} [I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2]^{-1/2} \right] \exp \left\{ -\frac{1}{2} \eta_i' \eta_i + t_2 (\eta_1 - \eta_2) (\eta_1 - \eta_2) \right\} \\
+ \frac{1}{2} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \} \\
+ \frac{1}{2} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \} \\
\cdot \frac{\partial}{\partial t_1} \left\{ -\frac{1}{2} \eta_i' \eta_i + t_2 (\eta_1 - \eta_2) \Gamma_2 (\eta_1 - \eta_2) \\
+ \frac{1}{2} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \} \right\}
\]

\[
= g(t_1, t_2) \text{tr}[(I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \Gamma_1] \\
+ \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \Gamma_1 (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \},
\]

\[
\frac{\partial L(t_1, t_2)}{\partial t_1} = L(t_1, t_2; \eta_1, \eta_2, \Gamma_1, \Gamma_2) = 2 \text{tr}\left[ (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \Gamma_1 \right] \right]^2 \\
+ 4 \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}' (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \Gamma_1 (I_p - 2t_1 \Gamma_1 - 2t_2 \Gamma_2)^{-1} \{ \eta_i - 2t_2 \Gamma_2 (\eta_1 - \eta_2) \}
\]

\[
\frac{\partial g(t_1, t_2)}{\partial t_2} = g(t_1, t_2) L(t_2, t_1; \eta_2, \eta_1, \Gamma_2, \Gamma_1).
\]

\[\square\]

References

[1] Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edision. Wiley, Hoboken, NJ.

[2] Fujikoshi, Y., Ulyanv, V. V. and Shimizu, R. (2010), *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, NJ.

[3] Marks, S. and Dunn, O. J. (1974). Discriminant function when covariance matrices are unequal. *J. Am. Statist. Assoc.*, 69, 555-559.

[4] Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[5] Wahl, P. W. and Kronmal, R. A. (1977). Discriminant function when covariance matrices are unequal and sample sizes are moderate. *Biometrics*, 33, 479-484.

[6] Wakaki, H. (1990). Comparison of linear and quadratic discriminant functions. *Biometrika*, 77, 227-229.