HIROSHIMA MATH. J.
26 (1996), 593–603

Asymptotic expansions for the best linear discriminant functions

Hirofumi WAKAKI

(Recieved June 2, 1995)

ABSTRACT. In the discriminant problem between two elliptical populations with different covariance matrices, we derive the asymptotic expansions of the expected misclassification probabilities of the sample best linear discriminant function. Using this result, we construct an estimator of the expected misclassification probabilities which are asymptotically less biased than the usual estimates. Asymptotic expansions of the bias of the estimated discriminant coefficients and the cut-off point are also derived.

1. Introduction

Consider the problem of classifying an observation X into one of two populations $\Pi_1 : E_p(\mu_1, \Gamma_1; h)$ and $\Pi_2 : E_p(\mu_2, \Gamma_2; h)$, where $E_p(\mu, \Gamma; h)$ is a p-dimensional elliptical distribution with probability density function

$$(1.1) \qquad |\Gamma|^{-1/2} h\{(x-\mu)'\Gamma^{-1}(x-\mu)\},$$

where h is a decreasing function, μ is a $p \times 1$ parameter vector and Γ is a positive definite $p \times p$ matrix. We assume that $E_p(\mu, \Gamma; h)$ has the covariance matrix. Then Π_j $(j = 1, 2)$ has the covariance matrix $\Sigma_j = \omega\Gamma_j$, where the constant ω is given by the characteristic function of $E_p(\mu, \Gamma; h)$ (cf. Kelker [2]). Considering some modification of h, we can assume without loss of generality that the constant $\omega = 1$.

In the case of normal populations, Anderson and Bahadur [1] derived the class of admissible linear discriminant procedures. The admissible linear discriminant rule is that an observation is classified into Π_1 if $X'b < c$, where b is a $p \times 1$ vector and c is a scalar given by

$$(1.2) \qquad b = (k_1\Sigma_1 + k_2\Sigma_2)^{-1}(\mu_2 - \mu_1), \qquad c = b'\mu_1 + k_1 b'\Sigma_1 b$$

and k_1 and k_2 are scalars chosen such that $k_1\Sigma_1 + k_2\Sigma_2$ is positive definite.

1991 *Mathematics Subject Classification.* 62H30, 62E20.

Key words and phrases. Best linear discrimination, asymptotic expansion, elliptical populations, different covariance matrices.

Let

$$y_j = k_j(b'\Sigma_j b)^{1/2}, \tag{1.3}$$

and P_j be the probability that an observation from Π_j is misclassified. Then $P_j = 1 - \Phi(y_j)$ where Φ is the standard normal distribution function (for details see [1]).

It is easily shown that the above rule is admissible linear for our problem. Let Ψ be the univariate marginal distribution function of $E_p(0, I; h)$, the spherical population. Then $P_j = 1 - \Psi(y_j)$. The expression of the probability density function of Ψ with using h is given by Wakaki [5].

Anderson and Bahadur [1] described three ways to determine k_1 and k_2. These methods are (1) *minimization of one misclassification probability for a specified probability of the other*, (2) *the minimax procedure* and (3) *minimization of the total misclassification probability for given a priori probabilities.* In this paper we deal with the minimax procedure since the minimax linear procedure is determined by y_1 and y_2 and does not depend on Ψ. On the other hand, the other two procedures depend on Ψ. Even for the normal case, the Bayes linear procedure (3) above is generally not unique, and sometimes one of the misclassification probabilities P_1 and P_2 becomes very large.

It is shown that the minimax linear discriminant rule is one of the above admissible rules with k_1 and k_2 which solves the equation $y_1 = y_2$. Since y_1 and y_2 are homogeneous of degree 0 in k_1 and k_2 and both scalars are positive, we can put $k_1 = k$ and $k_2 = 1 - k$ $(0 < k < 1)$.

When population parameters are unknown, these parameters should be estimated based on two samples of sizes n_j, one from each population Π_j. In this paper we use the usual sample mean vector $\overline{X}_j$ and the sample covariance matrix S_j. The sample minimax linear discriminant rule states that an observation is classified into Π_1 if $X'B < C$, where B is a $p \times 1$ vector and C is a scalar given by

$$B = (K_1 S_1 + K_2 S_2)^{-1}(\overline{X}_2 - \overline{X}_1), \qquad C = B'\overline{X}_1 + K_1 B' S_1 B, \tag{1.4}$$

with $K_1 = K$ and $K_2 = 1 - K$. Here, the scalar K is determined as follows.

Let Y_1 and Y_2 be functions of K and the sample estimates given by

$$Y_j = K_j(B' S_j B)^{1/2} \qquad (j = 1, 2). \tag{1.5}$$

Then K is a solution of the equation $Y_1^2 = Y_2^2$, with $(0 < K < 1)$.

When the training samples are given, each conditional misclassification probability P_j is given by $1 - \Psi(Z_j)$, where

$$Z_1 = (C - B'\mu_1)(B'\Sigma_1 B)^{-1/2} \qquad \text{and} \qquad Z_2 = (B'\mu_2 - C)(B'\Sigma_2 B)^{-1/2}. \tag{1.6}$$

A simple estimator of the expected misclassification probability is $1 - \Psi(Y_j)$, where the expectation is taken for the distribution of $\bar{X}_1, \bar{X}_2, S_1$ and S_2. In the following section, we see that $1 - \Psi(Y_j)$ is a biased estimator.

For the Fisher's linear discriminant function, McLachlan [4] modified the usual estimator of the expected misclassification probability, using asymptotic expansions. The resulted estimator has the bias of order $O(n^{-2})$, where $n = n_1 + n_2$. The purpose of this paper is to derive the asymptotic expansion of the expected misclassification probability of the sample minimax linear discriminant rule, and to construct it's estimator which is unbiased up to the order n^{-1}.

2. Expectation of $\Psi(Y_j)$

Before expanding the expected misclassification probability, we first derive the asymptotic expansion of the expectation of $\Psi(Y_j)$. Because of the invariance of Y_j under a group of affine transformations, we can without loss of generality assume that the following conditions hold true.

$$(2.1) \qquad \mu_1 + \mu_2 = 0, \qquad k_1\Sigma_1 + k_2\Sigma_2 = I,$$

and Σ_1 and Σ_2 are diagonal.

Let

$$(2.2) \qquad S_j = \Sigma_j + V_j \quad (j = 1, 2) \qquad \text{and} \qquad \bar{X}_2 - \bar{X}_1 = \delta + D,$$

where $\delta = \mu_2 - \mu_1$. Note that the coefficient vector of the minimax linear discriminant function is δ. Then V_j and D are $O_p(n^{-1/2})$ and the limiting joint distribution of $n^{1/2}V_j$ and $n^{1/2}D$ is independent normal. The basic expectation formulas are as follows. Let Ξ be any $p \times p$ symmetric matrix and α, ζ, η and ξ be any $p \times 1$ vector. Then

$$(2.3) \qquad \begin{aligned} \mathrm{E}[D'\Xi D] &= n_1^{-1}\mathrm{tr}(\Xi\Sigma_1) + n_2^{-1}\mathrm{tr}(\Xi\Sigma_2), \\ \mathrm{E}[(D'\alpha)(D'\zeta)] &= n_1^{-1}\alpha'\Sigma_1\zeta + n_2^{-1}\alpha'\Sigma_2\zeta, \\ \mathrm{E}[\alpha' V_j \Xi V_j \zeta] &= n_j^{-1}\{(2\kappa + 1)\alpha'\Sigma_j\Xi\Sigma_j\zeta + (\kappa + 1)\alpha'\Sigma_j\zeta\,\mathrm{tr}(\Xi\Sigma_j)\}, \\ \mathrm{E}[(\alpha' V_j\zeta)(\eta' V_j\xi)] &= n_j^{-1}\{\kappa(\alpha'\Sigma_j\zeta)(\eta'\Sigma_j\xi) + (\kappa + 1)(\alpha'\Sigma_j\eta)(\zeta'\Sigma_j\xi) \\ &\qquad + (\kappa + 1)(\alpha'\Sigma_j\xi)(\eta'\Sigma_j\zeta)\}, \end{aligned}$$

where κ is the kurtosis parameter.

Let

$$(2.4) \qquad K = k + K_f + K_s,$$

where $K_f = O_p(n^{-1/2})$ and $K_s = O_p(n^{-1})$. Considering the Taylor expansion of $\frac{1}{2}Y_j^2$ at $K=k$, $S_1=\Sigma_1$, $S_2=\Sigma_2$ and $\bar{X}_2-\bar{X}_1=\delta$, we obtain

$$\begin{aligned}\text{(2.5)}\qquad \tfrac{1}{2}Y_j^2 &= \tfrac{1}{2}k_j^2\beta_j + \{s_jk_j\beta_{12}K_f + F_j\}\\ &\quad + \{s_jk_jK_s + (\tfrac{3}{2}\beta_{\langle j\rangle 12} - \beta_{12})K_f^2 + G_jK_f + H_j\}\\ &\quad + \mathrm{O_p}(n^{-3/2}) \qquad (j=1,2),\end{aligned}$$

where $s_1=1$, $s_2=-1$, the notation $\langle j\rangle$ means that $\langle 1\rangle=2$ and $\langle 2\rangle=1$,

$$\text{(2.6)}\qquad \beta_{ijk\cdots l} = \delta'\Sigma_i\Sigma_j\Sigma_k\cdots\Sigma_l\delta \qquad (i,j,k,\cdots,l=1 \text{ or } 2),$$

and

$$\text{(2.7)}\qquad \begin{aligned} F_j &= k_j^2\{\delta'\Sigma_jD - k_{\langle j\rangle}\delta'\Sigma_jV_{\langle j\rangle}\delta,\\ G_j &= s_jk_j\textstyle\sum_{i=1}^{2}\{\delta'(\Sigma_{\langle i\rangle} - 2k_i\Sigma_1\Sigma_2)V_i\delta - k_i\delta'\Sigma_1V_i\Sigma_2\delta + \delta'\Sigma_1\Sigma_2D\},\\ H_j &= \tfrac{1}{2}k_j^2\{D'\Sigma_jD + k_j\delta'(I-2k_{\langle j\rangle}\Sigma_{\langle j\rangle})V_j^2\delta - k_1k_2\delta'V_j\Sigma_{\langle j\rangle}V_j\delta\\ &\quad + k_{\langle j\rangle}^2\delta'V_{\langle j\rangle}\Sigma_jV_{\langle j\rangle}\delta + 2k_{\langle j\rangle}^2\delta'\Sigma_jV_{\langle j\rangle}^2\delta\} + (\text{Remainder}).\end{aligned}$$

Here (Remainder) means the remainder term whose expectation is 0.

Since $\frac{1}{2}Y_1^2 - \frac{1}{2}Y_2^2 = 0$, from each terms of $\mathrm{O_p}(n^{-1/2})$ and the ones of $\mathrm{O_p}(n^{-1})$ we obtain

$$\text{(2.8)}\qquad \begin{aligned} K_f &= \beta_{12}^{-1}(F_2 - F_1),\\ K_s &= \beta_{12}^{-1}\{\tfrac{3}{2}(\beta_{112}-\beta_{122})K_f^2 + (G_2-G_1)K_f + H_2 - H_1\}.\end{aligned}$$

Substituting these equations in (2.5) gives the expansion of $\frac{1}{2}Y_j^2$ as

$$\text{(2.9)}\qquad \tfrac{1}{2}Y^2 = \tfrac{1}{2}Y_1^2 = \tfrac{1}{2}Y_2^2 = \tfrac{1}{2}y^2 + Y_f + Y_s + \mathrm{O_p}(n^{-3/2}),$$

where $y = y_1 = y_2$,

$$\text{(2.10)}\qquad \begin{aligned} Y_f &= k_1k_2\Big(\delta'D - \textstyle\sum_{i=1}^{2}\tfrac{1}{2}k_i\delta'V_i\delta\Big),\\ Y_s &= \tfrac{1}{2}\beta_{12}^{-1}\Big[\{\delta'(k_1I - k_2\Sigma_2)D\}^2 + \textstyle\sum_{i=1}^{2}\{\delta'(\tfrac{1}{2}k_i^2I - k_1k_2\Sigma_{\langle i\rangle})V_i\delta\}^2\Big]\\ &\quad + \tfrac{1}{2}k_1k_2\Big(D'D + \textstyle\sum_{i=1}^{2}k_i^2\delta'V_i^2\delta\Big) + (\text{Remainder}).\end{aligned}$$

A Taylor expansion of $\Psi(Y)$ gives

$$\begin{aligned}\text{(2.11)}\qquad \Psi(Y) &= \Psi(y) + \psi(y)y^{-1}Y_f\\ &\quad + [\psi(y)y^{-1}Y_s + \{\tfrac{1}{2}\psi'(y)y^{-2} - \tfrac{1}{8}\psi(y)y^{-3}\}Y_f^2],\end{aligned}$$

where ψ is the density function of Ψ and ψ' is its derivative. Note that we neglect the term of $O_p(n^{-3/2})$ at the right-hand side of (2.11). In the rest of this paper we neglect the terms of $O_p(n^{-3/2})$ and $O(n^{-3/2})$ in equations of asymptotic expansions. Using the formulas given by (2.3), the expectation of $\Psi(Y)$ can be expanded as

$$\mathrm{E}[\Psi(Y)] = \Psi(y) + \psi(y)y^{-1}\mathrm{E}[Y_s] + \{\tfrac{1}{2}\psi'(y)y^{-2} - \tfrac{1}{8}\psi(y)y^{-3}\}\mathrm{E}[Y_f^2], \tag{2.12}$$

where

$$\begin{aligned}\mathrm{E}[Y_s] = \sum\nolimits_{i=1}^{2} n_i^{-1}[\tfrac{1}{2}\{\gamma_i\beta_i k_i^3(\kappa+1) + \gamma_i k_i k_{\langle i\rangle} + \beta_i k_i^2 k_{\langle i\rangle}^2(3\kappa+2) \\ + k_{\langle i\rangle}(1-2k_1)\} + \tfrac{1}{2}\beta_{12}^{-1}\{\tfrac{1}{4}\beta_i^2 k_j^4(3\kappa+2) + \beta_i k_i^2 k_{\langle i\rangle}^2(3\kappa+2) \\ - \beta_i\beta_{i12}k_j^3 k_{\langle i\rangle}(\kappa+1) - \beta_{i12}k_i k_{\langle i\rangle}\}],\end{aligned} \tag{2.13}$$

$$\mathrm{E}[Y_f^2] = k_i^2 k_{\langle i\rangle}^2 \sum\nolimits_{i=1}^{2} n_i^{-1}\{\tfrac{1}{4}\beta_i^2 k_i^2(3\kappa+2) + \beta_i\}. \tag{2.14}$$

Here we used the notation

$$\gamma_{ijk\cdots l} = \mathrm{tr}(\Sigma_i\Sigma_j\Sigma_k\cdots\Sigma_l). \tag{2.15}$$

3. Asymptotic expansion of the expected misclassification probabilities

In this section we derive the asymptotic expansion of $\mathrm{E}[\Psi(Z_j)]$ $(j=1,2)$, where Z_j is given by (1.6). If we replace μ_1, μ_2, Σ_1 and Σ_2 with $\bar{X}_1, \bar{X}_2, S_1$ and S_2 then both Z_1 and Z_2 become Y. Therefore the difference between Z_j and Y is $O_p(n^{-1/2})$. Let

$$\bar{X}_j = \mu_j + D_j \qquad (j=1,2). \tag{3.1}$$

Then we can expand Z_j as

$$Z_j = y + Z_{fj} + Z_{sj}, \tag{3.2}$$

where

$$\begin{aligned} Z_{fj} &= y^{-1}k_j\sum\nolimits_{i=1}^{2} s_i(k_i\delta' D_i + \tfrac{1}{2}k_i^2\delta' V_i\delta), \\ Z_{sj} &= y^{-1}Y_s - \tfrac{1}{8}y^{-3}Y_f^2 + W_{sj}, \end{aligned} \tag{3.3}$$

with

$$\begin{aligned} W_{sj} = \tfrac{1}{8}y^{-3}k_j^4(7 - 4k_j + 2k_{\langle j\rangle})(\delta' V_j\delta)^2 + y^{-1}\{k_j^3\delta' V_j^2\delta - k_j D_j' D_j \\ + k_{\langle j\rangle}\beta_j^{-1}(\delta' D_j)^2\} + y^{-1}\beta_{12}^{-1}\{k_j^2\delta'\Sigma_{\langle j\rangle}V_j\delta\delta'(\tfrac{1}{2}k_j I - k_{\langle j\rangle}\Sigma_{\langle j\rangle})V_j\delta \\ + \delta'\Sigma_{\langle j\rangle}D_j\delta'(k_j I - k_{\langle j\rangle}\Sigma_{\langle j\rangle})D_j\}. \end{aligned} \tag{3.4}$$

A Taylor expansion of $\Psi(Z_j)$ gives that

$$\Psi(Z_j) = \Psi(y) + \psi(y)y^{-1}Z_{fj} + \{\psi(y)(y^{-1}Y_s - \tfrac{1}{8}y^{-3}Y_f^2 + W_{sj}) + \tfrac{1}{2}\psi'(y)Z_{fj}^2\}. \tag{3.5}$$

Comparing (3.5) with (2.11), we obtain the asymptotic expansion of the expected misclassification probability as

$$\mathrm{E}[\Psi(Z_j)] = \mathrm{E}[\Psi(Y)] + \psi(y)\,\mathrm{E}[W_{sj}] + \tfrac{1}{2}\psi'(y)\,\mathrm{E}[Z_{fj}^2 - y^{-2}Y_f^2], \tag{3.6}$$

where

$$\begin{aligned} \mathrm{E}[W_{sj}] &= n_j^{-1}[y^{-1}\{-\gamma_j k_j - 2\beta_{12}k_j^2 k_{\langle j\rangle}(2\kappa+1) + \beta_{12}^{-1}\beta_{j12}k_j\} \\ &\quad + y\{\gamma_j k_j(\kappa+1) - \tfrac{1}{4}k_{\langle j\rangle}(3\kappa+2) + \tfrac{1}{8}(29\kappa+14) \\ &\quad + \beta_{12}^{-1}\beta_{j12}k_j(\kappa+1)\}], \\ \mathrm{E}[Z_{fj}^2 - y^{-2}Y_f^2] &= n_j^{-1}(k_{\langle j\rangle} - k_j)\{1 + \tfrac{1}{4}y^2(3\kappa+2)\}. \end{aligned} \tag{3.7}$$

The equation (3.6) gives the bias of $\Psi(Y)$ as an estimator of the expected misclassification probability. Replacing the unknown parameters with those consistent estimators, we can modify the bias of $\Psi(Y)$ as in the following theorem.

THEOREM 3.1. *Let*

$$\hat{P}_j = 1 - \Psi(Y_j) + n^{-1}\{\psi(Y_j)A_j + \tfrac{1}{2}\psi'(Y_j)B_j\} \qquad (j = 1, 2), \tag{3.8}$$

where Y_j *is given by* (1.5),

$$\begin{aligned} A_j &= Y_j^{-1}\{-\hat{\gamma}_j K_j - 2\hat{\beta}_{12}K_j^2 k\langle j\rangle(2\hat{\kappa}+1) + \hat{\beta}_{12}^{-1}\hat{\beta}_{j12}K_j\} \\ &\quad + Y_j\{\hat{\gamma}_j K_j(\hat{\kappa}+1) - \tfrac{1}{4}k_{\langle j\rangle}(3\hat{\kappa}+2) + \tfrac{1}{8}(29\hat{\kappa}+14) + \hat{\beta}_{12}^{-1}\hat{\beta}_{j12}K_j(\hat{\kappa}+1)\}, \\ B_j &= (K_{\langle j\rangle} - K_j)\{1 + \tfrac{1}{4}Y_j^2(3\hat{\kappa}+2)\}, \end{aligned} \tag{3.9}$$

with a consistent estimator $\hat{\kappa}$ *of kurtosis parameter and*

$$\begin{aligned} \hat{\gamma}_1 &= \mathrm{tr}\{(K_1 I + K_2 S_2 S_1^{-1})^{-1}\}, \\ \hat{\gamma}_2 &= \mathrm{tr}\{(K_1 I + K_2 S_2 S_1^{-1})^{-1} S_2 S_1^{-1}\}, \\ \hat{\beta}_{12} &= B'(K_1 I + K_2 S_2 S_1^{-1})^{-2} S_2 S_1^{-1}(\bar{X}_2 - \bar{X}_1), \\ \hat{\beta}_{112} &= B'(K_1 I + K_2 S_2 S_1^{-1})^{-3} S_2 S_1^{-1}(\bar{X}_2 - \bar{X}_1), \\ \hat{\beta}_{212} &= B'(K_1 I + K_2 S_2 S_1^{-1})^{-3}(S_2 S_1^{-1})^2(\bar{X}_2 - \bar{X}_1). \end{aligned} \tag{3.10}$$

Then $\hat{P}_j$ *has the bias of order* $O(n^{-3/2})$ *as an estimator of the expected misclassification probability* $E[1-\Psi(Z_j)]$.

Note that we made the assumption (2.1) for the population parameters in the derivation of asymptotic expansions. But in the above theorem, we do not make assumptions as in (2.1). Consistent estimators (3.10) of γ's and β's are given as follows.

Let μ_1, μ_2, Σ_1 and Σ_2 be the original parameters. Then the transformed parameters μ_{1*}, μ_{2*}, Σ_{1*} and Σ_{2*} are

$$\begin{aligned} &\Sigma_{1*} = (k_1 I + k_2 \Lambda)^{-1}, \qquad \Sigma_{2*} = (k_1 I + k_2 \Lambda)^{-1}\Lambda \qquad \text{and} \\ &\mu_{1*} = -\mu_{2*} = (k_1 I + k_2 \Lambda)^{-1/2} H \Sigma_1^{-1/2}(\mu_1 - \mu_2)/2, \end{aligned} \tag{3.11}$$

where Λ is a diagonal matrix whose diagonal elements are the latent roots of $\Sigma_1^{-1}\Sigma_2$ and H is an orthogonal matrix such that $H'\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2}H = \Lambda$. Therefore $\beta_{ijk\cdots l}$ in (2.6) and $\gamma_{ijk\cdots l}$ in (2.14) are written with the original parameters as

$$\begin{aligned} \beta_{11\cdots122\cdots2} &= \delta'(\Sigma_{1*})^m(\Sigma_{2*})^n\delta \\ &= (\mu_2 - \mu_1)'(k_1\Sigma_1 + k_2\Sigma_2)^{-1}(k_1 I + k_2\Sigma_2\Sigma_1^{-1})^{-(m+n)} \\ &\quad \times (\Sigma_2\Sigma_1^{-1})^n(\mu_2 - \mu_1), \\ \gamma_{11\cdots122\cdots2} &= \mathrm{tr}\{(\Sigma_{1*})^m(\Sigma_{2*})^n\} \\ &= \mathrm{tr}\{(k_1 I + k_2\Sigma_2\Sigma_1^{-1})^{-(m+n)}(\Sigma_2\Sigma_1^{-1})^n\}. \end{aligned} \tag{3.12}$$

Replacing the original parameters with those sample estimators, we obtain (3.10).

4. Bias of the estimated discriminant function

In this section we derive asymptotic expansions of the bias of the estimated discriminant coefficient vector B and cut-off point C. The coefficient vector B can be expanded as

$$B = \delta + B_f + B_s, \tag{4.1}$$

where

$$\begin{aligned} B_f &= D - \{k_1 V_1 + k_2 V_2 + K_f(\Sigma_1 - \Sigma_2)\}\delta, \\ B_s &= \{k_1 V_1 + k_2 V_2 + K_f(\Sigma_1 - \Sigma_2)\}^2\delta - \{K_s(\Sigma_1 - \Sigma_2) + K_f(V_1 - V_2)\}\delta \\ &\quad - \{k_1 V_1 + k_2 V_2 + K_f(\Sigma_1 - \Sigma_2)\}D. \end{aligned} \tag{4.2}$$

Substituting (2.8) with (2.7) in (4.2) and taking the expectation by using (2.3), we can obtain the asymptotic bias of B as in the following theorem.

THEOREM 4.1. *The expectation of B can be expanded as*

$$\mathrm{E}[B] = \delta + \sum_{j=1}^{2} n_j^{-1}(b_{0j} + b_{1j}\Sigma_j + b_{2j}\Sigma_1\Sigma_2 + b_{3j}\Sigma_j\Sigma_1\Sigma_2)\delta, \tag{4.3}$$

where

$$\begin{aligned} b_{0j} = {} & \tfrac{1}{8}\beta_{12}^{-3}k_{\langle j\rangle}^{-2}\{3\beta_j^2\beta_{j12}k_j^4(3\kappa+2) - 12\beta_j\beta_{j12}^2k_j^3k_{\langle j\rangle}(\kappa+1) \\ & + 12\beta_j\beta_{j12}k_j^2 - 12\beta_{j12}^2k_jk_{\langle j\rangle}\} - \tfrac{1}{8}\beta_{12}^{-2}k_{\langle j\rangle}^{-2}\{\beta_j^2k_j^4(3\kappa+2) \\ & + 4\beta_j\beta_{j12}k_j^2(1-3k_j)k_{\langle j\rangle}(\kappa+1) + 16\beta_j\beta_{1122}k_j^2k_{\langle j\rangle}^2(\kappa+1) + 4\beta_jk_j^2 \\ & + 4\beta_{j12}(1-3k_j)k_{\langle j\rangle} + 16\beta_{1122}k_{\langle j\rangle}^2\} - \tfrac{1}{2}\beta_{12}^{-1}k_{\langle j\rangle}^{-2}\{\gamma_j\beta_jk_j^3k_{\langle j\rangle}(\kappa+1) \\ & + \gamma_jk_jk_{\langle j\rangle} - \gamma_{12}\beta_jk_{\langle j\rangle}^2(\kappa+1) - \gamma_{12}k_{\langle j\rangle}^2 + \beta_jk_j^2([4-3k_j]\kappa + 2k_{\langle j\rangle})k_{\langle j\rangle} \\ & + \beta_{j12}([4k_j^2-2]\kappa + 3k_j^2 - 1)k_{\langle j\rangle}^2 + (1-2k_j)k_{\langle j\rangle}\} + \gamma_jk_j^2(\kappa+1) \\ & + k_j([3-k_j]\kappa + 2 - k_j), \end{aligned} \tag{4.4}$$

$$\begin{aligned} b_{1j} = {} & -\tfrac{1}{8}\beta_{12}^{-3}k_{\langle j\rangle}^{-2}\{3\beta_j^2\beta_{j12}k_j^4(3\kappa+2) - 12\beta_j\beta_{j12}^2k_j^3k_{\langle j\rangle}(\kappa+1) \\ & + 12\beta_j\beta_{j12}k_j^2 - 12\beta_{j12}^2k_jk_{\langle j\rangle}\} + \tfrac{1}{8}\beta_{12}^{-2}k_{\langle j\rangle}^{-2}\{\beta_j^2k_j^3(2-k_j)(3\kappa+2) \\ & + 4\beta_j\beta_{j12}k_j^2(3k_j-5)k_{\langle j\rangle}(\kappa+1) + 16\beta_j\beta_{1122}k_j^2k_{\langle j\rangle}^2(\kappa+1) \\ & + 4\beta_jk_j(-k_j+2) + 4\beta_{j12}(3k_j-5)k_{\langle j\rangle} + 16\beta_{1122}k_{\langle j\rangle}^{-2}\} \\ & + \tfrac{1}{2}\beta_{12}^{-1}k_j^{-1}k_{\langle j\rangle}^{-2}\{\gamma_j\beta_jk_j^4k_{\langle j\rangle}(\kappa+1) + \gamma_jk_j^2k_{\langle j\rangle} - \gamma_{12}\beta_jk_jk_{\langle j\rangle}^2(\kappa+1) \\ & - \gamma_{12}k_jk_{\langle j\rangle}^2 + \beta_jk_j^2k_{\langle j\rangle}(-[3k_j^2 - 2k_j - 2]\kappa + 2(k_j+2)k_{\langle j\rangle}) \\ & + \beta_{j12}k_j(2[k_j^2-1]\kappa + 3k_j^2 - 1)k_{\langle j\rangle}^2 - (3k_j-2)k_{\langle j\rangle}\} \\ & - k_j([3k_j-1]\kappa + k_j), \end{aligned} \tag{4.5}$$

$$\begin{aligned} b_{2j} = {} & \tfrac{1}{4}\beta_{12}^{-2}k_{\langle j\rangle}^{-2}\{-\beta_j^2k_j^3k_{\langle j\rangle}(3\kappa+2) + 4\beta_j\beta_{j12}k_j^2k_{\langle j\rangle}^2(\kappa+1) - 4\beta_jk_jk_{\langle j\rangle} \\ & + 4\beta_{j12}k_{\langle j\rangle}^2\} + k_j^{-1}k_{\langle j\rangle}^{-2}\{\beta_jk_j^2([8k_j-5]\kappa + 5k_j - 3)k_{\langle j\rangle}^2 \\ & + (2k_j-1)k_{\langle j\rangle}^2, \end{aligned} \tag{4.6}$$

$$b_{3j} = -\beta_{12}^{-1}\{2\beta_jk_j^2(\kappa+1) + 1\}. \tag{4.7}$$

Theorem 4.1 shows that the direction of $\mathrm{E}[B]$ is not equal to the unknown best coefficient vector δ. This suggests to modify the bias of B by replacing the unknown parameters included in (4.4), (4.5), (4.6) and (4.7).

The cut-off point C can be expanded as

$$C = c + C_f + C_s, \tag{4.8}$$

where

$$\begin{aligned} C_f &= \tfrac{1}{2}\delta'(D_1 + D_2) + \tfrac{1}{2}\delta'\{(k_1 V_1 - k_2 V_2) + K_f(\Sigma_1 + \Sigma_2)\}\delta \\ &\quad + \delta'(k_1\Sigma_1 - k_2\Sigma_2)B_f, \\ C_s &= -\tfrac{1}{2}(D_1 - D_2)'(D_1 + D_2) - \tfrac{1}{2}\delta'\{(k_1 V_1 + k_2 V_2) + K_f(\Sigma_1 - \Sigma_2)\} \\ &\quad \times (D_1 + D_2) + \delta'(k_1 V_1 - k_2 V_2)B_f + \delta'(k_1\Sigma_1 - k_2\Sigma_2)B_s \\ &\quad + \tfrac{1}{2}B_f'(k_1\Sigma_1 - k_2\Sigma_2)B_f + K_f\{\tfrac{1}{2}\delta'(k_1 V_1 - k_2 V_2)\delta + \delta'(\Sigma_1 + \Sigma_2)B_f\} \\ &\quad + \tfrac{1}{2}K_s\delta'(\Sigma_1 + \Sigma_2)\delta. \end{aligned} \tag{4.9}$$

A calculation of the expectation of C_s gives the asymptotic bias of C as in the following theorem.

THEOREM 4.2. *The expectation of C can be expanded as*

$$\mathrm{E}[C] = c + \sum_{j=1}^{2} n_j^{-1} s_j c_j, \tag{4.10}$$

where

$$\begin{aligned} c_j = &-\tfrac{1}{16}\beta_{12}^{-3}k_{\langle j\rangle}^{-3}(1 - 2k_j k_{\langle j\rangle})\{3\beta_j^3\beta_{j12}k_j^4(3\kappa + 2) - 12\beta_j^2\beta_{j12}^2 k_j^3 k_{\langle j\rangle}(\kappa + 1) \\ &+ 12\beta_j^2\beta_{j12}k_j^2 - 12\beta_j\beta_{j12}^2 k_j k_{\langle j\rangle}\} + \tfrac{1}{16}\beta_{12}^{-2}k_{\langle j\rangle}^{-3}[3\beta_j^3 k_j^4(1 - 2k_j k_{\langle j\rangle}) \\ &\times (3\kappa + 2) - 4\beta_j^2\beta_{j12}k_j^2(2k_j^3 - 4k_j^2 - 5k_j + 3)k_{\langle j\rangle}(\kappa + 1) \\ &+ 16\beta_j^2\beta_{1122}k_j^2(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle}^2(\kappa + 1) \\ &- 4\beta_j^2 k_j^2\{k_j^2 k_{\langle j\rangle}^3(3\kappa + 2) - 3(1 - 2k_j k_{\langle j\rangle})\} \\ &- 4\beta_j\beta_{j12}\{4k_j^3 k_{\langle j\rangle}^4(\kappa + 1) + (2k_j^3 - 4k_j^2 - 5k_j + 3)k_{\langle j\rangle}\} \\ &+ 16\beta_j\beta_{1122}(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle}^2 + 16\beta_j k_j^2 k_{\langle j\rangle}^3 \\ &- 16\beta_{j12}k_j k_{j\langle j\rangle}^4] + \tfrac{1}{8}\beta_{12}^{-1}k_{\langle j\rangle}^{-3}[2\gamma_j\beta_j^2 k_j^3(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle}(\kappa + 1) \\ &+ 2\gamma_j\beta_j k_j(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle} - 2\gamma_{12}\beta_j^2(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle}^{-2}(\kappa + 1) \\ &- 2\gamma_{12}\beta_j(1 - 2k_j k_{\langle j\rangle})k_{\langle j\rangle}^2 - \beta_j^2 k_j^2(14k_j^3 - 25k_j^2 + 19k_j - 6)k_{\langle j\rangle}(3\kappa + 2) \end{aligned} \tag{4.11}$$

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

[4] G. J. McLachlan, An asymptotic unbiased technique for estimating the error rates in discriminant analysis, Biometrics **30** (1974), 239–249.

[5] H. Wakaki, Discriminant analysis under elliptical populations, Hiroshima Math. J., **24** (1994), 257–298.

Faculty of General Education
Ehime University
Bunkyo-Chou 3, *Matsuyama-shi*
Ehime 790, *Japan*