HIROSHIMA MATH. J.
24 (1994), 257–298

Discriminant analysis under elliptical populations

Hirofumi WAKAKI

(Received September 20, 1992)

0. Introduction

Consider independent random samples, of size n_j ($j = 1, 2$), from each of two p-variate populations Π_j having mean vectors μ_j and common covariance matrix Λ. Let the sample mean be denoted by $\bar{X}_j$ ($j = 1, 2$) and the pooled sample covariance matrix by S. Let X be an observation from one of the two populations. Fisher [7] showed that the linear combination of X which maximizes between sample variance relative to within samples variance is given by

$$(\bar{X}_1 - \bar{X}_2)' S^{-1} X, \tag{0.1}$$

which is known as Fisher's linear discriminant function (LDF). Welch [31] demonstrated that if both populations are assumed to be multivariate normal then the value of the log likelihood ratio in the two populations at any point X is given by

$$\lambda = \left\{X - \frac{1}{2}(\mu_1 + \mu_2)\right\}' \Lambda^{-1}(\mu_1 - \mu_2), \tag{0.2}$$

Therefore it can be shown that the optimal classification rule is to assign X into Π_1 (or Π_2) according to $\lambda > k$ (or $\lambda < k$). The cut point k is a constant depending on the relative costs of misclassification from each populations. Details of general principles of classification, and the derivation of the above rule are given in Chapter 6 of Anderson [2].

In practial situations the parameters are unknown, so the above rule must be modified. Wald [30] and Anderson [1] suggested replacing the unknown parameters by their sample estimators. Okamoto [24] derived asymptotic expansion formulas for the misclassification probabilities up to terms of the second order with respect to (n_1^{-1}, n_2^{-1}) under the assumption of normality. Siotani and Wang ([27], [28]) extended the formulas up to terms of the third order. A review of asymptotic expansions of classification statistics under normal populations is given by Siotani [26]. Chapter 9 of Siotani, Hayakawa and Fujikoshi [29] is also useful.

Under non-normal populations, several authors investigated the performance of the LDF. Lachenbruch, Sneeringer and Revo [20] have considered robustness of the LDF and the quadratic discriminant functions to three specific distributions. These distributions were generated from the normal distribution by using the non-linear transformations suggested by Johnson [15]. Their results indicated that the LDF was greatly affected by non-normality of the populations. On the other hand Balakrishnan and Kocherlakota [3] mentioned that the LDF is quite robust against the likelihood ratio rule in Monte Carlo simulations in which the mixtures of normal populations were taken. Nakanishi and Sato [23] also investigated the performance of the LDF and the quadratic discriminant function (QDF) for three types of non-normal distribution. Their purpose was a comparison of the LDF and the QDF. The results showed that the sign of the skewness of each populations and the kurtosis have essential effects. Koutras [18] obtained a general integral expression for evaluating the performance of the LDF with the population parameters under spherical distributions. He gave recurrence relations for certain special cases including the spherical gamma, Pearson VII, and generalized Laplace distributions. Krzanowski [19] gave a review of the work on the performance of the LDF when underlying assumptions are violated, which included the cases of unequal covariance matrices, continuous non-normal data, discrete data and mixtures of discrete and continuous variables.

In order to get robust discriminant functions, Randles et al. [25] considered to substitute M-estimators of the mean and the covariance matrices in the usual expressions for the linear and the quadratic discriminant functions. Their Monte Carlo results indicated lower misclassification probabilities compared to the LDF in cases of heavy-tailed or contaminated distributions. Broffitt, Clark and Lachenbruch [4] also investigated the method to use robust estimators, Huberized and trimed estimators of means and covariance matrices. However, none of their procedures produced a sufficient reduction in rates of misclassifications to counterbalance the added complexity of the discriminant rule.

In this paper we consider the classification problem when underlying assumptions may be violated.

In Part I we investigate the Fisher's linear discriminant function under elliptical populations with common covariance matrix. In Section 1 we give a simple expression of the conditional misclassification probabilities of the LDF. In Section 2, in order to derive asymptotic expansion formulas on misclassification probabilities, we derive an asymptotic expansion of the joint distribution of the sample mean and the sample covariance matrix under an elliptical population. In Section 3, we consider the conditional distribution of misclassification probabilities. The asymptotic expansions of the expected probabilities of each kind of misclassification are obtained. In the minimax criterion of the rule, we propose a loss of the estimators of unknown parameters. We also give an asymptotic expansion of the "risk" of the ordinal sample estimators in this framework. In Section 4, we give an estimator of the misclassification probabilities which is unbiased up to the order $(n_1 + n_2)^{-3/2}$.

In Part II, we consider to use M-estimators in order to get a robust classification rule. Huber [13] derived a robust M-estimator for location model. For an elliptical model Huber [14] derived a robust M-estimators of location and covariance matrix. For general parametric models Hample et al. [11] developed robust estimations using the influence functions. They obtained the B-robust M-estimator which has the smallest asymptotic variance subject to the bounded influence function. We apply their approach to our discriminant problem. In Section 5 we give a general setup of the discriminant problem. In Section 6 we prepare some definitions and lemmas related with the influence function under the case of two samples. In Section 7, we define a measure of sensitivity and a measure of efficiency of the estimator based on the loss function proposed in Section 3. In Section 8 we obtain the optimal M-estimators. In Section 9, we consider equivariant estimators. In the last section, we return to the elliptical model and apply the methods investigated in Sections 5–9 to it.

PART I. Fisher's linear discriminant function under elliptical populations

1. Minimax classification rule between two elliptical populations

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

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

where h is a decreasing function, μ is a $p \times 1$ parameter vector and Λ is a $p \times p$ positive definite matrix. It is known (cf. Kelker [17]) that the characteristic function of $E_p(\mu, \Lambda, h)$ has the form $\exp(it'\mu)\psi((t - \mu)'\Lambda(t - \mu))$. We assume that $\psi(s)$ is three times continuously differentiable at $s = 0$, which means that $E_p(\mu, \Lambda, h)$ has the 6th order moments. Then the covariance matrix is $\Omega = -2\psi'(0)\Lambda = \omega\Lambda$(say). We denote the unknown parameters as $\theta = (\mu_1, \mu_2, \Lambda)$ and Θ be the parameter space.

For any given parameter $\tau = (\eta_1, \eta_2, \Xi) \in \Theta$, we define the classification rule $R(\tau)$ as

$$(1.2)\qquad \text{assign } X \text{ into } \Pi_j \text{ if } (X-\bar{\eta})'\Xi^{-1}(\eta_j-\eta_{j'})>0 \qquad (j=1,2),$$

where $\bar{\eta}=(\eta_1+\eta_2)/2$ and $j'=3-j$. Let $P_j(\tau;\theta)$ $(j=1,2)$ be the probability of misclassifying X which belongs to Π_j.

THEOREM 1.1. *Let*

$$(1.3)\qquad c_j(\tau,\theta)=(\bar{\eta}-\mu_j)'\Xi^{-1}(\eta_j-\eta_{j'})/\|\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})\|$$

$(j=1,2)$, *and Q be a distribution function whose density function is*

$$(1.4)\qquad q(u)=\pi^{(p-1)/2}/\Gamma[(p-1)/2]\int_0^\infty s^{(p-3)/2}h(u^2+s)\,ds,$$

where Γ is the gamma function, and h is defined by (1.1). *Then the misclassification probabilities are expressed as*

$$(1.5)\qquad P_j(\tau;\theta)=Q\{c_j(\tau,\theta)\} \qquad (j=1,2).$$

PROOF. Let

$$(1.6)\qquad X=\Lambda^{1/2}Y+\mu_j.$$

Then Y is spherically distributed with the density function $h(Y'Y)$. We have

$$\begin{aligned}(1.7)\qquad P_j(\tau;\theta)&=\Pr\{(X-\bar{\eta})'\Xi^{-1}(\eta_j-\eta_{j'})<0\,|\,\Pi_j\}\\&=\Pr\{Y'\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})<(\bar{\eta}-\mu_j)'\Xi^{-1}(\eta_j-\eta_{j'})\}\\&=\Pr\{U<c_j(\tau,\theta)\},\end{aligned}$$

where

$$(1.8)\qquad U=Y'\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})/\|\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})\|.$$

Let H be an orthogonal matrix whose first row is

$$(1.9)\qquad \{\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})/\|\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})\|\}.$$

Since Y is spherical the distribution of U is the same as the one of

$$(1.10)\qquad Y'H\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})/\|\Lambda^{1/2}\Xi^{-1}(\eta_j-\eta_{j'})\|=Y_1,$$

where $Y=(Y_1,Y_2,\dots,Y_p)'$. Hence the distribution of U is the same as the marginal distribution of Y_1. In order to obtain the marginal dinsity function, we use Chu's representation (Theorem 1 of Chu [5]) of the density functin as

$$(1.11)\qquad h(Y'Y)=\int w(t)(2\pi)^{-p/2}t^{p/2}\exp\{-tY'Y/2\}\,dt,$$

where $w(t)$ is called the weighting function. Let $V=(Y_2,Y_3,\dots,Y_p)'$ then the marginal density function is expressed as

$$(1.12)\qquad q(Y_1)=\int\int w(t)(2\pi)^{-p/2}t^{p/2}\exp\{-tY'Y/2\}\,dt\,(dV)$$

$$= \int w(t)(2\pi)^{-1/2} t^{1/2} \exp\{-tY_1^2/2\}\, dt.$$

Using the expression (1.11) with $Y'Y = Y_1^2 + s$, we get

$$\int s^{(p-1)/2-1} h(Y_1^2 + s)\, ds \tag{1.13}$$

$$= \int s^{(p-1)/2-1} \int w(t)(2\pi)^{-p/2} t^{p/2} \exp\left\{-\frac{1}{2}t(Y_1^2 + s)\right\} dt\, ds$$

$$= \int w(t)(2\pi)^{-p/2} t^{p/2} \exp\left\{-\frac{1}{2}tY_1^2\right\} \int s^{(p-1)/2-1} \exp\left\{-\frac{1}{2}ts\right\} ds\, dt$$

$$= \int w(t)(2\pi)^{-p/2} t^{p/2} \exp\left\{-\frac{1}{2}tY_1^2\right\} \left(\frac{t}{2}\right)^{-(p-1)/2} \Gamma[(p-1)/2]\, dt.$$

Comparing this with (1.12) we obtain the marginal density functin as (1.4).

In the case of normal populations (1.4) is reduced to a standard normal density function, which can be checked easily with $h(s) = (2\pi)^{-p/2} \exp(-s/2)$.

THEOREM 1.2. *The classification rule $R(\theta)$ is minimax.*

PROOF. Since h is decreasing, the rule $R(\theta)$ is equivalent with a Bayes rule: assign X into Π_j if

$$h\{(X - \mu_j)'\Lambda^{-1}(X - \mu_j)\} > h\{(X - \mu_{j'})'\Lambda^{-1}(X - \mu_{j'})\}. \tag{1.14}$$

Therefore it is sufficient to show (cf. Anderson [2], page 203) that $P_1(\theta; \theta) = P_2(\theta; \theta)$, which is easily shown from Theorem 1.1 with

$$c_j(\theta, \theta) = -\frac{1}{2}(\mu_j - \mu_{j'})'\Lambda^{-1}(\mu_j - \mu_{j'})/\|\Lambda^{-1/2}(\mu_j - \mu_{j'})\| \tag{1.15}$$

$$= -\frac{1}{2}\|\Lambda^{-1/2}(\mu_1 - \mu_2)\| = -\Delta/2\,(\text{say}), \qquad (j = 1, 2).$$

In our notation Fisher's linear discrimination is expressed as $R(\hat{\theta}_s)$ with $\hat{\theta}_s = (\bar{X}_1, \bar{X}_2, \omega^{-1}S)$ where $\bar{X}_j$ $(j = 1, 2)$ is the sample mean and S is the pooled sample covariance matrix. The Theorem 1.2 shows that Fisher's linear discriminant function gives an asymptotically minimax rule in elliptical populations, since $\hat{\theta}_s$ asymptotically converges to θ.

2. Asymptotic expansion of the joint distribution of sample mean and sample covariance matrix from an elliptical population

Hayakawa and Puri [12] derived an asymptotic expansion of the distribution of sample covariance matrix under an elliptical population with mean 0. We deal with both the sample mean and the sample covariance matrix in the general case where the mean is unknown.

Let $X_1, X_2, \dots, X_n$ be an independent sample from $E_p(\mu, \Lambda, h)$ whose characteristic function is expressed as $\exp(it'\mu)\psi((t-\mu)'\Lambda(t-\mu))$. Assume that the covariance matrix Ω exists. Then $\Omega = \omega\Lambda$, where $\omega = -2\psi'(O)$. Denote the sample mean and the sample covariance matrix as $\bar{X}$ and S, respectively, and let

$$Z = n^{1/2}\Omega^{-1/2}(S - \Omega)\Omega^{-1/2} \tag{2.1}$$

and

$$Y = n^{1/2}\Omega^{-1/2}(\bar{X} - \mu). \tag{2.2}$$

Then the limiting distribution of Z and Y is mutually independent normal. The purpose of this section is to derive an asymptotic expansion of the joint distribution of Z and Y. Let

$$U_j = \Omega^{-1/2}(X_j - \mu), \qquad j = 1, \dots, n \tag{2.3}$$

and

$$\bar{U} = \frac{1}{n}\sum U_j. \tag{2.4}$$

Then Y and Z are expressed as

$$Y = n^{1/2}\bar{U} = n^{-1/2}\sum U_j \tag{2.5}$$

and

$$\begin{aligned} Z &= n^{1/2}\left\{\frac{1}{n-1}\sum(U_j - \bar{U})(U_j - \bar{U})' - I\right\} \\ &= \frac{n}{n-1}\{n^{-1/2}\sum U_j U_j' - n^{1/2}I - (n^{1/2}\bar{U}\bar{U}' - n^{-1/2}I)\} \\ &= \frac{n}{n-1}W - \frac{n}{n-1}n^{-1/2}(YY' - I), \end{aligned} \tag{2.6}$$

where

$$W = n^{-1/2}\sum(U_j U_j' - I). \tag{2.7}$$

First we consider the joint characteristic function of W and Y. The joint characteristic function is

(2.8)
$$\begin{aligned}\phi(T, \tau) &= \mathrm{E}[\operatorname{etr}\{iTW + i\tau Y'\}]\\ &= \mathrm{E}[\operatorname{etr}\{in^{-1/2}\textstyle\sum(TU_jU_j' - T + \tau U_j')\}]\\ &= [\mathrm{E}[\exp\{in^{-1/2}(-\operatorname{tr}(T) + U'TU + \tau'U)\}]]^n,\end{aligned}$$

where $T = \left(\frac{1}{2}(1 + \delta_{jk})t_{jk}\right)$, $\tau = (s_j)$ $(1 \leqq j, k \leqq p)$, δ_{jk} is Kronecker's delta and U is a spherical variable with characteristic function $\psi(\omega^{-1}\tau'\tau)$. By theorem 2 of Chu [5] U is represented as $U = (1/R)Z$, where Z is distributed as $N_p(0, I)$ and R is independent with Z. Therefore we have

(2.9)
$$\begin{aligned}&\mathrm{E}[\exp\{in^{-1/2}(U'TU + \tau'U)\}]\\ &= \mathrm{E}^*[\mathrm{E}[\exp\{in^{-1/2}(R^{-2}Z'TZ + R^{-1}\tau'Z)\}\,|\,R]]\\ &= \mathrm{E}^*\left[\,|I - 2in^{-1/2}R^{-2}T|^{-1/2}\exp\left\{-\frac{1}{2}n^{-1}R^{-2}\tau'(I - 2in^{-1/2}R^{-2}T)^{-1}\tau\right\}\right]\\ &= \mathrm{E}^*\left[\exp\left\{-\frac{1}{2}\log|I - 2in^{-1/2}R^{-2}T|\right.\right.\\ &\qquad\left.\left. - \frac{1}{2}n^{-1}R^{-2}\tau'(I - 2in^{-1/2}R^{-2}T)^{-1}\tau\right\}\right].\end{aligned}$$

The reason of superscript $*$ of the expectation is that the (probability) measure of R may be signed measure. The argument of the above exponential can be expanded as

(2.10)
$$\exp(n^{-1/2}F_1 + n^{-1}F_2 + n^{-3/2}F_3) + O(n^{-2}),$$

where

(2.11)
$$\begin{aligned}F_1 &= -i\operatorname{tr}(T) + iR^{-2}\operatorname{tr}(T),\\ F_2 &= -R^{-4}\operatorname{tr}(T^2) - \frac{1}{2}R^{-2}\tau'\tau,\\ F_3 &= -\frac{4}{3}iR^{-6}\operatorname{tr}(T^3) - iR^{-4}\tau'T\tau.\end{aligned}$$

Therefore

(2.12)
$$\begin{aligned}&\mathrm{E}[\exp\{in^{-1/2}(-\operatorname{tr}(T) + U'TU + \tau'U)\}]\\ &= \mathrm{E}^*\left[1 + n^{-1/2}F_1 + n^{-1}\left(\frac{1}{2}F_1^2 + F_2\right)\right.\\ &\qquad\left. + n^{-3/2}\left(\frac{1}{6}F_1^3 + F_1F_2 + F_3\right)\right] + O(n^{-2}).\end{aligned}$$

The characteristic function of U is

$$\psi(\omega^{-1}\tau'\tau) = \sum \frac{\psi^{(k)}(0)}{k!}(\omega^{-1}\tau'\tau)^k. \tag{2.13}$$

On the other hand we can express the characteristic function of U as

$$\begin{aligned}\psi(\omega^{-1}\tau'\tau) &= \mathrm{E}[\exp(i\tau'U)] \\ &= \mathrm{E}^*[\mathrm{E}[\exp(i\tau'ZR^{-1})\,|\,R]] \\ &= \mathrm{E}^*\left[\exp\left(-\frac{1}{2}\tau'\tau R^{-2}\right)\right] \\ &= \sum \frac{1}{k!}\mathrm{E}^*[R^{-2k}]\left(-\frac{1}{2}\right)^k(\tau'\tau)^k.\end{aligned} \tag{2.14}$$

Comparing coefficients of $\tau'\tau$ in (2.13) and (2.14) we obtain

$$\begin{aligned}&\mathrm{E}^*[R^{-2k}] = (-2)^k\omega^{-k}\psi^{(k)}(0) = \psi^{(k)}(0)/\{\psi'(0)\}^k, \\ &\mathrm{E}^*[R^{-2}] = 1, \\ &\mathrm{E}^*[R^{-4}] = \psi^{(2)}(0)/\{\psi'(0)\}^2 = \kappa + 1,\end{aligned} \tag{2.15}$$

where κ is the kurtosis parameter, and

$$\mathrm{E}^*[R^{-6}] = \psi^{(3)}(0)/\{\psi'(0)\}^3 = \psi_3 + 1 \text{ (say)}. \tag{2.16}$$

Using these formulas we obtain

$$\begin{aligned}&\mathrm{E}[\exp\{in^{-1/2}(-\mathrm{tr}(T) + U'TU + \tau'U)\}] \\ &= 1 + n^{-1}G_1 + n^{-3/2}G_2 + O(n^{-2}),\end{aligned} \tag{2.17}$$

where

$$\begin{aligned}G_1 &= \mathrm{E}^*\left[\frac{1}{2}F_1^2 + F_2\right] \\ &= -\frac{1}{2}\kappa\,\mathrm{tr}^2(T) - (\kappa+1)\,\mathrm{tr}(T^2) - \frac{1}{2}\tau'\tau, \\ G_2 &= \mathrm{E}^*\left[\frac{1}{6}F_1^3 + F_1F_2 + F_3\right] \\ &= -\frac{i}{6}(\psi_3 - 3\kappa)\,\mathrm{tr}^3(T) \\ &\quad -\frac{i}{2}\kappa\,\mathrm{tr}(T)\tau'\tau - i(\psi_3 - \kappa)\,\mathrm{tr}(T)\,\mathrm{tr}(T^2)\end{aligned} \tag{2.18}$$

$$-\frac{4}{3}i(\psi_3 + 1)\,\mathrm{tr}\,(T^3) - i(\kappa + 1)\tau' T\tau.$$

Here the notation $\mathrm{tr}^k(T)$ means $\{\mathrm{tr}\,(T)\}^k$. From (2.17) the characteristic function of W and Y can be expanded as

$$\begin{aligned}(2.19)\qquad \phi(T, \tau) &= \{1 + n^{-1}G_1 + n^{-3/2}G_2 + O(n^{-2})\}^n \\ &= \exp\,[n \log\{1 + n^{-1}G_1 + n^{-3/2}G_2 + O(n^{-2})\}] \\ &= \exp\{G_1 + n^{-1/2}G_2 + O(n^{-1})\}.\end{aligned}$$

Inverting $\phi(T, \tau)$, the joint density function of W and Y can be expressed as

$$\begin{aligned}(2.20)\qquad f(W, Y) &= (2\pi)^{-p(p+3)/4}\int \exp\{-i\,\mathrm{tr}\,(WT) - i\tau' Y + G_1\} \\ &\quad \cdot \{1 + n^{-1/2}G_2\}(dT)(d\tau) + O(n^{-1}),\end{aligned}$$

where $(dT) = \prod_j dt_{jj}\prod_{k<m} dt_{km}$. From (2.6) we have

$$(2.21)\qquad W = \frac{n-1}{n}Z + n^{-1/2}(YY' - I) = Z + n^{-1/2}(YY' - I) + O(n^{-1}).$$

Since the Jacobian of the translation (W, Y) to (Z, Y) is

$$(2.22)\qquad \{(n-1)/n\}^{p(p+1)/2} = 1 + O(n^{-1}),$$

the substitution of (2.21) to (2.20) gives an asymptotic expansion of the joint density function of Z and Y as

$$\begin{aligned}(2.23)\qquad f(Z, Y) &= (2\pi)^{-p(p+3)/4}\int \exp\{-i\,\mathrm{tr}\,(ZT) - i\tau' Y + G_1\} \\ &\quad \cdot [1 + n^{-1/2}\{G_2 - iY'TY + i\,\mathrm{tr}\,(T)\}](dT)(d\tau) + O(n^{-1}).\end{aligned}$$

Let

$$\begin{aligned}(2.24)\qquad T_1 &= (t_{11}, \ldots, t_{pp})', \\ T_2 &= (t_{12}, t_{13}, \ldots, t_{p-1,p})', \\ \Omega_1 &= 2(\kappa + 1)I_p + \kappa ll', \; l = (1, \ldots, 1)',\end{aligned}$$

and

$$(2.25)\qquad Z_1 = (z_{11}, \ldots, z_{pp})',\; Z_2 = (z_{12}, z_{13}, \ldots, z_{p-1,p})',$$

where $Z = (z_{kl})$, $(k, l = 1, 2, \ldots, p)$, as in Hayakawa and Puri [12]. Then the argument of the exponential is expressed as

$$(2.26)\qquad -iT_1'Z_1 - iT_2'Z_2 - i\tau' Y - \frac{1}{2}(T_1'\Omega_1 T_1 + (\kappa + 1)T_2'T_2 + \tau'\tau)$$

which implies that the limiting distribution of Z_1, Z_2 and Y is mutually independent normal with mean 0. The covariance matrices of Z_1, Z_2 and Y are Ω_1, $(\kappa+1)I_{p(p-1)/2}$ and I_p, respectively. Let $J_p = I_p - p^{-1}ll'$ then

(2.27) $$\Omega_1^{-1} = uI_p + (v-u)p^{-1}ll' = uJ_p + vp^{-1}ll',$$

where

(2.28) $$u = \frac{1}{2(\kappa+1)}, \quad v = \frac{1}{(p+2)\kappa+2}.$$

The expression (2.27) is useful in calculation of expectations since J_p and $p^{-1}ll'$ are idempotent, $J_p l = 0$ and $(p^{-1}ll')l = l$. The calculation of integrations in (2.23) gives the asymptotic expansion of the joint density function of Z and Y up to the order $n^{-1/2}$ as in the following theorem.

THEOREM 2.1. *Let Z and Y be random matrix and random vector given by* (2.1) *and* (2.2), *respectively. Then the joint density function of Z and Y can be expanded for large n as*:

(2.29) $$f(Z, Y) = (2\pi)^{-p(p+3)/4}|\Omega_1|^{-1/2}(\kappa+1)^{-p(p-1)/4} \cdot \exp\left[-\frac{1}{2}\left\{Z_1'\Omega_1^{-1}Z_1 + \frac{1}{\kappa+1}Z_2'Z_2 + Y'Y\right\}\right] \cdot [1 + n^{-1/2}g(Z, y) + O(n^{-1})],$$

where

(2.30) $$g(Z, Y) = a_1\,\mathrm{tr}\,(Z) + a_2\,\mathrm{tr}^3\,(Z) + a_3\,\mathrm{tr}\,(Z^3) + a_4\,\mathrm{tr}\,(Z)\,\mathrm{tr}\,(Z^2) + a_5 Y'Y\mathrm{tr}\,(Z) + a_6 Y'ZY$$

and

(2.31) $$a_1 = -\psi_3\left\{uv(4p+1-4p^{-1}) + v^2\left(\frac{1}{2}p + 3 + 4p^{-1}\right)\right\} + \kappa\left\{uv(2p-1) + v^2\left(\frac{3}{2}p+3\right) - v\left(\frac{1}{2}p+1\right)\right\} - 2uv(p+1-2p^{-1}) - 4v^2p^{-1},$$

$$a_2 = \psi_3\left\{\frac{8}{3}u^3p^{-2} - u^2v(p^{-1}+4p^{-2}) + v^3\left(\frac{1}{6} + p^{-1} + \frac{4}{3}p^{-2}\right)\right\} + \kappa\left\{u^2vp^{-1} - v^3\left(\frac{1}{2} + p^{-1}\right)\right\} + \frac{8}{3}u^3p^{-2} - 4u^2vp^{-2} + \frac{4}{3}v^3p^{-2}$$

$$a_3 = \frac{4}{3}(\psi_3 + 1)u^3,$$

$$a_4 = \psi_3\{-4u^3p^{-1} + u^2v(4p^{-1} + 1)\} - \kappa u^2v - 4u^3p^{-1} + 4u^2vp^{-1},$$

$$a_5 = \kappa\left\{-up^{-1} + v\left(\frac{1}{2} + p^{-1}\right)\right\},$$

$$a_6 = \kappa u.$$

In the case of normal population Y and Z is independent. Y has exactly normal distribution $N_p(0, I_p)$. Since $\kappa = \psi_3 = 0$ the marginal density function of Z can be expanded as

$$(2.32) \qquad 2^{-p(p+3)/4}\pi^{-p(p+1)/4}\,\mathrm{etr}\left(-\frac{1}{4}Z^2\right)$$
$$\cdot\left[1 + n^{-1/2}\left\{-\frac{1}{2}(p+1)\,\mathrm{tr}\,(Z) + \frac{1}{6}\,\mathrm{tr}\,(Z^3)\right\}\right] + O(n^{-1})$$

which was essentially given by Fujikoshi [8] (see also Siotani, Hayakawa and Fujikoshi [29], page 159).

3. On the conditional misclassification probabilities

We return to the classification problem of two elliptical populations in Section 1. Suppose that the training samples of size n_1 and n_2 from Π_1 and Π_2, respectively, are given. Let $\bar{X}_j$ be the sample mean and S_j be the sample covariance matrix from Π_j $(j = 1, 2)$. The pooled sample covariance matrix is given by

$$(3.1) \qquad S = (N-2)^{-1}\{(n_1 - 1)S_1 + (n_2 - 1)S_2\}$$

where $N = n_1 + n_2$ and $r_j = n_j/N$ $(j = 1, 2)$. In this section we consider the distributions of the conditional misclassification probabilities of Fisher's linear discrimination $R(\hat{\theta}_s)$, where $\hat{\theta}_s = (\bar{X}_1, \bar{X}_2, \omega^{-1}S)$. We modified S to $\omega^{-1}S$ in order to get a consistent estimator. However, this is just for convenience of calculations, since the factor ω^{-1} causes no change of the classification rule.

From Theorem 1 in Section 1, we know that the conditional misclassification probability $P_j(\hat{\theta}_s, \theta)$ is the function of $c_j(\hat{\theta}_s, \theta)$ $(j = 1, 2)$. Therefore we prepare the following lemma of Taylor expansion of $c_j(\tau, \theta)$ in a neighborhood of $\tau = \theta$ in order to investigate $P_j(\hat{\theta}_s, \theta)$ using an asymptotic expansion method.

LEMMA 3.1. *Let* $\tau = (\eta_1, \eta_2, \Xi) \in \Theta$, *where*

(3.2) $$\eta_j = \mu_j + \varepsilon_j (j = 1, 2), \quad \Xi = \Lambda + H.$$

Then the cut point $c_j(\tau, \theta)$ *given by* (1.3) *can be expanded for small* ε_j's *and* H *as*

(3.3)
$$\begin{aligned} & c_j(\tau; \theta) \\ &= -\frac{1}{2}\Delta + \frac{1}{2}\Delta^{-1}\xi_j'\Lambda^{-1/2}(\varepsilon_j + \varepsilon_{j'}) \\ &+ \frac{1}{4}\Delta^{-1}\{(\varepsilon_{j'} + 3\varepsilon_j)'\Lambda^{-1}(\varepsilon_j - \varepsilon_{j'}) - 4\xi_j'\Lambda^{-1/2}H\Lambda^{-1}\varepsilon_j \\ &+ \xi_j'\Lambda^{-1/2}H\Lambda^{-1}H\Lambda^{-1/2}\xi_j\} \\ &- \frac{1}{4}\Delta^{-3}\{\xi_j'\Lambda^{-1/2}(\varepsilon_j - \varepsilon_{j'}) - \xi_j'\Lambda^{-1/2}H\Lambda^{-1/2}\xi_j\} \\ &\cdot \{\xi_j'\Lambda^{-1/2}(3\varepsilon_j + \varepsilon_{j'}) - \xi_j'\Lambda^{-1/2}H\Lambda^{-1/2}\xi_j\} \\ &+ O(\|\tau - \theta\|^3), \end{aligned}$$

where $\xi_j = \Lambda^{-1/2}(\mu_j - \mu_{j'})$ $(j = 1, 2; j' = 3 - j)$ and $\Delta^2 = \xi_1'\xi_1$.

THEOREM 3.1. *Let*

(3.4) $$PN_j = N^{-1/2}\{P_j(\hat{\theta}_s, \theta) - P_j(\theta, \theta)\} \qquad (j = 1, 2).$$

Then the limiting distribution of PN_j *is* $N(0, v^2)$, *where*

(3.5) $$v^2 = \frac{1}{4}q(-\Delta/2)^2\omega r^{(1)},$$

q *is the density function given by* (1.4) *and* $r^{(1)} = r_1^{-1} + r_2^{-1}$.

PROOF. From Theorem 1.1 and Lemma 3.1 we obtain

(3.6)
$$\begin{aligned} &(\partial/\partial\eta_k)P_j(\tau, \theta)|_{\tau=\theta} = \frac{1}{2}q(-\Delta/2)\Delta^{-1}\xi_j'\Lambda^{-1/2} \qquad (k = 1, 2), \\ &(\partial/\partial\Xi)P_j(\tau, \theta)|_{\tau=\theta} = 0 (\in R^{p(p+1)/2}). \end{aligned}$$

Further, the limiting distributions of $N_j^{-1/2}(\bar{X}_j - \mu)$ is $N_j(0, r_j^{-1}\Omega)$. These shows the desired result (see Cramer [6], page 366).

In order to obtain the terms of $O(N^{-1/2})$, first we expand the joint characteristic function of PN_1 anf PN_2. Because the joint characteristic function gives distribution of any linear combinations of PN_1 anf PN_2. We will need the distribution of $(PN_1 - PN_2)/2$ in the last theorem of the present section.

LEMMA 3.2. *Let $\Psi(t)$ be the characteristic function of (PN_1, PN_2) where $t=(t_1, t_2)'$. Then $\Psi(t)$ can be expanded as*

$$\Psi(t) = \exp\left\{-\frac{1}{2}(t_1 - t_2)^2 v^2\right\} \cdot \{1 + N^{-1/2} i \sum t_k(b_0(t_1 - t_2)^2 + b_k)\} + O(N^{-1}), \tag{3.7}$$

where v^2 is given by (3.5),

$$b_0 = -\frac{1}{32} q_1^2 q_2 \omega^2 r^{(2)} \quad (r^{(2)} = r_1^{-2} + r_2^{-2}), \tag{3.8}$$

$$b_k = \frac{1}{8} q_2 \omega r^{(1)} + \frac{1}{4} q_1 \omega (3r_k^{-1} - r_{k'}^{-1})(p-1)\Delta^{-1} + \frac{1}{4} q_1 (p-1)(\kappa+1)\Delta \quad (k = 1, 2;\ k' = 3 - k)$$

and

$$q_1 = q(-\Delta/2), \quad q_2 = q'(-\Delta/2). \tag{3.9}$$

PROOF. Let

$$Y_j = n_j^{1/2} \Omega^{-1/2}(\bar{X}_j - \mu_j), \quad Z_j = n_j^{1/2} \Omega^{-1/2}(S_j - \Omega)\Omega^{-1/2}. \tag{3.10}$$

Then

$$\bar{X}_j = \mu_j + N^{-1/2} \omega^{1/2} \Delta^{1/2} r_j^{-1/2} Y_j, \quad S_j = \Omega^{1/2}[I + n_j^{-1/2} Z_j]\Omega^{1/2} \quad (j = 1, 2). \tag{3.11}$$

Since $(n_j - 1)/(N - 2) = r_j + O(N^{-1})$,

$$\begin{aligned} \omega^{-1} S &= \omega^{-1} \Omega^{1/2}(I + r_1 n_1^{-1/2} Z_1 + r_2 n_2^{-1/2} Z_2)\Omega^{1/2} + O_p(N^{-3/2}) \\ &= \Delta + N^{-1/2} \Delta^{1/2}(r_1^{1/2} Z_1 + r_2^{1/2} Z_2)\Delta^{1/2} + O_p(N^{-3/2}). \end{aligned} \tag{3.12}$$

From Lemma 3.1 we obtain

$$c_j(\hat{\theta}_s; \theta) = -\frac{1}{2}\Delta + N^{-1/2} c_j^{(1)} + N^{-1} c_j^{(2)} + O_p(N^{-3/2}), \tag{3.13}$$

where

$$c_j^{(1)} = \frac{1}{2}\Delta^{-1} \omega^{1/2} \xi_j'(r_j^{-1/2} Y_j + r_{j'}^{-1/2} Y_{j'}), \tag{3.14}$$

$$c_j^{(2)} = \frac{1}{4}\Delta^{-1}\omega(r_{j'}^{-1/2}Y_{j'} + 3r_j^{-1/2}Y_j)'(r_j^{-1/2}Y_j - r_{j'}^{-1/2}Y_{j'})$$
$$- \Delta^{-1}\omega^{1/2}\xi_j'(r_j^{1/2}Z_j + r_{j'}^{1/2}Z_{j'})r_j^{-1/2}Y_j$$
$$+ \frac{1}{4}\Delta^{-1}\xi_j'(r_j^{1/2}Z_j + r_{j'}^{1/2}Z_{j'})^2\xi_j$$
$$- \frac{1}{4}\Delta^{-3}\{\omega^{1/2}\xi_j'(r_j^{-1/2}Y_j - r_{j'}^{-1/2}Y_{j'}) - \xi_j'(r_j^{1/2}Z_j + r_{j'}^{1/2}Z_{j'})\xi_j\}$$
$$\cdot \{\omega^{1/2}\xi_j'(3r_j^{-1/2}Y_j + r_{j'}^{-1/2}Y_{j'}) - \xi_j'(r_j^{1/2}Z_j + r_{j'}^{1/2}Z_{j'})\xi_j\}.$$

Considering Taylor expansion of $P_j(\hat{\theta}_s, \theta) = Q(c_j(\hat{\theta}_s, \theta))$ at $c_j = -\Delta/2$, we obtain

$$\begin{aligned}(3.15)\qquad PN_j &= N^{1/2}[P_j(\hat{\theta}_s; \theta) - P_j(\theta; \theta)] \\ &= q_1 c_j^{(1)} + N^{-1/2}\left\{q_1 c_j^{(2)} + \frac{1}{2}q_2(c_j^{(1)})^2\right\} + O_p(N^{-1}).\end{aligned}$$

Therefore the characteristic function of PN_1 anf PN_2 can be expanded as

(3.16)

$$\begin{aligned}\Psi(t) &= \mathrm{E}[\exp(it_1 PN_1 + it_2 PN_2)] \\ &= \mathrm{E}\left[\exp\left[i\textstyle\sum_k t_k(q_1 c_k^{(1)} + N^{-1/2}\left\{q_1 c_k^{(2)} + \frac{1}{2}q_2(c_k^{(1)})^2\right\} + O_p(N^{-1})\right]\right] \\ &= \mathrm{E}\left[\exp\left\{i\textstyle\sum_k t_k q_1 \frac{1}{2}\omega^{1/2}\Delta^{-1}\xi_k'(r_k^{-1/2}Y_k + r_{k'}^{-1/2}Y_{k'})\right\}\right. \\ &\quad \left.\cdot\left[1 + N^{-1/2}i\textstyle\sum_k t_k\left\{q_1 c_k^{(2)} + \frac{1}{2}q_2(c_k^{(1)})^2\right\}\right]\right] + O(N^{-1}).\end{aligned}$$

Taking the expectation by using the joint density of (Z_j, Y_j) given by (2.29), we can see that the characteristic function can be reduced to (3.7).

Lemma 3.2 shows that the joint limiting distribution of PN_1 and PN_2 is $N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\right]$, which is degenerate (cf. Muirhead [22], page 4). More preciously, $PN_2 = -PN_1 + O_p(N^{-1/2})$. This means that variation of the estimator causes P_1 get larger, smaller the P_2. Therefore the total misclassification probability $(P_1 + P_2)/2$ (assuming equal prior probabilities) is stable.

THEOREM 3.2. *The marginal distribution function of* PN_j $(j = 1, 2)$ *can be expanded as*

$$(3.17)\qquad \begin{aligned}&\Pr\{PN_j < x\}\\ &= \Pr\{N^{-1/2}[P_j(\hat{\theta}_s;\theta) - P_j(\theta;\theta)] < x\}\\ &= \Phi(x/v) + N^{-1/2}\phi(x/v)/v\{(x^2/v^2 - 1)\tilde{b}_0 - b_j\} + O(N^{-1}),\end{aligned}$$

where Φ and ϕ are the distribution function and the density function of $N(0, 1)$, respectively, b_j are given by (3.8) *and*

$$(3.18)\qquad \tilde{b}_0 = b_0/v^2 = -\frac{1}{8}q_2\omega r^{(2)}/r^{(1)}.$$

PROOF. From Lemma 3.2 the characteristic function of PN_j is given by

$$(3.19)\qquad \begin{aligned}\Psi_j(t_j) &= \exp\left\{-\frac{1}{2}t_j^2v^2\right\}\\ &\quad\cdot\{1 + N^{-1/2}it_j(b_0t_j^2 + b_j)\} + O(N^{-1}).\end{aligned}$$

The inversion of Ψ_j gives the expansion of the distribution function.

COROLLARY 3.2. *The expected misclassification probabilities can be expanded as*

$$(3.20)\qquad \mathrm{E}[P_j(\hat{\theta}_s;\theta)] = Q(-\Delta/2) + N^{-1}b_j + O(N^{-3/2}) \qquad (j = 1, 2),$$

where Q is the distribution function given in Theorem 1.1 *and b_j is given by* (3.8).

PROOF. The expectation of PN_j is given by $i^{-1}\Psi_j'(0) = N^{-1/2}b_j + O(N^{-1})$.

When we consider $R(\hat{\theta})$ as an estimator of the minimax rule $R(\theta)$, we may use

$$(3.21)\qquad \begin{aligned}L(\hat{\theta},\theta) &= \max\{P_1(\hat{\theta},\theta), P_2(\hat{\theta},\theta)\} - P_1(\theta,\theta)\\ &= \max\{P_1(\hat{\theta},\theta), P_2(\hat{\theta},\theta)\} - P_2(\theta,\theta)\end{aligned}$$

as a natural loss of $\hat{\theta}$. Hence one of the important criteria on the goodness of $\hat{\theta}$ is given by $\mathrm{E}[L(\hat{\theta},\theta)]$ which may be called the risk of $\hat{\theta}$ in minimax classification.

THEOREM 3.3. *The risk of $\hat{\theta}_s$ can be expanded as*

$$(3.22)\qquad \mathrm{E}[L(\hat{\theta}_s,\theta)] = N^{-1/2}(2/\pi)^{1/2}v + N^{-1}\frac{1}{2}(b_1 + b_2) + O(N^{-3/2}),$$

where v is given by (3.5) *and b_j $(j = 1, 2)$ is given by* (3.8).

PROOF. Using the equation $\max(a, b) = (a + b)/2 + |a + b|/2$, we get

$$(3.23)\qquad N^{1/2}\mathrm{E}[L(\hat{\theta}_s,\theta)] = \mathrm{E}[\max(PN_1, PN_2)]$$

$$= \frac{1}{2} \mathrm{E}[(PN_1 + PN_2) + |PN_1 - PN_2|].$$

From Lemma 3.2 the characteristic function of $(PN_1 - PN_2)/2$ is given by

$$\Psi\{(t/2, -t/2)'\} = \exp\left\{-\frac{1}{2}t^2 v^2\right\} \cdot \left\{1 + N^{-1/2}\frac{i}{2}t(b_1 - b_2)\right\} + O(N^{-1}). \tag{3.24}$$

It's inversion gives an asymptotic expansion of the density function of $(PN_1 - PN_2)/2$ as

$$v^{-1}\phi(x/v)\left\{1 - N^{-1/2}\frac{1}{2}(b_1 - b_2)/v^2 \cdot x\right\} + O(N^{-1}), \tag{3.25}$$

and hence

$$\begin{aligned} \mathrm{E}\left[\frac{1}{2}|PN_1 - PN_2|\right] &= \int |x|/v\left\{1 - N^{-1/2}\frac{1}{2}(b_1 - b_2)/v^2 \cdot x\right\}\phi(x/v)\,dx + O(N^{-1}) \\ &= \int (2\pi)^{-1/2}v^{-1}|x| \exp\{-x^2/(2v^2)\}\,dx + O(N^{-1}) \\ &= (2/\pi)^{1/2} v[-\exp\{-x^2/(2v^2)\}]_0^{\infty} + O(N^{-1}) \\ &= (2/\pi)^{1/2} v + O(N^{-1}). \end{aligned} \tag{3.26}$$

From Corollary 3.2 we get

$$\mathrm{E}[(PN_1 + PN_2)/2] = N^{-1/2}(b_1 + b_2)/2 + O(N^{-1}). \tag{3.27}$$

Substituting (3.26) and (3.27) into (3.23) we obtain the result (3.22).

In the case of normal population

$$\kappa = 0,\ \omega = 1,\ q_1 = \phi(-\Delta/2),\ q_2 = \Delta/2\phi(-\Delta/2). \tag{3.28}$$

Therefore, the coefficients v, $\tilde{b}_0$ and b_j $(j = 1, 2)$ are reduced as follows:

$$\begin{aligned} v^2 &= \frac{1}{4}\{\phi(-\Delta/2)\}^2 r^{(1)}, \\ \tilde{b}_0 &= -\frac{1}{16}\Delta r^{(2)}/r^{(1)}\phi(-\Delta/2), \\ b_j &= \left\{\frac{1}{16}\Delta r^{(1)} + \frac{1}{4}(p-1)\Delta^{-1}(3r_j^{-1} - r_{j'}^{-1}) + \frac{1}{4}(p-1)\Delta\right\}\phi(-\Delta/2). \end{aligned} \tag{3.29}$$

4. Estimation of the misclassification probabilities

In this section we consider the problem of estimating the expected misclassification probabilities, which is expanded as (3.20).

In order to get an unbiased estimator of these probabilities, we prepare the next lemma.

LEMMA 4.1. *Consider an estimator $Q(-\hat{\Delta}/2)$ of misclassification probability $Q(-\Delta/2)$ of $R(\theta)$, where Q is given in Theorem 1.1 and*

$$\hat{\Delta}^2 = (\bar{X}_1 - \bar{X}_2)'(\omega^{-1}S)^{-1}(\bar{X}_1 - \bar{X}_2). \tag{4.1}$$

Then the bias is given by

$$\begin{aligned} &E[Q(-\hat{\Delta}/2)] - Q(-\Delta/2) \\ &= N^{-1}\left[q_2\left\{\frac{1}{8}r^{(1)}\omega + \frac{1}{32}(3\kappa+2)\Delta^2\right\} - \frac{1}{4}q_1(p-1)r^{(1)}\omega\Delta^{-1}\right. \\ &\quad \left. + q_1\left\{-\frac{1}{4}(p+2)\kappa - \frac{1}{4}(p+1) + \frac{1}{16}(3\kappa+2)\right\}\Delta\right] + O(N^{-3/2}). \end{aligned} \tag{4.2}$$

PROOF. Using (3.11) and (3.12), $\hat{\Delta}$ is expanded as

$$\hat{\Delta}^2 = \Delta^2 + N^{-1/2}\delta_1 + N^{-1}\delta_2 + O_p(N^{-3/2}), \tag{4.3}$$

where

$$\begin{aligned} \delta_1 &= \{2\omega^{1/2}\xi_1'(r_1^{-1/2}Y_1 - r_2^{-1/2}Y_2) - \xi_1'(r_1^{1/2}Z_1 + r_2^{1/2}Z_2)\xi_1\}, \\ \delta_2 &= \{\omega(r_1^{-1/2}Y_1 - r_2^{-1/2}Y_2)'(r_1^{-1/2}Y_1 - r_2^{-1/2}Y_2) \\ &\quad - 2\omega^{1/2}\xi_1'(r_1^{1/2}Z_1 + r_2^{1/2}Z_2)(r_1^{-1/2}Y_1 - r_2^{-1/2}Y_2) \\ &\quad + \xi_1'(r_1^{1/2}Z_1 + r_2^{1/2}Z_2)^2\xi_1\}. \end{aligned} \tag{4.4}$$

The expansion of $\hat{\Delta}^2$ implies that

$$\begin{aligned} -\hat{\Delta}/2 &= -\Delta/2 - \frac{1}{4}N^{-1/2}\Delta^{-1}\delta_1 \\ &\quad - N^{-1}\left(\frac{1}{4}\Delta^{-1}\delta_2 - \frac{1}{16}\Delta^{-3}\delta_1^2\right) + O_p(N^{-3/2}). \end{aligned} \tag{4.5}$$

Therefore Taylor expansion of Q gives

$$\begin{aligned} Q(-\hat{\Delta}/2) &= Q(-\Delta/2) \\ &\quad + q_1\left\{-\frac{1}{4}N^{-1/2}\Delta^{-1}\delta_1 - N^{-1}\left(\frac{1}{4}\Delta^{-1}\delta_2 - \frac{1}{16}\Delta^{-3}\delta_1^2\right)\right\} \end{aligned} \tag{4.6}$$

$$+\frac{1}{2}q_2\left\{\frac{1}{4}N^{-1/2}\Delta^{-1}\delta_1\right\}^2+O_p(N^{-3/2}),$$

where q_1 and q_2 are given by (3.9). Taking the expectation, we get the desired result.

From Corollary 3.2 and the above lemma we can get an estimator of the misclassification probabilities for $R(\hat{\theta}_s)$ as the following theorem.

THEOREM 4.1. *Let* $\hat{\Delta}^2$ *be given by* (4.1), *then*

$$Q(-\hat{\Delta}/2)+N^{-1}\left[-\frac{1}{32}q_2(3\kappa+2)\hat{\Delta}^2+q_1\omega r_j^{-1}(p-1)\hat{\Delta}^{-1}+q_1\hat{\Delta}\left\{\frac{1}{2}p(\kappa+1)+\frac{1}{16}\kappa-\frac{1}{8}\right\}\right] \tag{4.7}$$

is an unbiased estimator of the expected misclassification probability $\mathrm{E}[P_j(\hat{\theta}_s;\theta)]$ $(j=1, 2)$ *up to the order* $N^{-3/2}$.

In the normal case, (4.7) is reduced to

$$\Phi(-\hat{\Delta}/2)+N^{-1}\left\{-\frac{1}{32}\hat{\Delta}^3+r_j^{-1}(p-1)\hat{\Delta}^{-1}+\frac{1}{8}(4p-1)\hat{\Delta}\right\}\phi(-\hat{\Delta}/2), \tag{4.8}$$

which agrees with the result of McLachlan [21].

When κ is unknown, we need to replace κ in (4.7) by an estimate $\hat{\kappa}$.

PART II. Robust estimators in discriminant analysis

5. A general setup of estimation problem in discriminant analysis

In Part II we consider the classification problem under a general setup. Suppose that the population Π_j $(j=1, 2)$ has the density function $f(x;\eta_j)$, where the unknown parameter $\eta_j(\in H)$ is a $(q+r)$-dimensional vector. We assume that the last r elements of η_1 and η_2 are equal. So that we denote $\eta_j=(\zeta_j', \xi')'$ $(j=1, 2)$ and its parameter space as $H=Z\times\Xi$, where $Z\subset R^q$ and $\Xi\subset R^r$. We also use notations $\theta=(\zeta_1', \zeta_2', \xi')'$ and $\Theta=H\times H\times\Xi$. The sample space is written as $\Omega\subset R^p$. In the case of elliptical populations, $q=p$, $r=p(p+1)/2$, $\eta_j=(\mu_j, \Lambda)$ and

$$f(x;\eta_j)=|\Lambda|^{-1/2}h\{(x-\mu_j)'\Lambda^{-1}(x-\mu_j)\} \qquad (j=1, 2). \tag{5.1}$$

We identify the minimax rule with the region of the sample space in which the observation is assigned to Π_1. The minimax region is given by

(5.2) $$R(\theta) = \{x; f(x;\eta_1)/f(x;\eta_2) > k(\theta)\}.$$

Here $k(\theta)$ is obtained from the equation

(5.3) $$F(R(\theta)^c; \eta_1) = F(R(\theta); \eta_2),$$

and $F(\ ;\eta)$ is the probability measure corresponding to the density function $f(x;\eta)$. We often use the notation $F_{j\theta}$ instead of $F(\ ;\eta_j)$.

In Part I, we considered to use the sample mean and the pooled sample covariance matrix to estimate the minimax classification rule for elliptical populations. It is also natural to use the maximum likelihood estimator to estimate the minimax region $R(\theta)$. However, in general, the maximum likelihood estimator is not robust against deviations from the assumptions. For example, the sample mean is known to be sensitive to outliers. For general parametric models Hampel et al. [11] developed robust estimations using the influence function. The influence function is a standardized asymptotic bias of the estimator caused by one outlier. In our problem bias of $R(\hat{\theta})$ is important rather than $\hat{\theta}$ itself. Before considering the influence to $R(\hat{\theta})$, first we describe some definitions and properties related with the influence function of the estimator in the case of two samples.

6. Definitions and properties related with the influence function

The purpose of this section is to prepare some definitions and properties related with the influence function for constructing robust M-estimators used to obtain a robust discriminant rule in the following sections. In this section we modify or generalize the works included in chapter 4 of Hampel et al. [11].

Suppose that we have training samples

(6.1) $$X_1^{(j)}, X_2^{(j)}, \ldots, X_{n(j)}^{(j)}$$

from Π_j ($j = 1, 2$). The corresponding empirical distribution is given by

(6.2) $$F_{j,n(j)} = \frac{1}{n(j)} \sum_{k=1}^{n(j)} \nabla(X_k^{(j)}),$$

where $\nabla(x)$ is the point math 1 in x. We consider the estimators of θ expressed by functionals, i.e., $\hat{\theta} = T[F_{1,n(1)}, F_{2,n(2)}]$ with some functional T: domain $(T) \to \Theta$. The domain of T is the set of all pairs of distributions for which T is defined. We denote the corresponding parts of T with ζ_1, ζ_2 and ξ as $T_\zeta^{(1)}$, $T_\zeta^{(2)}$ and T_ξ, respectively. We also use the notation $T_\eta^{(j)} = (T_\zeta^{(j)}, T_\xi)$. It is said that an estimator T is Fisher consistent (Kallianpur and Rao [16]) if

(6.3) $$T[F_{1\theta}, F_{2\theta}] = \theta \qquad \text{for all } \theta \in \Theta.$$

DEFINITION 6.1. The *influence functions* of T at $[F_1, F_2]$ are defined by

$$IF_1(u, T; F_1, F_2) = (\partial/\partial h)^+ T[F_1^{u,h}, F_2] \tag{6.4}$$

and

$$IF_2(u, T; F_1, F_2) = (\partial/\partial h)^+ T[F_1, F_2^{u,h}], \tag{6.5}$$

where $(\partial/\partial h)^+$ is right derivative at $h = 0$ and $F_j^{u,h} = (1 - h)F_j + hV(u)$.

The influence function was invented by Hampel ([9], [10]) in order to investigate the infinitesimal behavior of real-valued functionals. We shortly denote the influence function at $[F_{1\theta}, F_{2\theta}]$ as $IF_j(u, T; \theta)$ $(j = 1, 2)$.

THEOREM 6.1. (Hampel et al. [11], page 196) *Let $F_{1,n(1)}$ and $F_{2,n(2)}$ be the empirical distributions of the samples from F_1 and F_2, respectively. Let $\hat{\theta} = T[F_{1,n(1)}, F_{2,n(2)}]$ and $\theta = T[F_1, F_2]$ then the limiting distribution of $N^{1/2}(\hat{\theta} - \theta)$, with increasing sample sizes and with keeping $n(1)/n(2)$ constant, is $N[0, V(T; \theta)]$, where $N = n(1) + n(2)$,*

$$V(T; \theta) = r_1^{-1} V_1(T; \theta) + r_2^{-1} V_2(T; \theta), \tag{6.6}$$

$r_j = n(j)/N$ and

$$V_j(T; \theta) = \int IF_j(u, T; \theta) IF_j(u, T; \theta)' dF(u; \eta_j) \qquad (j = 1, 2). \tag{6.7}$$

DEFINITION 6.2. Let ψ_1 and ψ_2 be functions on the product space $\Omega \times \Theta$ to Θ. Then the *M-estimator given by $\psi = [\psi_1, \psi_2]$* is defined by the implicit equations:

$$\textstyle\sum_{j=1}^{2} r_j \int \psi_j(x, T) dF_j(x) = 0. \tag{6.8}$$

For the training samples or the empirical distributions $F_{1,n(1)}$ and $F_{2,n(2)}$ the above equation with $r_j = n(j)/N$ is equivalent with

$$\textstyle\sum_{j=1}^{2} \sum_{k=1}^{n(j)} \psi_j(X_k^{(j)}; T) = 0. \tag{6.9}$$

Note that the maximum likelihood estimator is an M-estimator. This is seen by taking $\psi_j = s_j(x, \theta)$ $(j = 1, 2)$,

$$s_1(x, \theta) = \begin{pmatrix} s(x, \zeta_1) \\ 0 \\ s(x, \xi) \end{pmatrix} \quad \text{and} \quad s_2(x, \theta) = \begin{pmatrix} 0 \\ s(x, \zeta_2) \\ s(x, \xi) \end{pmatrix} \tag{6.10}$$

where $s(x, \eta) = (s(x, \zeta)', s(x, \xi)')' = (\partial/\partial\eta) \log f(x; \eta)$.

It is known that the M-estimator given by ψ is Fisher consistent if

(6.11) $$\sum_{j=1}^{2} r_j \int \psi_j(x;\theta) dF_{j\theta}(x) = 0 \quad (j = 1, 2) \text{ for all } \theta.$$

THEOREM 6.2. *Assume the equation*:

(6.12) $$(\partial/\partial\tau')\int \psi_j(x, \tau) dF_j(x) = \int \dot{\psi}_j(x, \tau) dF_j(x) \quad (j = 1, 2)$$

holds, where $(\partial/\partial\tau')g(\tau) = (\partial g/\partial\tau_1, \partial g/\partial\tau_2, \ldots, \partial g/\partial\tau_{2q+r})$ *and* $\dot{\psi}_j(x, \tau) = (\partial/\partial\tau')\psi_j(x, \tau)$, *a* $(2q + r) \times (2q + r)$ *matrix. Then the influence unction of the M-estimator is given by*

(6.13) $$IF_j(u, T; F_1, F_2) = M(\psi; F_1, F_2)^{-1} r_j\{\psi_j(u, T) - \int \psi_j(u, T) dF_j(u)\}$$

$(j = 1, 2)$, *where M is a* $(2q + r) \times (2q + r)$ *matrix defined as*:

(6.14) $$M(\psi; F_1, F_2) = -\sum_{j=1}^{2} r_j \int \dot{\psi}_j(u, T) dF_j(u)$$

and $\dot{\psi}_j(u, \tau) = (\partial/\partial\tau')\psi_j (j = 1, 2)$.

PROOF. Let $\theta^{u,h} = T[F_1^{u,h}, F_2]$ and $\theta = T[F_1, F_2]$. Then

(6.15) $$r_1 \int \psi_1(x; \theta^{u,h}) dF_1^{u,h}(x) + r_2 \int \psi_2(x; \theta^{u,h}) dF_2(x) = 0.$$

Take the derivatives of both sides at $h = 0$, then we get

(6.16) $$\begin{aligned} r_1\{\psi_1(u;\theta) &- \int \psi_1(x;\theta) dF_1(x) \\ &+ \int \dot{\psi}_1(x;\theta) dF_1(x) IF_1(u, T; F_1, F_2)\} \\ &+ r_2 \int \dot{\psi}_2(x;\theta) dF_2(x) IF_1(u, T; F_1, F_2) = 0. \end{aligned}$$

Hence we get

(6.17) $$M(\psi; F_1, F_2) IF_1(u, T; F_1, F_2) = r_1\{\psi_1(u;\theta) - \int \psi_1(x;\theta) dF_{1\theta}\},$$

which gives the desired result for $IF_1(u, T; F_1, F_2)$. The result for $IF_2(u, T; F_1, F_2)$ is similarly obtained.

LEMMA 6.1. *Suppose that T is Fisher consistent and that the equation*:

(6.18) $$(\partial/\partial\theta')\int \psi_j(x, \theta) dF_{j\theta}(x) = \int (\partial/\partial\theta')\{\psi_j(x, \theta) f(x, \eta_j)\} dx$$

and (6.11) *hold. Then it holds that*

(6.19) $$\sum_j \int IF_j(x, T; \theta) s_j(s; \theta)' dF_{j\theta}(x) = I.$$

PROOF. Take the derivative of both sides of equation in (6.11), then we get

(6.20) $$\sum_{j=1}^{2} r_j\{\int \dot{\psi}_j(x;\theta) dF_{j\theta}(x) + \int \psi_j(x;\theta) s_j(x;\theta)' dF_{j\theta}(x)\} = 0.$$

Therefore $M(\psi; \theta) = M(\psi; F_{1\theta}, F_{2\theta})$ can be expressed as:

$$(6.21) \qquad M(\psi;\theta) = \sum r_j \int \psi_j(x;\theta) s_j(x;\theta)' dF_{j\theta}(x).$$

From Theorem 6.2 and $\int s(x;\eta)dF(x;\eta) = 0$, we get the desired result.

7. Measures of efficiency and robustness of the estimators in classification

We are interesting in obtaining an estimator $R(\hat{\theta})$ of the minimax discriminant region $R(\theta)$ with certain optimalities in classification problem. First we investigate $R(\hat{\theta})$ such that $R(\hat{\theta})$ minimizes the maximum of two kind of the misclassification probabilities. For such a purpose we define a loss $L(\hat{\theta};\theta)$ of $\hat{\theta}$ at θ by

$$(7.1) \qquad \begin{aligned} L(\hat{\theta};\theta) &= \max\{F(R(\hat{\theta})^c;\eta_1), F(R(\hat{\theta});\eta_2)\} - F(R(\theta)^c;\eta_1) \\ &= \max\{F(R(\hat{\theta})^c;\eta_1), F(R(\hat{\theta});\eta_2)\} - F(R(\theta);\eta_2). \end{aligned}$$

LEMMA 7.1. *Suppose that $F(R(\tau);\eta)$ is c^1-class as a function of τ, for any η. Then the following equation holds.*

$$(7.2) \qquad [(\partial/\partial\tau)F(R(\tau)^c;\eta_1)]_{\tau=\theta} + k(\theta)[(\partial/\partial\tau)F(R(\tau);\eta_2)]_{\tau=\theta} = 0.$$

PROOF. The region $R(\theta)$ is also a Bayes rule when the prior probabilities from Π_1 and Π_2 are $1/\{1+k(\theta)\}$ and $k(\theta)/\{1+k(\theta)\}$, respectively. So that the function $F(R(\tau)^c;\eta_1) + k(\theta)F(R(\tau);\eta_2)$ is minimized at $\tau = \theta$.

THEOREM 7.1. *Let $\hat{\theta} = T[F_{1,n(1)}, F_{2,n(2)}]$. Suppose that the limiting distribution of $N^{1/2}(\hat{\theta}-\theta)$ is $N_{q+r}(0, V(T;\theta))$ and the condition of Lemma 7.1 holds. Then $N^{1/2}L(\hat{\theta};\theta)$ is asymptotically distributed as the same distribution as*

$$(7.3) \qquad \{D(\theta)'V(T;\theta)D(\theta)\}^{1/2} \max\{U, -k(\theta)U\},$$

where

$$(7.4) \qquad D(\theta) = [(\partial/\partial\tau)F(R(\tau);\eta_2)]_{\tau=\theta}$$

and U is a standard normal variable.

PROOF. Consider Taylor expansions of $F(R(\tau)^c;\eta_1)$ *and* $F(R(\tau);\eta_2)$ at $\tau = \theta$ and use Lemma 7.1. Then we get

$$(7.5) \qquad \begin{aligned} L(\hat{\theta};\theta) = \max\{&[(\partial/\partial\tau')F(R(\tau)^c;\eta_1)]_{\tau=\theta}(\hat{\theta}-\theta), \\ &[(\partial/\partial\tau')F(R(\tau);\eta_2)]_{\tau=\theta}(\hat{\theta}-\theta)\} + o_p(\|\hat{\theta}-\theta\|) \\ = \max\{&-k(\theta)D(\theta)'(\hat{\theta}-\theta), D(\theta)'(\hat{\theta}-\theta)\} + o_p(\|\hat{\theta}-\theta\|). \end{aligned}$$

From the assumption of asymptotic normality of $\hat{\theta}$ the limiting distribution of $N^{1/2}D(\theta)'(\hat{\theta}-\theta)$ is $N(0, D(\theta)'V(T;\theta)D(\theta))$, which shows the desired result.

COROLLARY 7.1. *Under the assumption of Theorem* 7.1, *the expectation of* $N^{1/2}L(\hat{\theta};\theta)$ *under the limiting distribution is given by*

$$(7.6)\qquad (2\pi)^{-1/2}\{1+k(\theta)\}\{D(\theta)'V(T;\theta)D(\theta)\}^{1/2}.$$

PROOF. The expectation is easily obtained with the use of

$$(7.7)\qquad \max\{U,\ -k(\theta)U\}=\frac{1}{2}\{1-k(\theta)\}U+\frac{1}{2}\{1+k(\theta)\}|U|.$$

(7.6) may be called as "an asymptotic risk" of the estimator. Therefore we define a measure $e^d(T;\theta)$ of efficiency of an estimator by

$$(7.8)\qquad e^d(T;\theta)=\{D(\theta)'V(T;\theta)D(\theta)\}^{-1},$$

in the situation where our purpose is to estimate the minimax regions. The superscript "d" means that the measure is defined for discrimination problem. The large value of $e^d(T;\theta)$ means small asymptotic misclassification probabilities. For an M-estimator corresponding to ψ we also denote the efficiency as $e^d(\psi;\theta)$.

Next we consider the robustness of an estimator $T[F_{1,n(1)}, F_{2,n(2)}]$. Suppose that the $n(1)$-th value $X_{n(1)}$ was an outlier. Then the influence on our loss is expressed as

$$(7.9)\qquad \begin{aligned}&\{L(T[F_{1,n(1)}, F_{2,n(2)}];\theta)-L(T[F_{1,n(1)-1}, F_{2,n(2)}];\theta)\}/\{1/n(1)\}\\ &=\left\{L\left(\left[\left(1-\frac{1}{n(1)}\right)F_{1,n(1)-1}+\frac{1}{n(1)}V(X_{n(1)}),\ F_{2,n(2)}\right];\theta\right)\right.\\ &\qquad\left.-L(T[F_{1,n(1)-1}, F_{2,n(2)}];\theta)\right\}/\{1/n(1)\},\end{aligned}$$

where the denominator means the ratio of outlier in the sample. Replacing $F_{1,n(1)-1}$ and $F_{2,n(2)}$ with their limiting distributions $F_{1\theta}$ and $F_{2\theta}$, respectively, $X_{n(1)}$ with u, and $1/n(1)$ with h, we obtain

$$(7.10)\qquad \{L(T[F_{1\theta}^{u,h}, F_{2\theta}];\theta)-L(T[F_{1\theta}, F_{2\theta}];\theta)\}/h.$$

Let h tend to zero, then we can formulate an influence function of an estimator in the situation where we want to estimate the minimax discriminant regions as follows.

DEFINITION 7.1. The *influence functions* of an estimator T at θ corresponding to Π_1 and Π_2 are defined as

$$(7.11)\qquad IF_1^d(u,T;\theta)=(\partial/\partial h)^+L(T[F_{1\theta}^{u,h}, F_{2\theta})$$

and

(7.12) $$IF_2^d(u, T; \theta) = (\partial/\partial h)^+ L(T[F_{1\theta}, F_{2\theta}^{u,h}).$$

Using the chain rule and Lemma 7.1, we get

(7.13) $$IF_j^d(u, T; \theta) = \max\{D(\theta)' IF_j(u, T; \theta), -k(\theta)D(\theta)' IF_j(u, T; \theta)\}.$$

We define a gross-error sensitivity of T at θ as

(7.14) $$\gamma_j^d(T; \theta) = \sup_u IF_j^d(u, T; \theta) \qquad (j = 1, 2).$$

Hampel et al. [11] defined three types of gross-error sensitivity, i.e., the unstandardized gross-error sensitivity, the self-standardized sensitivity and the information-standardized sensitivity, for multidimensional estimators (see [11], page 228–229). For each sensitivity, B-robustness of an estimator means that its sensitivity is finite. In our situation, $\gamma_j^d(T; \theta)$ measures a robustness of a discriminant rule obtained by using the estimator T. Therefore we say that T is D-robust if γ_s^d's are finite. If $k(\theta) = 1$, then (7.14) is reduced to

(7.15) $$\gamma_j^d(T; \theta) = \sup_u |D(\theta)' IF_j(u, T; \theta)|.$$

This suggests that the gross-error sensitivity of the estimator should be defined according to the purpose of estimation.

8. The optimal D-robust M-estimators

In the previous section we obtained a measure $\gamma_j^d(\psi; \theta)$ $(j = 1, 2)$ of the robustness and a measure $e^d(\psi; \theta)$ of the efficiency. It is impossible to obtain the M-estimator which minimizes $\gamma_j^d(\psi; \theta)$ and maximizes $e^d(\psi; \theta)$, simultaneously. Therefore we consider to maximize the efficiency $e^d(\psi; \theta)$ in certain class of φ-functions whose gross-error sensitivity $\gamma_j^d(\psi; \theta)$ is less than some given constant. We say that an M-estimator is optimal D-robust if it attains the maximum in certain class. The purpose of this section is to construct the ψ-functions which give the optimal D-robust M-estimator.

Let Ψ be a class of ψ, pairs of ψ-functions such that the conditions (6.11), (6.12) and (6.18) hold and the integral:

(8.1) $$\int \psi_j(x; \eta_j)\psi_j(x; \eta_j)' dF(x; \eta_j) \qquad (j = 1, 2)$$

exists. In this class we want to maximize the efficiency $e^d(\psi; \theta)$ subject to $\gamma_j^d(\psi; \theta) \leqq c_j$ for given constant c_j $(j = 1, 2)$. The next theorem shows that if $c_1 = c_2 = \infty$ the maximum is attained by maximum likelihood estimator of θ.

THEOREM 8.1. *Suppose that the score functions* $[s_1(x, \theta), s_2(x, \theta)]$ *belong to* Ψ. *Let* $J(\theta) = r_1 J_1(\theta) + r_2 J_2(\theta)$, *where*

$$J_j(\theta) = \int s_j(x;\theta)s_j(x;\theta)'dF(x;\eta_j) \qquad (j = 1, 2). \tag{8.2}$$

Then the asymptotic covariance matrix of the maximum likelihood estimator is $J(\theta)^{-1}$ and $V(T;\theta) - J(\theta)^{-1}$ is positive semi definite for all M-estimator T corresponding to ψ which belongs to Ψ.

PROOF. From Theorem 6.2 and Lemma 6.1, the influence function of the maximum likelihood estimator is shown to be $r_j J(\theta)^{-1} s_j(u, \theta)$. From Theorem 6.1, we obtain that the asymptotic covariance matrix is $J(\theta)^{-1}$. Let X and Y be independent random vectors distributed as $F_{1\theta}$ and $F_{2\theta}$, respectively. Define a $4(q+r)$-dimensional random vector U by

$$U = \begin{pmatrix} IF_1(X, T;\theta) \\ IF_2(Y, T;\theta) \\ s_1(X;\theta) \\ s_2(Y;\theta) \end{pmatrix}. \tag{8.3}$$

For any $(q+r)$-dimensional vector a, let

$$h = (r_1^{-1/2}a', e_2^{-1/2}a', -r_1^{1/2}a'J(\theta)^{-1}, -r_2^{1/2}a'J(\theta)^{-1})'. \tag{8.4}$$

From (6.7) the covariance matrix of U is given by

$$\mathrm{Cov}(U) = \tag{8.5}$$

$$\begin{pmatrix} V_1(T;\theta) & & & \text{symmetric} \\ 0 & V_2(T;\theta) & & \\ \int IF_1(X, T;\theta)s_1(X;\theta)'dF_{1\theta}(X) & 0 & J_1(\theta) & \\ 0 & \int IF_2(X, T;\theta)s_2(X;\theta)'dF_{2\theta}(X) & 0 & J_2(\theta) \end{pmatrix}$$

From Theorem 6.1 and (6.19) we obtain

$$\begin{aligned} h'\,\mathrm{Cov}(U)h &= r_1^{-1}a'V_1(T;\theta)a + r_2^{-1}a'V_2(T;\theta)a \\ &\quad + r_1 aJ(\theta)^{-1}J_1(\theta)J(\theta)^{-1}a' + r_2 a'J(\theta)^{-1}J_2(\theta)J(\theta)^{-1}a \\ &\quad - 2a'\{\textstyle\int IF_1(x, T;\theta)s_1(x;\theta)'dF_{1\theta}(x)\}J(\theta)^{-1}a \\ &\quad - 2a'\{\textstyle\int IF_2(x, T;\theta)s_2(x;\theta)'dF_{2\theta}(x)\}J(\theta)^{-1}a \\ &= a'V(T;\theta)a - a'J(\theta)^{-1}a \geqq 0. \end{aligned} \tag{8.6}$$

For finite c_j $(j = 1, 2)$, if the score functions are not bounded we must modify the maximum likelihood estimator. Let θ be an arbitrary fixed point in Θ and T be a M-estimator given by some pair ψ in Ψ. We use abbreviations in the rest of this section as $F_j(x) = F_{j\theta}(x)$, $s_j(x) = s_j(x, \theta)$ and $IF_j(x) = IF_j(x, T; \theta)$ $(j = 1, 2)$.

From Theorem 6.2 and Lemma 6.1, it is shown that IF_j's must satisfy

$$\int IF_j(x)\, dF_j(x) = 0 \qquad (j = 1, 2) \tag{8.9}$$

and

$$\sum_j IF_j(x) s_j(x)' dF_j(x) = I. \tag{8.10}$$

Let A be an arbitrary $(2q + r) \times (2q + r)$ matrix and let a_j $(j = 1, 2)$ be any vectors. Then using (8.9) and (8.10) we have

$$\begin{aligned} &\sum_j r_j^{-1} \int |D(\theta)' \{IF_j(u) - r_j A(s_j(u) - a_j)\}|^2\, dF_j(u) \\ &= \sum_j r_j^{-1} D(\theta)' \int \{IF_j(u) IF_j(u)' + r_j^2 A(s_j(u) - a_j)(s_j(u) - a_j)' A' \\ &\quad - r_j IF_j(u)(s_j(u) - a_j)' A' - r_j A(s_j(u) - a_j) IF_j(u)'\}\, dF_j(u) D(\theta) \\ &= \sum_j D(\theta)' [r_j^{-1} V_j(T; \theta) + r_j A \{J_j(\theta) + a_j a_j'\} \\ &\quad - \int IF_j(u) s_j(u)' dF_j(u) A' - A \int s_j(u) IF_j(u)' dF_j(u)] D(\theta) \\ &= e^d(T; \theta)^{-1} + D(\theta)' \{AJ(\theta)A' - A' + A + r_1 a_1 a_1' + r_2 a_2 a_2'\} D(\theta). \end{aligned} \tag{8.11}$$

Therefore the maximization of $e^d(T; \theta)$ with respect to T is equivalent with the minimization of

$$\sum_j r_j^{-1} \int |D(\theta)' \{IF_j(u) - r_j A(s_j(u) - a_j)\}|^2\, dF_j(u) \tag{8.12}$$

with respect to $IF_1(u)$ and $IF_2(u)$, the influence functions of T. The condition $\gamma_j^d(T; \theta) \leqq c_j$ is written as

$$-c_j/k(\theta) \leqq D(\theta)' IF_j(u) \leqq c_j \qquad \text{for all } u. \tag{8.13}$$

Therefore the minimum is attained if

$$D(\theta)' IF_j(u) = h[r_j D(\theta)' A\{s_j(u) - a_j\}; c_j, -c_j/k(\theta)], \tag{8.14}$$

where h is a translated Huber function defined as

$$h(x; \alpha, \beta) = \begin{cases} \alpha & \text{if } \alpha < x \\ x & \text{if } \beta < x \leqq \alpha. \\ \beta & \text{if } x < \beta \end{cases} \tag{8.15}$$

If $\beta = -\alpha$ then (8.15) agrees with the original Huber function. The following theorem gives a way of constructing ψ-functions whose influence functions satisfy (8.14).

THEOREM 8.2. *Define* $\psi_j(x; A, a_j)$ $(j = 1, 2)$ *by*

$$\psi_j(x; A, a_j) = h_g[\{I - P(\theta)\}A\{s_j(x) - a_j\}; \tilde{c}_j/r_j] + D(\theta)\{D(\theta)'D(\theta)\}^{-1}h[D(\theta)'A\{s_j(x) - a_j\}; c_j/r_j, -c_j/\{r_j k(\theta)\}], \tag{8.16}$$

where $\tilde{c}_j$ $(j = 1, 2)$ *is appropriately chosen constant,* $P(\theta)$ *is a projection matrix given by*

$$P(\theta) = D(\theta)\{D(\theta)'D(\theta)\}^{-1}D(\theta)' \tag{8.17}$$

and h_g *is a generalized Huber function in* R^{2q+r} *defined as*

$$h_g(U; c) = U \cdot \min\{1, c/\|U\|\}. \tag{8.18}$$

If a system of equations for A, a_1 *and* a_2:

$$\int \psi_j(x; A, a_j)dF_{j\theta}(x) = 0 \qquad (j = 1, 2), \tag{8.19}$$

$$\sum_j r_j \int \psi_j(x; A, a_j)s_j(x; \theta)'dF_{j\theta}(x) = I \tag{8.20}$$

has a solution, $A = A_\theta$, $a_j = a_{j\theta}$ $(j = 1, 2)$, *then* $\psi = [\psi_1(x; A_\theta, a_{2\theta}), \psi_2(x; A_\theta, a_{2\theta})]$ *gives the M-estimator which maximize the efficiency.*

PROOF. From Theorem 6.2 and Lemma 6.1, we obtain that the influence function $IF_j(x, T; \theta)$ is equal to ψ_j $(j = 1, 2)$ which is constructed as to satisfy (8.14).

We note that the first term of (8.18) has no effect on either the efficiency and the gross-error sensitivity in discrimination, but for finite samples, both the risk and the influence of each sample point depend not only on the second term but also on the first term.

If $k(\theta) = 1$, (8.16) is written as

$$\psi_j(x; A, a_j) = \{I - P(\theta)\}A\{s_j(x) - a_j\}W_j^o(x; A, a_j) + P(\theta)A\{s_j(x) - a_j\}W_j^d(x; A, a_j) \qquad (j = 1, 2), \tag{8.21}$$

where

$$W_j^o(x; A, a_j) = \min[1, \tilde{c}_j/\|r_j\{I - P(\theta)\}A\{s_j(x) - a_j\}\|] \tag{8.22}$$

and

$$W_j^d(x; A, a_j) = \min[1, c_j/|r_j D(\theta)'A\{s_j(x) - a_j\}|]. \tag{8.23}$$

Using

$$M_j^o = \int \{s_j(x) - a_j\}\{s_j(x) - a_j\}'W_j^o(x; A, a_j)dF_{j\theta}(x) \tag{8.24}$$

and

$$M_j^d = \int \{s_j(x) - a_j\}\{s_j(x) - a_j\}' W_j^d(x; A, a_j) dF_{j\theta}(x) \qquad (j = 1, 2), \tag{8.25}$$

the system of equation (8.19) and (8.20) can be written as

$$\textstyle\sum_j r_j [\{I - P(\theta)\} A M_j^o + P(\theta) A M_j^d] = I \tag{8.26}$$

and

$$\begin{aligned} Aa_j = \{I - P(\theta)\} A \int s_j(x) W_j^o(x; A, a_j) dF_{j\theta}(x) / \int W_j^o(x; A, a_j) dF_{j\theta}(x) \\ + P(\theta) A \int s_j(x) W_j^d(x; A, a_j) dF_{j\theta}(s) / \int W_j^d(x; A, a_j) dF_{j\theta}(x). \end{aligned} \tag{8.27}$$

Since $W_j^o(x; A, a_j)$ depends only on $\{I - P(\theta)\}A$ and $\{I - P(\theta)\}Aa_j$, and $W_j^d(x; A, a_j)$ depends only on $P(\theta)A$ and $P(\theta)Aa_j$, we can divide the system of equation and the estimation equation into orthogonal-part and discriminant-part as in the following lemma.

LEMMA 8.1. *If a system of equation for A^o and a_j^o $(j = 1, 2)$:*

$$\textstyle\sum_j r_j A^o M_j^o = I, \tag{8.28}$$

$$a_j^o = \int s_j(x) W_j^o(x; A, a_j) dF_{j\theta}(x) / \int W_j^o(x; A, a_j) dF_{j\theta}(x) \tag{8.29}$$

has a solution, and a system of equation for A^d and a_j^d $(j = 1, 2)$:

$$\textstyle\sum_j r_j A^d M_j^d = I, \tag{8.30}$$

$$a_j^d = \int s_j(x) W_j^d(x; A, a_j) dF_{j\theta}(x) / \int W_j^d(x; A, a_j) dF_{j\theta}(x) \tag{8.31}$$

has a solution, then a solution for (8.26) *and* (8.27) *is given by*

$$A = \{I - P(\theta)\} A^o + P(\theta) A^d, \quad Aa_j = \{I - P(\theta)\} Aa_j^o + P(\theta) Aa_j^d. \tag{8.32}$$

Further, the estimation equation (6.8) *for $\psi_j(x; A, a_j)$ $(j = 1, 2)$ is equivalent with*

$$\{I - P(\theta)\} \textstyle\sum_j r_j \int A^o \{s_j(x) - a_j^o\} W_j^o(x; A^o, a_j^o) dF_j(x) = 0 \tag{8.33}$$

and

$$P(\theta) \textstyle\sum_j r_j \int A^d \{s_j(x) - a_j^d\} W_j^d(x; A^d, a_j^d) dF_j(x) = 0. \tag{8.34}$$

9. Equivariant *M*-estimator

Consider a group of transformation on the sample space:

$$\mathscr{A} = \{\alpha : \Omega \longrightarrow \Omega\}. \tag{9.1}$$

Suppose the model $\{F(x; \eta); \eta \in H\}$ is invariant under $\mathscr{A}$, that is, every $\alpha \in \mathscr{A}$

and $\eta \in H$ determine a unique element in H, denoted by $\bar{\alpha}\eta$, such that $\tilde{\alpha}F(\ ;\eta) = F(\ ;\bar{\alpha}\eta)$, where $\tilde{\alpha}F$ is the distribution of αX with X being distributed as F. We denote $\bar{\alpha} = (\bar{\alpha}^{(\zeta)'}, \bar{\alpha}^{(\xi)'})'$ corresponding to $\eta = (\zeta', \xi')'$. We assume that $\bar{\alpha}^{(\xi)}\eta$ depends only on ξ. Then we can define the transformation g_α associated with α on whole parameter space Θ as

$$g_\alpha \theta = \begin{pmatrix} \bar{\alpha}^{(\zeta)}\eta_1 \\ \bar{\alpha}^{(\zeta)}\eta_2 \\ \bar{\alpha}^{(\xi)}\xi \end{pmatrix}. \tag{9.2}$$

We assume that g_α is differentiable with θ.

We say that an estimator T is equivariant if $T[\tilde{\alpha}F_1, \tilde{\alpha}F_2] = g_\alpha T[F_1, F_2]$ for all α.

LEMMA 9.1. (Hampel et al. [11], page 259) *If T is equivariant then*

$$IF_j(\alpha u, T; \tilde{\alpha}F_1, F_2) = [\partial g_\alpha/\partial\theta']IF_j(u, T; F_1, F_2) \qquad (j = 1, 2), \tag{9.3}$$

where $[\partial g_\alpha/\partial\theta']$ is the derivative of $g_\alpha\tau$ at $\tau = \theta$.

PROOF. Because of $(\tilde{\alpha}F)^{\alpha u, h} = \alpha(F^{u,h})$ and equivariance, we get

$$T[(\tilde{\alpha}F_1)^{\alpha u, h}, \tilde{\alpha}F_2] = g_\alpha T[F_1^{u,h}, F_2]. \tag{9.4}$$

Take the right derivative of both sides with using chain rule, then we get the desired result.

THEOREM 9.1. *The efficiency and the gross-error sensitivities defined by* (7.8) *and* (7.14), *respectively, are invariant if T is equivariant.*

PROOF. The misclassification probability of Π_2 can be written as

$$\begin{aligned} F(R(\tau); \eta_2) &= \Pr\{\alpha X \in \alpha R(\tau); \eta_2\} \\ &= F\{\alpha R(\tau); \bar{\alpha}\eta_2\}. \end{aligned} \tag{9.5}$$

So that $\alpha R(\theta)$ gives the minimax region for $F_{1,g_\alpha\theta}$ and $F_{2,g_\alpha\theta}$, which implies $R(g_\alpha\theta) = \alpha R(\theta)$ with probability 1. If the support of $f(x;\eta)$ does not depend on η, then

$$F(R(\tau); \eta_2) = F(R(g_\alpha\tau); \bar{\alpha}\eta_2). \tag{9.6}$$

Take the derivative of both sides at $\tau = \theta$, then we get

$$D(\theta) = [\partial g_\alpha/\partial\theta']D(g_\alpha\theta). \tag{9.7}$$

From Lemma 9.1

(9.8) $$D(\theta)' IF_j(u, T; \theta) = D(g_\alpha \theta)' IF_j(\alpha u, T; g_\alpha \theta) \qquad (j = 1, 2),$$

which implies the invariance of $e^d(T; \theta)$ and $\gamma_j^d(T; \theta)$.

The M-estimator corresponding to ψ is equivalent if the equation

(9.9) $$\sum_{j=1}^{2} r_j \int \psi_j(x, \tau) dF_j(x) = 0$$

is equivalent with

(9.10) $$\sum_{j=1}^{2} r_j \int \psi_j(\alpha x, g_\alpha \tau) dF_j(x) = 0.$$

Let $\delta(\theta)$ be the maximal invariant function on Θ under the group of transformation $\mathscr{G} = \{g_\alpha; \alpha \in \mathscr{A}\}$. Let $\Theta_\delta = \{\theta \in \Theta; \delta(\theta) = \delta\}$, the orbit of Θ such that the value of the maximal invariant function is δ. Let θ_δ be an arbitrary fixed element of Θ. Then there is a transformation $\alpha (= \alpha_\theta$, say) on Ω such that $g_\alpha \theta = \theta_{\delta(\theta)}$.

THEOREM 9.2. *For ψ which defines an equivariant M-estimator, define $\varphi = [\varphi_1, \varphi_2]$ by*

(9.11) $$\varphi_j(x, \theta) = \psi_j(\alpha_\theta x, \theta_{\delta(\theta)}) \qquad (j = 1, 2),$$

then φ defines the same M-estimator as ψ.

PROOF. Let τ be a solution of (9.9). Substitution of $\alpha = \alpha_\tau$ in (9.10) gives

(9.12) $$\sum_{j=1}^{2} r_j \int \varphi_j(x, \tau) dF_j(x) = 0.$$

Similarly a solution of (9.12) is also a solution of (9.9).

The way of constructing the optimal D-robust equivalent M-estimator can be partitioned to three steps.

Step 1. Find the maximal invariant δ and choose an appropriate set of θ_δ.

Step 2. For each θ_δ, compute the likelihood scores, and calculate the matrix $A(\theta_\delta)$ and $a_j(\theta_\delta)$ $(j = 1, 2)$ described in Theorem 8.2.

Step 3. By using Theorem 9.2, define the ψ-functions for all θ.

If $k(\theta) = 1$, then Lemma 8.1 is useful in reducing the ψ-functions to simple forms.

10. The optimal D-robust M-estimators in elliptical opulations

In this section we construct the optimal D-robust equivariant M-estimators in elliptical populations along the steps described in the previous section.

We return to the elliptical model (1.1) considered in Part I. Since the symmetric matrix Λ contains redundance, we parametrize the model as

(10.1) $$f(x;\eta_j) = |\Lambda|^{-1/2}h\{(x-\mu_j)'\Lambda^{-1}(x-\mu_j)\} \qquad (j=1,2),$$

where $\eta_j = \{\mu_j', \operatorname{vecs}(\Lambda)'\}'$, with operator vecs defined as follows (Hampel et al. [11], page 272):
If S is a symmetric matrix, let $\operatorname{vecs}(S)$ be the vector

(10.2) $$\operatorname{vecs}(S) = (s_{11}/2^{1/2},\dots,s_{pp}/2^{1/2}, s_{21}, s_{31},\dots,s_{p,p-1})'.$$

The whole parameter is $\theta = (\mu_1', \mu_2', \operatorname{vecs}(\Lambda)')'$. For any $2p + p(p+1)/2$-dimensional vector α, p-dimensional vector β_1, β_2 and $p \times p$ symmetric matrix Γ, we often use the notation $\alpha = (\beta_1, \beta_2, \Gamma)$ instead of writing $\alpha = (\beta_1', \beta_2', \operatorname{vecs}(\Gamma)')'$.

From Theorem 1.2, we obtain that $k(\theta) = 1$, the minimax discriminant region $R(\theta)$ is given by

(10.3) $$R(\theta) = \{x; (x-\bar{\mu})'\Lambda^{-1}(\mu_1-\mu_2) > 0\},$$

and

(10.4) $$F(R(\tau);\eta_2) = P_2(\tau;\theta) = Q\{c_2(\tau,\theta)\},$$

where $c_2(\tau, \theta)$ is given by (1.3) and Q is the distribution function whose desity function is given by (1.4). From Lemma 3.1, the derivative of $c_2(\tau, \theta)$ is given by

(10.5) $$[(\partial/\partial\tau)c_2(\tau,\theta)]_{\tau=\theta} = -\frac{1}{2}[\xi, \xi, \mathbf{0}],$$

where

(10.6) $$\xi = \Delta^{-1}\Lambda^{-1/2}(\mu_1-\mu_2).$$

The results (10.4) and (10.5) imply

(10.7) $$D(\theta) = -\frac{1}{2}q_1[\xi, \xi, \mathbf{0}],$$

where q_1 is given by (3.9).

The model is invariant under the affine group of transformation $\mathscr{A}$ on the sample space $\Omega = R^p$, where

(10.8) $$\mathscr{A} = \{\alpha = (L, b); L \text{ is a } p\times p \text{ nonsingular matrix}, b\in R^p\},$$

and $ax = Lx + b$ for $x \in \Omega$. The induced transformation on $H = \{\eta\}$ and Θ are $\bar{\alpha}\eta = (L\mu + b, L\Lambda L')$ and $g_\alpha\theta = (L\mu_1 + b, L\mu_2 + b, L\Lambda L')$, respectively. It is known that the maximal invariant of Θ under G is

(10.9) $$\Delta^2 = (\mu_1-\mu_2)'\Lambda^{-1}(\mu_1-\mu_2)$$

(cf. Muirhead [22] page 220). Θ is partitioned by G as

$$\Theta = \bigcup_{\Delta > 0} \Theta_\Delta. \tag{10.10}$$

where each orbit is defined as

$$\Theta_\Delta = \{\theta = (\mu_1, \mu_2, \Lambda); (\mu_1 - \mu_2)'\Lambda^{-1}(\mu_1 - \mu_2) = \Delta^2\}. \tag{10.11}$$

For each orbit Θ_Δ's, we choose $\theta_\Delta = (\delta, 0, I_p)$, where $\delta = (\Delta, 0, \dots, 0)'$. We denote $\eta_\delta = (\delta, I_p)$ and $\eta_0 = (0, I_p)$.

The transformation α_θ such that the induced transformation transforms θ to θ_Δ, defined in the previous section, is

$$\alpha_\theta = (H\Lambda^{-1/2}, \ -H\Lambda^{-1/2}\mu_2), \tag{10.12}$$

where H is an orthogonal matrix whose first row is ξ', where ξ is given by (10.6). We shortly denote the induced transformation by α_θ as g_θ. From (10.7) and (10.6) the projection matrix $P(\theta_\Delta)$ is given by

$$P(\theta_\Delta) = \frac{1}{2}\begin{pmatrix} U & U & 0 \\ U & U & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{10.13}$$

where U is a $p \times p$ matrix whose 1–1 element is 1 and other all elements are 0.

The score function $s(x, \eta)$ is given by

$$s(x, \eta) = [\Lambda^{-1}(x - \mu)w(v), \Lambda^{-1}(x - \mu)(x - \mu)'\Lambda^{-1}w(v) - \Lambda^{-1}], \tag{10.14}$$

where

$$v = (x - \mu)'\Lambda^{-1}(x - \mu) \tag{10.15}$$

and

$$w(v) = -2(d/dv)\{\log h(v)\}. \tag{10.16}$$

Therefore we obtain $s_1(x, \theta_\Delta) = s_1(x - \delta, \theta_0)$ and $s_2(x, \theta_\Delta) = s_2(x, \theta_0)$, where

$$s_1(z, \theta_0) = \begin{pmatrix} zw(z'z) \\ 0 \\ \operatorname{vecs}\{zz'w(z'z) - I_p\} \end{pmatrix} \tag{10.17}$$

and

$$s_2(x, \theta_0) = \begin{pmatrix} 0 \\ zw(z'z) \\ \operatorname{vecs}\{zz'w(z'z) - I_p\} \end{pmatrix}. \tag{10.18}$$

Let the optimal estimation function be $[\psi_1, \psi_2]$. From Theorem 8.2 and Lemma 8.1 we obtain

$$\begin{aligned} \psi_j(x, \theta_\Delta) &= P(\theta_\Delta)A^d\{s_j(x, \theta_\Delta) - a_j^d\} W_j^d(x, \theta_\Delta) \\ &\quad + \{I - P(\theta_\Delta)\}A^o\{s_j(x, \theta_\Delta) - a_j^o\} W_j^o(x, \theta_\Delta) \end{aligned} \tag{10.19}$$

where

$$W_j^d(x, \theta_\Delta) = \min[1, c_j/|r_j D(\theta_\Delta)' A^d\{s_j(x, \theta_\Delta) - a_j^d\}|] \tag{10.20}$$

and

$$W_j^o(x, \theta_\Delta) = \min[1, \tilde{c}_j/\| r_j\{I - P(\theta_\Delta)\}A^o\{s_j(x, \theta_\Delta) - a_j^o\}\|] \tag{10.21}$$

$(j = 1, 2)$, and where A^h, a_j^h $(h = d, o; j = 1, 2)$ is a solution of the system of equations:

$$a_j^h = \int s_j(x, \theta_\Delta) W_j^h(x; \theta_\Delta) dF_{j\theta_\Delta}(x) / \int W_j^h(x; \theta_\Delta) dF_{j\theta_\Delta}(x) \tag{10.22}$$

and

$$P(\theta_\Delta)A^d \textstyle\sum_j r_j M_j^d = P(\theta_\Delta), \quad \{I - P(\theta_\Delta)\}A^o \textstyle\sum_j r_j M_j^o = I - P(\theta_\Delta) \tag{10.23}$$

with

$$M_j^h = \int \{s_j(x, \theta_\Delta) - a_j^h\}\{s_j(x, \theta_\Delta) - a_j^h\}' W_j^h(x, \theta_\Delta) dF_{j\theta_\Delta}(x). \tag{10.24}$$

Let

$$c_j^d = c_j/\{r_j \| D(\theta_\Delta)\|\} = 2^{1/2} c_j/(r_j q_j) \quad \text{and} \quad c_j^o = \tilde{c}_j/r_j \quad (j = 1, 2). \tag{10.25}$$

Then the weighting function $W_j^h(x, \theta_\Delta)$'s are written as

$$W_1^h(x, \theta_\Delta) = \tilde{W}_1^h(x - \delta, \theta_0) \tag{10.26}$$

and

$$W_2^h(x, \theta_\Delta) = \tilde{W}_2^h(x, \theta_0) \qquad (h = d, o) \tag{10.27}$$

where

$$\tilde{W}_j^d(z, \theta_0) = \min[1, c_j^d/\| P(\theta_\Delta)A^d\{s_j(z, \theta_0) - a_j^d\}\|\} \tag{10.28}$$

and

$$\tilde{W}_j^o(z, \theta_0) = \min[1, c_j^o/\|\{I - P(\theta_\Delta)\}A^o\{s_j(z, \theta_0) - a_j^o\}\|\}. \tag{10.29}$$

The system of equations for A^h and $a_j^h(h = d, o; j = 1, 2)$ are given by

(10.30) $$a_j^h = \int s_j(z, \theta_0)\tilde{W}_j^h(z; \theta_0)dF(z; \eta_0)/\int \tilde{W}_j^h(z; \theta_0)dF(z; \eta_0)$$

and

(10.31) $$P(\theta_\Delta)A^d\sum_j r_j M_j^d = P(\theta_\Delta), \quad \{I - P(\theta_\Delta)\}A^o\sum_j r_j M_j^o = I - P(\theta_\Delta)$$

with

(10.32) $$M_j^h = \int \{s_j(z, \theta_0) - a_j^h\}\{s_j(z, \theta_0) - a_j^h\}'\tilde{W}_j^h(z; \theta_0)dF(z; \eta_0).$$

Therefore A^h and a_j^h $(h = d, o; j = 1, 2)$ do not depend on Δ for given c_j^d and c_j^a $(j = 1, 2)$.

The optimal values of A^h and a_j^h $(h = d, o; j = 1, 2)$ may be calculated by following iterative method.

The j-th influence function $(j = 1, 2)$ of the maximal likelihood estimator at $F(\ ; \eta_\delta)$ and $F(\ ; \eta_0)$ is given by $r_j J(\theta_\Delta)^{-1}s_j(x, \theta_\Delta)$ where $J(\theta_\Delta) = r_1 J_1(\theta_\Delta) + r_2 J_2(\theta_\Delta)$ with

(10.33) $$J_j(\theta_\Delta) = \int s_j(z, \theta_0)s_j(z, \theta_0)'dF(z; \eta_0).$$

Therefore we set the starting values $A^{h(0)} = J(\theta_\Delta)^{-1}$ and $a_j^{h(0)} = 0$ $(h = d, o; j = 1, 2)$. For the k-th value $A^{h(k)}$ and $a_j^{h(k)}$, the k-th weighting functions are defined as

(10.34) $$\tilde{W}_j^{d(k)}(z, \theta_0) = \min[1, c_j^d/\| P(\theta_\Delta)A^{d(k)}\{s_j(z, \theta_0) - a_j^{d(k)}\}\|\}$$

and

(10.35) $$\tilde{W}_j^{o(k)}(z, \theta_0) = \min[1, c_j^o/\| \{I - P(\theta_\Delta)\}A^{o(k)}\{s_j(z, \theta_0) - a_j^{o(k)}\}\|\}.$$

The $(k + 1)$th values are given by

(10.36) $$a_j^{h(k+1)} = \int s_j(z, \theta_0)\tilde{W}_j^{h(k)}(z, \theta_0)dF_j(z; \eta_0)/\int \tilde{W}_j^{h(k)}(z, \theta_0)dF_j(z; \eta_0)$$

and

(10.37) $$A^{h(k+1)} = [\sum r_j M_j^{h(k+1)}]^{-1}.$$

where

(10.38) $$M_j^{h(k+1)} = \int \{s_j(z, \theta_0) - a_j^{h(k+1)}\}\{s_j(z, \theta_0) - a_j^{h(k+1)}\}'\tilde{W}_j^{h(k)}(z, \theta_0)dF_j(z; \eta_0)$$

$(h = d, o; j = 1, 2)$.

Actual calculations on the first few cycles of the above iterative process show that $\tilde{W}_j^{h(k)}(z, \theta_0)$ is a function of z_1^2 and $\|z_2\|^2$ where $z = (z_1, z_2')'$ with

z_2 is a $(p-1) \times 1$ vector, and that $a_j^{h(k)} = [0, 0, (\alpha_j^{h(k)} - 1)I_p]$ and $A^{h(k)} = \text{Diag}\{a_{\mu 1,1}^{h(k)}, a_{\mu 1,2}^{h(k)} I_{p-1}, a_{\mu 2,1}^{h(k)}, a_{\mu 2,2}^{h(k)} I_{p-1}, A_\Lambda^{h(k)}\}$ with some constants $\alpha_j^{h(k)}$, $a_{\mu 1,1}^{h(k)}$, $a_{\mu 1,2}^{h(k)}$, $a_{\mu 2,1}^{h(k)}$, $a_{\mu 2,2}^{h(k)}$ and a semi-d-type matrix $A_\Lambda^{h(k)}$. Here $\text{Diag}\{B_1, B_2, \ldots, B_k\}$, for scalar or square matrix B_j's $(j = 1, 2, \cdots k)$, means a square matrix given by

$$\text{Diag}\{B_1, B_2, \cdots B_k\} = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & B_k \end{pmatrix}, \tag{10.39}$$

and the semi-d-type matrix is defined as follows.

DEFINITION 10.1. The *semi-d-type matrix D* of order $p \geqq 2$, given by five numbers $d_{\lambda,1}$, $d_{\lambda,2}$, d_ν, d_ρ and d_τ, is defined as

$$D = \begin{pmatrix} D_1 & 0 & 0 \\ 0 & d_{\lambda,1} I_{p-1} & 0 \\ 0 & 0 & d_{\lambda,2} I_{(p-1)(p-2)/2} \end{pmatrix} \begin{matrix} \} \; p \\ \} \; (p-1) \\ \} \; (p-1)(p-2)/2 \end{matrix} \tag{10.40}$$

where D_1 is a $p \times p$ matrix given by

$$D_1 = \begin{pmatrix} d_\nu & d_\rho & \cdots & \cdots & d_\rho \\ d_\rho & d_{\nu,2} & d_{\rho,2} & \cdots & d_{\rho,2} \\ \vdots & d_{\rho,2} & d_{\nu,2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & d_{\rho,2} \\ d_\rho & d_{\rho,2} & \cdots & d_{\rho,2} & d_{\nu,2} \end{pmatrix} \tag{10.41}$$

with

$$d_{\nu,2} = d_{\lambda,2} + d_{\rho,2}, \quad d_{\rho,2} = \frac{1}{p-1}(d_\tau - d_{\lambda,2}). \tag{10.42}$$

We use the notation $A = D^*[d_{\lambda,1}, d_{\lambda,2}, d_\nu, d_\rho, d_\tau]$ which means that A is a semi-d-type matrix given by $d_{\lambda,1}$, $d_{\lambda,2}$, d_ν, d_ρ and d_τ.

If the sequence $[a_1^{h(k)}, a_2^{h(k)}, A^{h(k)}]$ converges to $[a_1^h, a_2^h, A^h]$ for $h = d$ and o, then a_j^h and A^h $(h = d, o; j = 1, 2)$ are the optimal values and have the same form as $a_j^{h(k)}$ and $A^{h(k)}$, respectively. Therefore we guess that $a_j^h = [0, 0, (a_j^h - 1)I_p]$ and $A^h = \text{Diag}\{a_{\mu 1,1}^h, a_{\mu 1,2}^h I_{p-1}, a_{\mu 2,1}^h, a_{\mu 2,2}^h I_{p-1}, A_\Lambda^h\}$ with some constans α_j^h, $a_{\mu k,l}^h$'s $(k, l = 1, 2)$ and $A_\Lambda^h = D^*[a_{\lambda,1}^h, a_{\lambda,2}^h, a_\nu^h, a_\rho^h, a_\tau^h]$. In order to simplify the ψ-functions we prepare the following lemma which is a modified vergion of Lemma 1 given by Hample et al. [11] (page 276).

LEMMA 10.1. *Let $\Lambda(z)$ be a $p \times p$ symmetric matrix whose elements are functions of $z = (z_1, z_2')'$, where z_1 is a scalar and z_2 is a $(p-1)$-dimensional vector. If $\Lambda(z)$ satisfies the equation*

$$\Lambda(\tilde{\Gamma} z) = \tilde{\Gamma}\Lambda(z)\tilde{\Gamma}', \tag{10.43}$$

where $\tilde{\Gamma} = \mathrm{Diag}\{1, \Gamma\}$ for all $(p-1) \times (p-1)$ orthogonal matrix Γ. then

$$\Lambda(z) = \begin{pmatrix} \psi_{\Lambda,1}(z_1, z_2'z_2) & \text{symmetric} \\ z_2\psi_{\Lambda,2}(z_1, z_2'z_2) & z_2z_2'\psi_{\Lambda,3}(z_1, z_2'z_2) - I\psi_{\Lambda,4}(z_1, z_2'z_2) \end{pmatrix} \tag{10.44}$$

for some functions $\psi_{\Lambda,k}(x, y)$'s ($k = 1, 2, 3$ and 4).

Using the fact that $aI - bz_2z_2'$ has the latent roots a of multiplicity $p-2$ and $a - bz_2'z_2$, the norm of vecs $\{\Lambda(z)\}$ is given by

$$\begin{aligned} \|\mathrm{vecs}\,\{\Lambda(z)\}\|^2 &= \frac{1}{2}\mathrm{tr}\,\{\Lambda(z)^2\} \\ &= \frac{1}{2}[\psi_{\Lambda,1}^2 + z_2'z_2\psi_{\Lambda,2}^2 + (p-2)\psi_{\Lambda,4}^2 + \{\psi_{\Lambda,4} - z_2'z_2\psi_{\Lambda,3}\}^2]. \end{aligned} \tag{10.45}$$

LEMMA 10.2. *Let $\Lambda(z)$ has the form given by Lemma 10.1, and $A = D^*[a_{\lambda,1}, a_{\lambda,2}, a_\nu, a_\rho, a_\tau]$. Let $\Xi(z)$ be a symmetric matrix given by* vecs $\{\Xi(z)\} = A$ vecs $\{\Lambda(z)\}$. *Then $\Xi(z)$ has the same form as one of $\Lambda(z)$ with $\psi_{\Xi,k}$'s ($k = 1, 2, 3$ and 4), where*

$$\psi_{\Xi,1}(x, y) = a_\nu\psi_{\Lambda,1}(x, y) + a_\rho\{y\psi_{\Lambda,3}(x, y) - (p-1)\psi_{\Lambda,4}(x, y)\}, \tag{10.46}$$

$$\psi_{\Xi,2}(x, y) = a_{\lambda,1}\psi_{\Lambda,2}(x, y), \quad \psi_{\Xi,3}(x, y) = a_{\lambda,2}\psi_{\Lambda,3}(x, y) \tag{10.47}$$

and

$$\begin{aligned} \psi_{\Xi,4}(x, y) = &- a_\rho\psi_{\Lambda,1}(x, y) + \frac{1}{p-1}a_{\lambda,2}y\psi_{\Lambda,3}(x, y) \\ &- \frac{1}{p-1}a_\tau\{y\psi_{\Lambda,3}(x, y) - (p-1)\psi_{\Lambda,4}(x, y)\} \end{aligned} \tag{10.48}$$

From (10.17), (10.18) and the above lemma,we obtain

$$A^h\{s_1(z, \theta_0) - a_1^h\} = \begin{pmatrix} \mu_1^h(z) \\ 0 \\ \mathrm{vecs}\,\{\Lambda_1^h(z)\} \end{pmatrix} \tag{10.49}$$

and

$$(10.50) \qquad A^h\{s_2(z, \theta_0) - a_2^h\} = \begin{pmatrix} 0 \\ \mu_2^h(z) \\ \text{vecs}\,\{\Lambda_2^h(z)\} \end{pmatrix},$$

where

$$(10.51) \qquad \mu_j^h(z) = \begin{pmatrix} a_{\mu j,1}^h z_1 w(z'z) \\ a_{\mu j,2}^h z_2 w(z'z) \end{pmatrix} \qquad (j = 1, 2)$$

and

$$(10.52) \qquad \Lambda_j^h(z) = \begin{pmatrix} \psi_{\Lambda j,1}^h(z) & \text{symmetric} \\ z_2 \psi_{\Lambda j,2}^h(z) & z_2 z_2' \psi_{\Lambda j,3}^h(z) - I \psi_{\Lambda j,4}^h(z) \end{pmatrix}$$

with

$$(10.53) \qquad \psi_{\Lambda j,1}^h(z) = a_\nu^h\{z_1^2 w(z'z) - \alpha_j^h\} + a_\rho^h\{z_2' z_2 w(z'z) - (p-1)\alpha_j^h\},$$

$$(10.54) \qquad \psi_{\Lambda j,2}^h(z) = a_{\lambda,1}^h z_1 w(z'z), \quad \psi_{\Lambda j,3}^h(z) = a_{\lambda,2}^h w(z'z)$$

and

$$(10.55) \qquad \psi_{\Lambda j,4}^h(z) = -\, a_\rho^h\{z_1^2 w(z'z) - a_j^h\} - \frac{1}{p-1} a_\tau^h\{z_2' z_2 w(z'z) - (p-1)\alpha_j^h\},$$

$$+ \frac{1}{p-1} a_{\lambda,2}^h z_2' z_2 w(z'z) \qquad (h = d, o\,; j = 1, 2).$$

Therefore, from (10.45), we can see that the norm of $A^h\{s_j(z, \theta_0) - a_j^h\}$ is a function of z_1^2 and $z_2' z_2$, which is given as follows.

$$(10.56) \qquad \| A^h\{s_j(z, \theta_0) - a_j^h\} \|^2$$
$$= (a_{\mu j,1}^h)^2 z_1^2 w(z'z)^2 + (a_{\mu j,2}^h)^2 z_2' z_2 w(z'z)^2$$
$$+ \frac{1}{2}[\{z_1^2 w(z'z) - \alpha_j^h\}^2 \{(a_\nu^h)^2 + (p-1)(a_\rho^h)^2\}$$
$$+ \{z_2' z_2 w(z'z) - (p-1)\alpha_j^h\}^2 \{(a_\rho^h)^2 + 1/(p-1)(a_\tau^h)^2\}$$
$$+ 2\{z_1^2 w(z'z) - \alpha_j^h\}\{z_2' z_2 w(z'z) - (p-1)\alpha_j^h\}\{a_\nu^h a_\rho^h + a_\tau^h a_\rho^h\}$$
$$+ z_1^2 z_2' z_2 w(z'z)^2 (a_{\lambda,1}^h)^2 + \| z_2 \|^4 w(z'z)^2 (p-2)/(p-1)(a_{\lambda,2}^h)^2].$$

From (10.13), (10.49), (10.50) and (10.51)

$$\| P(\theta_\Delta)\{A^h s_j(z, \theta_0) - a_j^h\} \|^2 = \frac{1}{2}(a_{\mu j,1}^h)^2 z_1^2 w(z'z)^2. \tag{10.57}$$

Therefore $\tilde{W}_j^h(z, \theta_0)$ is a function of z_1^2 and $\|z_2\|^2$ $(h = d, o; j = 1, 2)$.

In order to get the system of equations for α_j^h, $a_{\mu j,k}^h$, $a_{\lambda,k}^h$, a_ν^h and a_τ^h $(h = d, o; j, k = 1, 2)$, we use the following two lemmas.

LEMMA 10.3. *Let* $M^h = r_1 M_1^h + r_2 M_2^h$, *where* M_j^h $(j = 1, 2)$ *is given by* (10.32). *Then*

$$M^h = \mathrm{Diag}\,\{m_{\mu 1,1}^h, m_{\mu 1,2}^h I_{p-1}, m_{\mu 1,1}^h, m_{\mu 1,2}^h I_{p-1}, M_\Lambda^h\}, \tag{10.58}$$

with semi-d-type matrix M_Λ^h *given by*

$$M_\Lambda^h = D^*[m_{\lambda,1}^h, m_{\lambda,2}^h, m_\nu^h, m_\rho^h, m_\tau^h], \tag{10.59}$$

where

$$\begin{aligned}
m_{\mu j,1}^h &= r_j \int z_1^2 w(z'z)^2 \tilde{W}_j^h(z; \theta_0) dF(z; \eta_0), \\
m_{\mu j,2}^h &= r_j \frac{1}{p-1} \int z_2' z_2 w(z'z)^2 \tilde{W}_j^h(z; \theta_0) dF(z; \eta_0), \\
m_{\lambda,1}^h &= \sum r_j \frac{1}{p-1} \int z_1^2 z_2' z_2 w(z'z)^2 \tilde{W}_j^h(z; \theta_0) dF(z; \eta_0), \\
m_{\lambda,2}^h &= \sum r_j \frac{1}{(p-1)(p+1)} \int \|z_2\|^4 w(z'z)^2 W_j^h(z; \theta_0) dF(z; \eta_0), \\
m_\nu^h &= \sum r_j \frac{1}{2} \int \{y_1^2 w(y_j' y_j) - \alpha_{j,1}^h\}^2 \tilde{W}_j^h(z; \theta_0) dF(z; \eta_0), \\
m_\rho^h &= \sum r_j \frac{1}{2} \int \{z_1^2 w(z'z) - \alpha_{j,1}^h\} \\
&\quad \cdot \left\{\frac{1}{p-1} z_2' z_2 w(z'z) - \alpha_{j,2}^h\right\} \tilde{W}_j^h(z; \theta_0) dF(z; \eta_0), \\
m_\tau^h &= \sum r_j \frac{1}{2}(p-1) \int \left\{\frac{1}{(p-1)} z_2' z_2 w(z'z) - \alpha_{j,2}^h\right\}^2 W_j^h(z; \theta_0) dF(z; \eta_0).
\end{aligned} \tag{10.60}$$

LEMMA 10.4. *Let* M_Λ *be a semi-d-type matrix.* $D^*[m_{\lambda,1}, m_{\lambda,2}, m_\nu, m_\rho, m_\tau]$. *Then the inverse matrix is also semi-d-type, which is given by*

$$M_\Lambda^{-1} = D^*[m_{\lambda,1}^{-1}, m_{\lambda,2}^{-1}, m_\tau/\gamma, -m_\rho/\gamma, m_\tau/\gamma], \tag{10.61}$$

where

$$\gamma = m_\nu m_\tau - (p-1) m_\rho^2. \tag{10.62}$$

Now the system of equations (10.22) and (10.23) with (10.20), (10.21) and (10.24) can be reduced to the following system of equations.

$$\alpha_{j,1}^h = \int z_1^2 w(z'z)\tilde{W}_j^h(z,\theta_0)dF(z;\eta_0)/\int \tilde{W}_j^h(z,\theta_0)dF(z;\eta_0),$$

(10.63)
$$\alpha_{j,2}^h = \frac{1}{p-1}\int z_2'z_2 w(z'z)\tilde{W}_j^h(z,\theta_0)dF(z;\eta_0)/\int \tilde{W}_j^h(z,\theta_0)dF(z;\eta_0),$$
$$a_{\mu j,k}^h = (m_{\mu j,k}^h)^{-1},\ a_{\lambda,k}^h = (m_{\lambda,k}^h)^{-1},$$
$$a_\nu^h = m_\tau^h/\gamma^h,\ a_\rho^h = -m_\rho^h/\gamma^h \quad \text{and} \quad a_\tau^h = m_\nu^h/\gamma^h$$

($h = d, o; j, k = 1, 2$), where

(10.64) $$\gamma^h = m_\nu^h m_\tau^h - (p-1)m_\rho.$$

Here $m_{\mu j,k}^h,\cdots$ are given by (10.60) and

(10.65) $$\tilde{W}_j^h(z,\theta_0) = \min\,[1,\ c_j^h/v_j^h(z_1^2, z_2'z_2)],$$

where the function $v_j^h(x, y)$ is non-negative and is given by

(10.66) $$\{v_j^h(x,y)\}^2 = \frac{1}{2}(a_{\mu j,1}^h)^2 xw(x+y)^2$$

and

(10.67)
$$\{v^o(x,y)\}^2 = \frac{1}{2}(a_{\mu j,1}^o)^2 xw(x+y)^2 + (a_{\mu j,2}^o)^2 yw(x+y)^2$$
$$+\frac{1}{2}[\{xw(x+y)-\alpha_j^o\}^2\{(a_\nu^o)^2+(p-1)(a_\rho^o)^2\}$$
$$+\{yw(x+y)-(p-1)\alpha_j^o\}^2\{(a_\rho^o)^2+1/(p-1)(a_\tau^o)^2\}$$
$$+2\{xw(x+y)-\alpha_j^o\}\{yw(x+y)-(p-1)\alpha_j^o\}\{a_\nu^o a_\rho^o + a_\tau^o a_\rho^o\}$$
$$+xyw(x+y)^2(a_{\lambda,1}^o)^2 + y^2w(x+y)^2(p-2)/(p-1)(a_{\lambda,2}^o)^2].$$

From Theorem 9.2, the optimal estimation functions are $\psi_j(\alpha_\theta x, \theta_\Delta)$ ($j = 1, 2$), where α_θ and $\psi_j(x, \theta_\Delta)$ are given by (10.12) and (10.19), respectively. We can easily check

(10.68) $$\alpha_\theta x - \delta = H\Lambda^{-1/2}(x-\mu_1) \quad \text{and} \quad \alpha_\theta x = H\Lambda^{-1/2}(x-\mu_2),$$

and the first elements of $\alpha_\theta x - \delta$ and $\alpha_\theta x$ are given by

(10.69) $$\xi'\Lambda^{-1/2}(x-\mu_1) \quad \text{and} \quad \xi'\Lambda^{-1/2}(x-\mu_2),$$

respectively, where ξ is given by (10.6). Therefore using Theorem 8.2, Lemma

8.1 and Theorem 9.2 we obtain the estimation equations which define the optimal D-robust equivariant M-estimator of θ as in the following theorem.

THEOREM 10.1. *Suppose that* α_j^h, $a_{\mu j,k}^h$, $a_{\lambda,k}^h$, a_ν^h *and* a_τ^h $(h = d, o; j, k = 1, 2)$ *solves the system of equations given by* (10.63). *Then the following system of equations for* $T = (\hat{\mu}_1, \hat{\mu}_2, \hat{\Lambda})$ *defines the optimal D-robust equivariant M-estimator.*

$$
\begin{aligned}
&\textstyle\sum_j r_j a_{\mu j,1}^d \int \hat{y}_j' \hat{\xi} w(\hat{y}_j' \hat{y}_j) W_j^d(x) dF_j(x) = 0, \\
&r_1 a_{\mu 1,1}^o \int \hat{y}_1' \hat{\xi} w(\hat{y}_1' \hat{y}_1) W_1^o(x) dF_1(x) \\
&= r_2 a_{\mu 2,2}^o \int \hat{y}_2' \hat{\xi} w(\hat{y}_2' \hat{y}_2) W_2^o(x) dF_2(x), \\
&\textstyle\sum_j r_j a_{\mu j,2}^o \int (I - \hat{\xi}\hat{\xi}') \hat{y}_j w(\hat{y}_j' \hat{y}_j) W_j^d(x) dF_j(x) = 0, \\
&\textstyle\sum_j r_j \int [a_\rho^o \{\hat{y}_j'(I - \hat{\xi}\hat{\xi}')\hat{y}_j + a_\nu^o (\hat{y}_j'\hat{\xi})^2] w(\hat{y}_j'\hat{y})^2 W_j^o(x) dF_j(x) \\
&= \textstyle\sum_j r_j \alpha_j^o \{a_\nu^o + (p-1) a_\rho^o\} \int W_j^o(x) dF_j(x), \\
&\textstyle\sum_j r_j a_{\lambda,1}^o \int (I - \hat{\xi}\hat{\xi}') \hat{y}_j \hat{y}_j' \hat{\xi} w(\hat{y}_j' \hat{y}_j) W_j^o(x) dF_j(x) = 0, \\
&\textstyle\sum_j r_j a_{\lambda,2}^o \int (I - \hat{\xi}\hat{\xi}') \hat{y}_j \hat{y}_j' (I - \hat{\xi}\hat{\xi}') w(\hat{y}_j' \hat{y}_j) W_j^o(x) dF_j(x) \\
&= \textstyle\sum_j r_j \int \left[\left\{ \frac{1}{p-1} (a_{\lambda,2}^o - a_\tau^o) \hat{y}_j'(I - \hat{\xi}\hat{\xi}') \hat{y}_j - a_\rho^o (\hat{y}_j'\hat{\xi})^2 \right\} w(\hat{y}_j' \hat{y}_j) \right. \\
&\qquad \left. + \alpha_j^o (a_\tau^o + a_\rho^o) \right] W_j^o(x) dF_j(x) (I - \hat{\xi}\hat{\xi}'),
\end{aligned}
\tag{10.70}
$$

where

$$\hat{y}_j = \hat{\Lambda}^{-1/2}(x - \hat{\mu}_j), \tag{10.71}$$

$$\hat{\xi} = \hat{\Lambda}^{-1/2}(\hat{\mu}_1 - \hat{\mu}_2) / \| \hat{\Lambda}^{-1/2}(\hat{\mu}_1 - \hat{\mu}_2) \| \tag{10.72}$$

and

$$W_j^h(x) = \min [1, c_j^h / v_j^h \{(\hat{y}_j'\hat{\xi})^2, \hat{y}_j'(I - \hat{\xi}\hat{\xi}')\hat{y}_j\}] \tag{10.73}$$

with $v_j^h(x, y)$ *given by* (10.66) *and* (10.67) $(h = d, o; j = 1, 2)$.

The question of existence and uniqueness of the above estimator should be answered for each model, i.e., for each function h in (10.1). These problems are remained for further study.

Acknowledgment

The author would like to express his sincere gratitude to Professor Y. Fujikoshi, Hiroshima University, for his valuable comments and encouragement.

References

[1] T. W. Anderson, Classification by Multivariate Analysis, Psychometrika, **16** (1951), 631–650.

[2] T. W. Anderson, An Introduction to Multivariate Analysis, Wiley, New York, (1958).

[3] N. Balakrishnan and S. Kocherlakota, Robustness to nonnormality of the linear discriminant function: mixtures of normal distributions, Comm. Statist.-Theo. Meth., **14**(2) (1985), 465–478.

[4] B. Broffitt, W. R. Clark and P. A. Lachenbruch, The effect of Huberizing and triming on the quadratic discriminant function, Comm. Stat.-Theo. Meth., **A9**(1) (1980), 13–25.

[5] K. C. Chu, Estimation and decision for linear systems with elliptical random processes, IEEE Trans. Automatic Control, **AC-18** (1973), 499–505.

[6] H. Cramer, Mathematical Methods of Statistics. Princeton Univ. Press, 1946.

[7] R. A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen., **7** (1936), 179–188.

[8] Y. Fujikoshi, An asymptotic expansion for the distributions of the latent roots of the Wishart matrix with multiple population roots, Ann. Inst. Statist. Math., **29** A (1977), 379–387.

[9] F. R. Hampel, Contributions to the theory of robust estimation, Ph. D. thesis, University of California, Berkeley, (1968).

[10] F. R. Hampel, The influence curve and its role in robust estimation, J. Am. Statist. Assoc. **69** (1974), 383–393.

[11] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, Robust Statistics, Wiley, New York, 1986.

[12] T. Hayakawa and M. L. Puri, Asymptotic distributions of likelihood ratio criteria for testing latent roots and latent vectors of a covariance matrix under an elliptical population, Biometrika, **72** (1985), 331–338.

[13] P. J. Huber, A robust estimation of a location parameter, Ann. Math. Statist., **35** (1964), 73–101.

[14] P. J. Huber, Robust covariances, Statistical Decision Theory and Related Topics II (Shanti S. Gupta and David S. Moore, ed.), Academic Press, New York, (1977), 165–191.

[15] S. Johnson, Systems of frequency curves generated by methods of translation, Biometrika, **36** (1949), 149–176.

[16] G. Kallianpur and C. R. Rao, On Fisher's lower bound to asymptotic variance of a consistent estimate, Sankhyā A **15** (1955), 331–342.

[17] D. Kelker, Distribution theory of spherical distributions and a location-scale parameter generalization, Sankhyā A, **32** (1970), 419–430.

[18] M. Koutras, On the performance of the linear discriminant function for spherical distributions, J. Multiv. Anal., **22** (1987), 1–12.

[19] W. J. Krzanowski, The performance of Fisher's linear discriminant function under non-optimal conditions, Technometrics, **19** (1977), 191–200.

[20] P. A. Lachenbruch, C. Sneeringer and L. T. Revo, Robustness of the linear and quadratic discriminant function to certain types of non-normality, Comm. Statist.-Theo. Meth., **1**(1) (1973), 39–56.

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

[22] R. J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

[23] H. Nakanishi and Y. Sato, The performance of the linear and quadratic discriminant

functions for three types of non-normal distribution, Comm. Statist.-Theo. Meth., **14**(5) (1985), 1181–1200.

[24] M. Okamoto, An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Statist., **34** (1963), 1286–1301. (Correction: 39 (1968), 1358–1359).

[25] R. H. Randles, J. D. Broffitt, J. S. Ramberg and R. B. Hogg, Generalized linear and quadratic discriminant functions using robust estimators, J. Am. Statist. Assoc., **73** (1978), 564–568.

[26] M. Siotani, Large Sample Approximation and Asymptotic Expansions of Classification Statistics, Handbook of Statistics, Vol. **2** (P. R. Krishnaiah and L. N. Kanal, ed.), North-Holland Publishing Company (1982), 61–100.

[27] M. Siotani and R. H. Wang, Further Expansion Formulae for Error Rates and Comparison of the W-procedure and the Z-procedure in discriminant analysis, Tech. Repo. No. 33, Dept. Statist., Kansas State Univ., Manhattan, KS, (1975).

[28] M. Siotani and R. H. Wang, Asymptotic expansions for error rates and comparison of the W-procedure and the Z-procedure in discriminant analysis, Multivariate Analysis-IV (P. R. Krishnaiah, ed.), North-Holland, Amsterdam, (1977), 525–545.

[29] M. Siotani, T. Hayakawa and Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, Ohaio, (1985).

[30] A. Wald, On a statistical problem arising in the classification of an individual into two groups, Ann. Math. Statist., **15** (1944), 145–163.

[31] B. L. Welch, Note on discriminant functions, Biometrika **31** (1939), 218–220.

Department of Computer Science and Intelligent Systems
Faculty of Engineering
Oita University