8221221

Ramos, Juan Enrique

ESTIMATING THE PROBABILITIES OF MISCLASSIFICATION IN DISCRIMINANT ANALYSIS

Iowa State University

PH.D. 1982

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

Estimating the probabilities of misclassification

in discriminant analysis

by

Juan Enrique Ramos

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1982

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
	A. The Classification Problem	1
	B. Literature Review	6
	C. An Overview of the Present Research	10
II.	THE LIKELIHOOD RATIO TEST STATISTIC FOR CLASSIFICATION	12
	A. Derivation of the Statistic and its Properties	12
	B. Expressions for the Probabilities of Misclassification	19
	C. Selection of a Cutoff Point	23
III.	ESTIMATORS FOR THE UNCONDITIONAL PROBABILITY OF MISCLASSIFICATION	33
	A. Introduction	33
	B. Estimators Based on the Normal Distribution	36
	C. Asymptotic Biases	39
	D. Asymptotic Mean Square Errors	41
	E. Jackknife Estimator	53
IV.	COMPARISONS AND CONCLUSIONS	65
	A. Comparisons with Respect to Asymptotic Bias	65
	B. Comparisons with Respect to Asymptotic Mean Square Error	67
	C. A Monte Carlo Study	78
	D. Conclusions	83
V.	BIBLIOGRAPHY	85
VI.	ACKNOWLEDGEMENTS	88

LIST OF TABLES

		Page
Table 2.1.	Mahalanobis distance between the two populations and corresponding overall probability of misclassification	27
Table 2.2.	Sample sizes and dimension	28
Table 2.3.	Values of $P(0) - P(k_o)$ when $p = 3$	29
Table 2.4.	Values of $P(0) - P(k_o)$ when $p = 4$	30
Table 2.5.	Values of $P(0) - P(k_o)$ when $p = 8$	31
Table 3.1.	Estimators for $P_2^*(Z)$ and the corresponding first order terms	38
Table 4.1.	Values of $AMSE(Q_{DS}) - AMSE(Q_{MOS})$	73
Table 4.2.	Values of $AMSE(Q_{MOS}) - AMSE(Q_{MC})$	75
Table 4.3.	Values of $AMSE(Q_{DS}) - AMSE(Q_{MC})$	76
Table 4.4.	Average values of $\|P_2^*(Z) - Q_t\|$ when $p = 1$	81
Table 4.5.	Average values of $\|P_2^*(Z) - Q_t\|$ when $p = 2$	82

I. INTRODUCTION

A. The Classification Problem

Let X, a p x 1 vector, be an observation which is known to have come from one of two populations, denoted by π_1 and π_2, with density functions $f_1(X)$ and $f_2(X)$, respectively. The problem of classifying the observation consists of deriving a procedure to assign X into one of the two populations and evaluating the performance of the proposed procedure.

Anderson (1958) shows that when the population density functions are known and for given prior probabilities and costs of misclassification, the Bayes procedure classifies X into π_1 if

$$\frac{f_1(X)}{f_2(X)} \geq k, \tag{1.1}$$

where k is a constant depending on the prior probabilities and the costs of misclassification; if the ratio in (1.1) is less than k, X is classified into π_2.

Under the assumptions that the populations are distributed as multivariate normal, denoted by $N(\mu_i, \Sigma)$, i = 1, 2, and the populations parameters μ_1, μ_2 and Σ are known, the Bayes classification rule reduces to:

$$\begin{aligned} &\text{classify } X \text{ in } \pi_1 \quad \text{if} \quad R(X) \geq \log k, \\ &\text{classify } X \text{ in } \pi_2 \quad \text{if} \quad R(X) < \log k, \end{aligned} \tag{1.2}$$

where

$$R(X) = (X - \tfrac{1}{2}(\mu_1 + \mu_2))'\Sigma^{-1}(\mu_1 - \mu_2) \tag{1.3}$$

is known as the best linear discriminant function.

When the population densities are $N(\mu_i, \Sigma)$, $i = 1, 2$, but the parameters μ_1, μ_2 and Σ are unknown, we can no longer use the procedure given in (1.2). For this case, no best classification rule has been found yet. In order to classify the observation X, information must be obtained about the unknown parameters by taking random samples of sizes N_1 and N_2 from populations π_1 and π_2. The classification rule will be based on the data collected in the samples.

Suppose that we have a random sample $X_{11},\ldots,X_{1N_1}$ from Π_1 and an independent random sample $X_{21},\ldots,X_{2N_2}$ from π_2. The usual estimators of the parameters μ_1, μ_2 and Σ are $\bar{X}_1$, $\bar{X}_2$ and S respectively, where

$$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{n_i} X_{ij}, \quad i = 1, 2 \tag{1.4}$$

and

$$S = \frac{1}{n} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i), \tag{1.5}$$

with $n = N_1 + N_2 - 2$.

Any classification rule based on $\bar{X}_1$, $\bar{X}_2$ and S will be denoted by $C(\bar{X}_1, \bar{X}_2, S)$.

A widely used classification procedure is derived by simply replacing μ_1, μ_2 and Σ by $\bar{X}_1$, $\bar{X}_2$ and S in (1.2). This rule is given as follows:

$$\begin{aligned} &\text{classify } X \text{ in } \pi_1 \quad \text{if} \quad W \geq \log k, \\ &\text{classify } X \text{ in } \pi_2 \quad \text{if} \quad W < \log k, \end{aligned} \tag{1.6}$$

where

$$W = (X - \tfrac{1}{2}(\bar{X}_1 + \bar{X}_2))' S^{-1} (\bar{X}_1 - \bar{X}_2). \tag{1.7}$$

This is known as the Anderson's classification statistic.

An alternative procedure to classify observations in the case when the populations are normally distributed and with unknown parameters, is the likelihood ratio test procedure. The classification problem is viewed as a testing problem. The null hypothesis states that $X_{11}, \ldots, X_{1N_1}$ are distributed as $N(\mu_1, \Sigma)$ and X, $X_{21}, \ldots, X_{2N_2}$ are distributed as $N(\mu_2, \Sigma)$. The alternative hypothesis states that X, $X_{11}, \ldots, X_{1N_1}$ are distributed as $N(\mu_1, \Sigma)$ and $X_{21}, \ldots, X_{2N_2}$ are distributed as $N(\mu_2, \Sigma)$. The likelihood ratio test, see Anderson (1958), leads to the following procedure:

$$\begin{aligned} &\text{classify } X \text{ in } \pi_1 \quad \text{if} \quad L \leq (\eta - 1)n, \\ &\text{classify } X \text{ in } \pi_2 \quad \text{if} \quad L > (\eta - 1)n, \end{aligned} \tag{1.8}$$

where η is a given constant and L is the likelihood ratio test statistic given by

$$L = \frac{N_1}{N_1+1}(X-\bar{X}_1)' S^{-1} (X-\bar{X}_1) - \eta \frac{N_2}{N_2+1}(X-\bar{X}_2)' S^{-1} (X-\bar{X}_2). \tag{1.9}$$

Han (1979) examined several different approaches to the classification problem. One of them, using Bayesian arguments, also leads to the L classification statistic.

When $\eta = 1$, L reduces to the John's classification statistic, John (1960), given by

$$Z = \frac{N_1}{N_1+1}(X-\bar{X}_1)'S^{-1}(X-\bar{X}_1) - \frac{N_2}{N_2+1}(X-\bar{X}_2)'S^{-1}(X-\bar{X}_2) \tag{1.10}$$

and the classification rule in (1.8) simplifies to

$$\begin{aligned} &\text{classify } X \text{ in } \pi_1 \quad \text{if} \quad Z \leq 0, \\ &\text{classify } X \text{ in } \pi_2 \quad \text{if} \quad Z > 0. \end{aligned} \tag{1.11}$$

The performance of a given classification rule, $C(\bar{X}_1, \bar{X}_2, S)$ can be evaluated by examining the probabilities of misclassification. Three different pairs of probabilities are of interest here. The first one consists of the optimum probability of misclassification,

$$P_2^{**} = P[f_1(X)/f_2(X) \geq k \mid X\varepsilon\ \pi_2] \tag{1.12}$$

and

$$P_1^{**} = P[f_1(X)/f_2(X) < k \mid X\varepsilon\ \pi_1] \tag{1.13}$$

These are the probabilities of misclassifying an observation coming from π_1 or from π_2 when the population densities are completely specified. When $f_i(X)$ is $N(\mu_i, \Sigma)$, $i = 1, 2$, both (1.12) and (1.13) reduce to

$$P_2^{**}(R(X)) = \Phi(-\delta/2), \tag{1.14}$$

where k is equal to zero, δ^2 is the Mahalanobis distance between the two populations, defined as

$$\delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) \tag{1.15}$$

and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

The next pair of probabilities of misclassification is given when the parameters are unknown and estimated by samples. The unconditional probabilities of misclassification are given by

$$P_2^*(C) = P_2^*[C(\bar{X}_1,\bar{X}_2,S) \text{ classifies } X \text{ into } \pi_1 \mid X\varepsilon\ \pi_2] \tag{1.16}$$

and

$$P_1^*(C) = P_1^*[C(\bar{X}_1,\bar{X}_2,S) \text{ classifies } X \text{ into } \pi_2 \mid X\varepsilon\ \pi_1]. \tag{1.17}$$

Finally, the conditional probabilities of misclassification, are defined as,

$$P_2(C) = P_2[C(\bar{X}_1,\bar{X}_2,S) \text{ classifies } X \text{ in } \pi_1 \mid X\varepsilon\ \pi_2,\ \bar{X}_1,\bar{X}_2,S] \tag{1.18}$$

and

$$P_1(C) = P_1[C(\bar{X}_1,\bar{X}_2,S) \text{ classifies } X \text{ in } \pi_2 \mid X\varepsilon\ \pi_1,\ \bar{X}_1,\bar{X}_2,S]. \tag{1.19}$$

When the Z classification statistic is used, the unconditional probabilities of misclassification are

$$P_2^*(Z) = P[Z \leq 0 \mid X\varepsilon\ \pi_2] \tag{1.20}$$

and

$$P_1^*(Z) = P[Z > 0 \mid X\varepsilon\ \pi_1]. \tag{1.21}$$

The conditional probabilities of misclassification are

$$P_2(Z) = P[Z \leq 0 \mid X\varepsilon\ \pi_2,\ \bar{X}_1, \bar{X}_2, S] \tag{1.22}$$

$$P_1(Z) = P[Z > 0 \mid X\varepsilon\ \pi_1,\ \bar{X}_1, \bar{X}_2, S] \tag{1.23}$$

A concept which is usually of interest is the overall error rate, defined as

$$\frac{1}{2}(P_1^*(C) + P_2^*(C)) \tag{1.24}$$

Note that all three types of probabilities P, P^* and P^{**} are functions of the parameters μ_1, μ_2 and Σ. When the parameters are unknown, the probabilities are also unknown and the problem of estimating these probabilities arises. In this thesis, we shall consider the estimation of the probability of misclassifying the observation to π_1, when it comes from π_2. The estimation of the probability of misclassification when X comes from π_1, can easily be obtained by interchanging the subscripts.

B. Literature Review

The problem of classification of observations has been the subject of considerable amount of research since Fisher (1936) introduced the linear discriminant function. An extensive bibliography on this area has been published in a review paper by DasGupta (1973) and in the book by Lachenbruch (1975).

A large portion of the work done is related to Anderson's classification statistic W. Initially, the aim was directed mostly towards obtaining the exact distribution of W; see for example Wald (1944) and Sitgreaves (1952). Since Sitgreaves (1961) concluded that the exact distribution is too complicated to be useful numerically, Okamoto (1963) gave an asymptotic expansion for the distribution of W up to the second order with respect to N_1^{-1}, N_2^{-1} and n^{-1}. Also, Anderson (1973) gave another expansion for the distribution of the studentized W.

More recently, attention has been given in the literature to the problem of estimating the probabilities of misclassification since they provide a way to evaluate the performance of the classification procedure. A bibliography on this subject is given by Toussaint (1974).

Smith (1947) proposed the first estimator for the probabilities of misclassification. This estimator, denoted as P_R, is defined as the proportion of observations in the sample from population π_2 that are misclassified when using $C(\bar{X}_1,\bar{X}_2,S)$. It has the advantages of being simple to compute and does not require any distributional assumptions. However, it is seriously biased and gives too optimistic results when N_1 and N_2 are small.

Hills (1966) considered the problem of estimating the three types of probabilities of misclassification defined above. He obtained some inequalities on these probabilities and examined the performance of some estimators when the populations are not normal.

Lachenbruch (1967) suggested a modification of Smith's estimator to correct for bias. He proposed a jackknife type classification procedure, known in the literature as Lachenbruch's leaving one-out method. This procedure produces an estimator of the probabilities of misclassification having second order bias without requiring any distributional assumptions. This estimator will be denoted by P_L.

Lachenbruch and Mickey (1968) used a simulation study to compare the performance of several estimators of the conditional probability of misclassification when Anderson's classification statistic is used. They compared the P_R and P_L estimators and four others, denoted by Do, DS, O and OS, which assume that the populations are normally distributed. The estimators Do and DS are obtained by replacing δ^2, in the expression for the optimal probability of misclassification given in (1.14), by D^2 and $((n-p-1)/n)D^2$ respectively, where

$$D^2 = (\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2) \tag{1.25}$$

is the sample Mahalanobis distance between the two populations. The estimators O and OS are similarly obtained, except that δ^2 is replaced by D^2 and $((n-p-1)/n)D^2$ in Okamoto's asymptotic expansion for the probability of misclassification. By examining the overall performance, the estimators P_R and Do were the worst; the best estimator was the OS estimator, except for small δ^2 and small sample sizes.

Broffitt and Williams (1973) obtained the UMVUE for the expectations of P_R and P_L. They also gave exact expressions for the unconditional probabilities of misclassification using Anderson's W statistic.

McLachlan (1974b) verified analytically most of the results in Lachenbruch and Mickey's simulation study. Assuming that the populations are normally distributed, he derived an asymptotic expansion for the AMSE for each one of the estimators Do, DS, O and OS. The estimators are then compared by using the AMSE. He also proposed another estimator, denoted by M, which has third order bias with respect to N_1^{-1}, N_2^{-1} and n^{-1}. The performance of M in terms of AMSE is similar to that of the OS estimator.

A criterion proposed by Kudo (1959) and John (1960) as an alternative to the W classification procedure is the Z procedure, given in (1.10). DasGupta (1965) proved that this procedure is admissible and minimax in the class of invariant procedures.

The exact distribution of the Z statistic has not been obtained yet. Memon and Okamoto (1971) gave an asymptotic expansion for the distribution function of Z with respect to N_1^{-1}, N_2^{-1} and n^{-1}.

Based on this expansion and on Okamoto's expansion for the probability of misclassification when using W, Memon and Okamoto (1971) concluded that when the error rates for the Z and W statistics are compared, the Z procedure is better.

Siotani and Wang (1977) compared the W and the Z statistics with respect to the error rates, including the third order terms in the expansions. They obtained a set of values of N_1, N_2, p and δ^2 for which the Z procedure is better than the W procedure.

The Z and W statistics have also been compared in other respects. Aitchison et al. (1977) compared two methods of estimating density functions, known as the estimative method and the predictive density function method. They found that, in relation to the Kullback and Liebler measure of closeness, the predictive density function method is better than any other method of estimating densities. Moran and Murphy

(1979) showed that when the predictive density function method is used to classify observations, it reduces to the L classification statistic. It is well-known that the estimative method reduces to the W classification statistic.

Han (1979) gave an expression for the conditional probability of misclassification when using the L statistic, given $\bar{X}_1$, $\bar{X}_2$ and S.

C. An Overview of the Present Research

The primary goal of this thesis is to study the John's classification statistic and the estimation of the probability of misclassification. In Chapter II, John's classification statistic is obtained from two different approaches to the classification problem. First, it is derived as a special case of the likelihood ratio test statistic. Then, it is obtained from a Bayesian point of view, using the predictive density function concept. In Section B, expressions for various probabilities of misclassification are given. The Memon-Okamoto (1971) asymptotic expansions for the unconditional probabilities of misclassification are of special interest and will be extensively used in this thesis. In Section C of Chapter II, we examine the problem of selecting a cutoff point. The commonly used cutoff point is zero. An alternative cutoff point is given so that it minimizes the overall error rate. We show that, with respect to overall error rate, the zero cutoff point is not too far away from the optimum.

In Chapter III, Section B, several estimators for the unconditional probability of misclassification, $P_2^*(Z)$, are proposed. All these estimators are based on the normal distribution. Expressions for the asymptotic bias and for the asymptotic mean square error (AMSE) of each one of the proposed estimators are given in Sections C and D, respectively. In Section E of Chapter III, a jackknife estimator is proposed and its properties are investigated.

In Chapter IV, the estimators proposed in Chapter III are compared, first with respect to their asymptotic bias and then with respect to their AMSE. In Section C of this chapter, two additional estimators for the unconditional probability of misclassification are given; these two estimators, do not require any distributional assumptions. A simulation study is then used to compare all estimators. General conclusions about the relative superiority of the estimators are given in Section D.

II. THE LIKELIHOOD RATIO TEST STATISTIC FOR CLASSIFICATION

A. Derivation of the Statistic and its Properties

The likelihood ratio test approach considers the problem of classification of observations as one of testing the null hypothesis that $X_{11},\ldots,X_{1N_1}$ are drawn from $N(\mu_1, \Sigma)$ and $X, X_{21},\ldots,X_{2N_2}$ are drawn from $N(\mu_2, \Sigma)$ versus the alternative hypothesis that $X, X_{11},\ldots,X_{1N_1}$ are drawn from $N(\mu_1, \Sigma)$ and $X_{21},\ldots,X_{2N_2}$ are drawn from $N(\mu_2, \Sigma)$ with μ_1, μ_2 and Σ being unspecified. Following Anderson (1958), it can be shown that the likelihood ratio criterion is the $(N_1 + N_2 + 1)/2$ th power of the ratio

$$\frac{n + \frac{N_1}{N_1 + 1}(X - \bar{X}_1)'S^{-1}(X - \bar{X}_1)}{n + \frac{N_2}{N_2 + 1}(X - \bar{X}_2)'S^{-1}(X - \bar{X}_2)}, \tag{2.1}$$

where $\bar{X}_1$, $\bar{X}_2$, S and n are defined in (1.4) and (1.5).

The region of classifying the observation into π_1 is then equivalent to the region of rejection of the null hypothesis. This region contains all those X for which the ratio in (2.1) is smaller than a certain constant η. This leads directly to the likelihood ratio test procedure which is given as

$$\begin{aligned} &\text{Classify } X \text{ in } \pi_1 \text{ if } L \leq (\eta - 1)n, \\ &\text{Classify } X \text{ in } \pi_2 \text{ if } L > (\eta - 1)n, \end{aligned} \tag{2.2}$$

where

$$L = \frac{N_1}{N_1 + 1}(X - \bar{X}_1)'S^{-1}(X - \bar{X}_1) - \eta \frac{N_2}{N_2 + 1}(X - \bar{X}_2)'S^{-1}(X - \bar{X}_2) \quad (2.3)$$

is the likelihood ratio test statistic. For the special case when $\eta = 1$, it is obvious that we obtain the classification procedure given in (1.11) with classification statistic Z, defined in (1.10).

An alternative derivation of the likelihood ratio test statistic is obtained using the predictive density function approach to the problem of classification. This is a Bayesian approach and requires the assessment of a prior distribution for the unknown parameters μ_1, μ_2 and Σ. The solution to the problem of classification of observations is obtained when the likelihood functions in (1.1) are replaced by the respective predictive density functions, defined (see Aitchison, 1975) as follows:

Let f be a future experiment with class of density functions $\{p(z|\theta);\ \theta\varepsilon\Theta\}$ and sample space Z. Let e be an informative experiment with class of density functions $\{p(y|\theta);\ \theta\varepsilon\Theta\}$ and sample space Y. For each $\theta\varepsilon\Theta$, denote its prior density function as $p(\theta)$ and assume that f and e are independent experiments. Then, the predictive density function is

$$p(z|y) = \int_{\Theta} p(z|\theta)p(\theta|y)d\theta \quad (2.4)$$

where $p(\theta|y)$ is the posterior density function. Note that $p(z|y)$ is the posterior Bayes estimator of the density function $p(z|\theta)$.

Suppose that the informative experiment e consists of random samples $X_{i1},\ldots,X_{iN_i}$ from $N(\mu_i, \Sigma)$ $i = 1,2$. Then $\bar{X}_1$, $\bar{X}_2$ and S are independently distributed, with $\bar{X}_i$ distributed as $N(\mu_i, \frac{1}{N_i}\Sigma)$, $i = 1,2$ and nS as a Wishart with n degrees of freedom. Denote their joint distribution as $f(\bar{X}_1, \bar{X}_2, S)$. Let $\theta = (\mu_1, \mu_2, \Sigma^{-1})$ and assume its prior density $g(\theta)$ is such that $g(\theta)\,d\theta$ is proportional to $|\Sigma|^{\frac{1}{2}(p-1)} d\mu_1 d\mu_2 d\Sigma^{-1}$ with $p < n$. Then, the posterior density of θ, $h(\theta|\bar{X}_1, \bar{X}_2, S)$ is proportional to $g(\theta)\, f(\bar{X}_1, \bar{X}_2, S)$ and the predictive density function for π_i is

$$q_i(X|\bar{X}_1, \bar{X}_2, S) = \int_\Theta p_i(X|\theta) h(\theta|\bar{X}_1, \bar{X}_2, S) d\theta; \; i = 1,2 \qquad (2.5)$$

Aitchison and Dunsmore (1975) showed that the $q_i(X|\bar{X}_1, \bar{X}_2, S)$ $i = 1,2$ are equal to

$$\frac{\Gamma(\frac{n+1}{2})\pi^{-p/2}}{\Gamma(\frac{n-p+1}{2})\left|(1+\frac{1}{N_i})nS\right|^{\frac{1}{2}}\{(1+(X-\bar{X}_i)'[(1+\frac{1}{N_i}nS]^{-1}(X-\bar{X}_i)]\}^{(n+1)/2}}. \qquad (2.6)$$

When the $q_i(X|\bar{X}_1, \bar{X}_2, S)$ are substituted for the $f_i(X)$, $i = 1,2$ respectively in (1.1) we obtain that the predictive density function approach to the problem of classification reduces to

$$\begin{aligned} &\text{Classify } X \text{ in } \pi_1 \quad \text{if} \quad L \leq (\eta^* - 1)n, \\ &\text{Classify } X \text{ in } \pi_2 \quad \text{if} \quad L > (\eta^* - 1)n, \end{aligned} \qquad (2.7)$$

where $\eta^* = \left(\frac{k\,N_2(N_1 + 1)}{N_1(N_2 + 1)}\right)^{-\frac{2}{n+1}}$.

Han (1979) showed that a procedure called the Best invariant estimative method also gives the L classification statistic. This will not be presented here.

Now let us consider the properties of the statistic L.

First, we can rewrite it as

$$L = c_1 m_1 - \eta c_2 m_2,$$

where,

$$c_i = \frac{N_i}{N_i + 1}, \quad i = 1,2$$

and
$$m_i = (X - \bar{X}_i)' S^{-1} (X - \bar{X}_i) \quad i = 1,2. \tag{2.8}$$

Then we can show that,

$$E(L|X\varepsilon\pi_1) = \frac{1}{n-p-1}[p(1 - \eta) - c_2\delta^2]$$

$$E(L|X\varepsilon\pi_2) = \frac{1}{n-p-1}[p(1 - \eta) + c_1\delta^2]. \tag{2.9}$$

This can be established since $\bar{X}_1$, $\bar{X}_2$ and S are mutually independent and $E(S^{-1}) = \frac{1}{n-p-1}\Sigma^{-1}$, see Press (1972). Using these results, we have that

$$
\begin{aligned}
E(c_1 m_1 | X\varepsilon\pi_2) &= c_1 E[(X-\bar{X}_1)' S^{-1} (X-\bar{X}_1)] \\
&= c_1 E[tr\{(X-\bar{X}_1)' S^{-1} (X-\bar{X}_1)] \\
&= c_1 E[tr\{S^{-1} (X-\bar{X}_1)(X-\bar{X}_1)'] \\
&= c_1 tr\{E[S^{-1} (X-\bar{X}_1)(X-\bar{X}_1)']\} \\
&= c_1 tr\{E(S^{-1}) E[(X-\bar{X}_1)(X-\bar{X}_1)']\} \\
&= c_1 tr\{(\frac{\Sigma^{-1}}{n-p-1})((\mu_1-\mu_2)(\mu_1-\mu_2)' + \frac{N_2+1}{N_2}\Sigma) \\
&= \frac{p}{n-p-1} + \frac{c_1\delta^2}{n-p-1} .
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E(\eta c_2 m_2 | X\varepsilon\pi_2) &= \eta c_2 tr\{E(S^{-1}) E[(X-\bar{X}_1)(X-\bar{X}_1)']\} \\
&= \eta c_2 tr\ \{(\frac{1}{n-p-1}\Sigma^{-1})(\frac{N_2+1}{N_2}\Sigma)\} \\
&= \eta p/(n-p-1).
\end{aligned}
$$

Then,

$$
E(L | X\varepsilon\pi_2) = \frac{1}{n-p-1}[p(1-\eta) + c_1\delta^2].
$$

The derivation of $E(L|X\varepsilon\pi_1)$ is omitted since it is very similar to the procedure just outlined.

An alternative expression for the likelihood ratio test statistic for classification is given by

$$L = b[(X-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))'S^{-1}(X-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))-\alpha_1(\alpha_1+1)D^2] \tag{2.10}$$

where,

$$b = c_1 - \eta c_2,$$

$$\alpha_1 = \frac{\eta c_2}{b}$$

and (2.11)

$$D^2 = (\bar{X}_1-\bar{X}_2)'S^{-1}(\bar{X}_1-\bar{X}_2)$$

This result can be obtained by setting $h_1 = c_1\bar{X}_1 - c_2\bar{X}_2$ and $h_2 = c_1\bar{X}_1'S^{-1}\bar{X}_1-\eta c_2\bar{X}_2'S^{-1}\bar{X}_2$ and expanding L as

$$\begin{aligned}
L &= c_1(X-\bar{X}_1)'S^{-1}(X-\bar{X}_1)-\eta c_2(X-\bar{X}_2)'S^{-1}(X-\bar{X}_2) \\
&= c_1(X'S^{-1}X-2X'S^{-1}\bar{X}_1+\bar{X}_1'S^{-1}\bar{X}_1)-\eta c_2[X'S^{-1}X-2X'S^{-1}\bar{X}_2+\bar{X}_2'S^{-1}\bar{X}_2] \\
&= bX'S^{-1}X-2X'S^{-1}h_1+h_2 \\
&= b[X'S^{-1}X-2(\tfrac{1}{b})X'S^{-1}h_1+h_2/b] \\
&= b[(X-(\tfrac{1}{b})h_1)'S^{-1}(X-(\tfrac{1}{b})h_1)-(\tfrac{1}{b})^2h_1'S^{-1}h_1+\frac{h_2}{b}] \\
&= b[(X-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))'S^{-1}(X-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))-\alpha_1(\alpha_1+1)D^2]
\end{aligned}$$

since

$$\begin{aligned}
X-(\tfrac{1}{b})h_1 &= x - \frac{c_1\bar{X}_1-\eta c_2\bar{X}_2}{c_1-\eta c_2} \\
&= X - \bar{X}_1(\frac{c_1}{c_1-\eta c_2}) + \bar{X}_2\ (\frac{\eta c_2}{c_1-\eta c_2})
\end{aligned}$$

$$= X - \bar{X}_1 \left(1 + \frac{\eta c_2}{c_1 - c_2}\right) + \bar{X}_2 \left(\frac{\eta c_2}{c_1 - c_2}\right)$$

$$= X - \bar{X}_1 + \alpha_1(\bar{X}_2 - \bar{X}_1)$$

and $b^* = \frac{1}{b^2}\,[bh_2 - h_1' S^{-1} h_1]$

$$= \frac{1}{b^2}\,[bh_2 - (c_1\bar{X}_1 - \eta c_2\bar{X}_2)' S^{-1}(c_1\bar{X}_1 - \eta c_2\bar{X}_2)]$$

$$= \frac{1}{b^2}\,[(c_1 - \eta c_2)(c_1\bar{X}_1' S^{-1}\bar{X}_1 - \eta c_2\bar{X}_2' S^{-1}\bar{X}_2) -$$

$$c_1^2\bar{X}_1' S^{-1}\bar{X}_1 - \eta^2 c_2^2\bar{X}_2' S^{-1}\bar{X}_2 + 2\eta c_1 c_2\bar{X}_1' S^{-1}\bar{X}_2]$$

$$= \frac{1}{(c_1 - \eta c_2)^2}\,[-\eta c_1 c_2\bar{X}_1' S^{-1}\bar{X}_1 - \eta c_1 c_2\bar{X}_2' S^{-1}\bar{X}_2 + 2\eta c_1 c_2\bar{X}_1' S^{-1}\bar{X}_2]$$

$$= \frac{-\eta c_1 c_2}{(c_1 - \eta c_2)^2}\,(\bar{X} - \bar{X}_1)' S^{-1}(\bar{X}_2 - \bar{X}_1)$$

$$= -\,\alpha_1(\alpha_1 + 1)D^2.$$

Using this alternative form, we can now obtain the conditional characteristic function of L, given $\bar{X}_1$, $\bar{X}_2$, and S. Following Johnson and Kotz (1972), we have

$$\phi_L(t) = E[e^{itL} \mid X \varepsilon \pi_2,\ \bar{X}_1,\ \bar{X}_2,\ S]$$

$$= \phi_Q(tb) \cdot e^{-itb\alpha_1(\alpha_1+1)D^2}$$

where $Q = [(X - \bar{X}_1) + \alpha_1(\bar{X}_2 - \bar{X}_1)]' S^{-1}[(X - \bar{X}_1) + \alpha_1(\bar{X}_2 - \bar{X}_1)]$. Hence, the expression for $\phi_L(t)$ is

$$\exp\{-itb\alpha_1(\alpha_1+1)D^2 - \frac{1}{2}\sum_{j=1}^{p} w_j^2 + \frac{1}{2}\sum_{j=1}^{p} \frac{w_j^2}{1-2itb\lambda_j}\} \prod_{j=1}^{p}(1-2itb\lambda_j)^{-\frac{1}{2}} \quad (2.12)$$

with the w_j and λ_j, $j = 1,\ldots,p$ given in the above reference. Conditional moments can then be generated from (2.12) in the usual way or from the cumulant generating function given also in Johnson and Kotz (1972).

B. Expressions for the Probabilities of Misclassification

Let us consider first the case when Σ is known. The likelihood ratio criterion leads directly to the rule:

$$\begin{aligned} &\text{Classify } X \text{ in } \pi_1 \quad \text{if} \quad Z_o \leq 0, \\ &\text{Classify } X \text{ in } \pi_2 \quad \text{if} \quad Z_o > 0, \end{aligned} \quad (2.13)$$

where,

$$Z_o = c_1(X-\bar{X}_1)'\Sigma^{-1}(X-\bar{X}_1) - c_2(X-\bar{X}_2)'\Sigma^{-1}(X-\bar{X}_2) \quad (2.14)$$

Geisser (1964) obtained the predictive density functions for π_i $i = 1,2$. They are proportional to

$$\left(\frac{N_i}{N_i+1}\right)^{\frac{1}{2}p} |\Sigma|^{-\frac{1}{2}} \exp\{-\frac{N_i}{2(N_i+1)}(X-\bar{X}_i)'\Sigma^{-1}(X-\bar{X}_i)\}. \quad (2.15)$$

Therefore, the predictive density function approach also leads to a classification rule which uses the Z_o classification statistic.

The conditional probability of misclassification is then given by

$$P_2(Z_o) = P[Z_o \leq 0 | X\varepsilon\pi_2, \bar{X}_1, \bar{X}_2].$$

Using the alternative expression for L given in (2.10),

$$P_2(Z_o) = P[(X-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))'\Sigma^{-1}(X-\bar{X}_1+\alpha_1(\bar{X}-\bar{X}_1))\leq\alpha_1(\alpha_1+1)D^2|\bar{X}_1,\bar{X}_2,X\varepsilon\pi_2]$$

$$= P[\chi^{2'}(p,\lambda)\leq\alpha_1(\alpha_1+1)D^2|\bar{X}_1,\bar{X}_2,X\varepsilon\pi_2], \qquad (2.16)$$

where

$$\lambda = (\mu_2-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))'\Sigma^{-1}(\mu_2-\bar{X}_1+\alpha_1(\bar{X}_2-\bar{X}_1))$$

and $\chi^{2'}(P,\lambda)$ denotes a noncentral chi-square random variable with p degrees of freedom and noncentrality parameter λ. Note that (2.16) depends on μ_2, which is unknown, and hence must be estimated.

The unconditional probability of misclassification $P_2^*(Z_o) = E[P_2(Z_o)]$ was obtained by John (1960), in terms of an infinite series. An asymptotic expansion of this probability with respect to N_1^{-1} and N_2^{-1} is given by Memon and Okamoto (1971).

We now consider the Σ unknown case. It has been shown in Section A of Chapter II that the likelihood ratio criterion and the predictive density approach to the classification problem both lead to the rule which uses the L classification statistic defined in (2.3). Using the alternative expression for L in (2.10), the derivation of the exact distribution of the likelihood ratio test statistic for classification could be pursued applying some of the results on the distribution of indefinite quadratic forms. Johnson and Kotz (1972) present several methods to deal with this problem. In spite of this, the exact distribution would be of little help since it is very complicated.

Given $\bar{X}_1$, $\bar{X}_2$ and S, the conditional probability of misclassification

$$P_2(L) = P[L \leq (\eta-1)n \mid X\varepsilon\pi_2, \bar{X}_1, \bar{X}_2, S] \tag{2.17}$$

was obtained by Han (1979).

Although there is no asymptotic expansion available in the literature for the distribution of the L statistic, Memon and Okamoto (1971) have obtained an asymptotic expansion with respect to N_1^{-1}, N_2^{-1} and n^{-1} for the distribution function of Z. This is an asymptotic expansion for the distribution function of L when $\eta = 1$.

Using this expansion, the unconditional probabilities of misclassification when classifying with the Z statistic are given as

$$P(Z \leq 0 \mid X\varepsilon\pi_2) = \Phi(-\tfrac{\delta}{2}) + \frac{a_2}{N_1} + \frac{a_1}{N_2} + \frac{a_3}{n} + \frac{b_2}{N_1^2} + \frac{b_1}{N_2^2} + \frac{b_{12}}{N_1N_2} + \frac{b_{23}}{N_1n} + \frac{b_{13}}{N_2n} + \frac{b_3}{n^2} + 0_2 \tag{2.18}$$

and

$$P(Z > 0 \mid X\varepsilon\pi_1) = \Phi(-\tfrac{\delta}{2}) + \frac{a_1}{N_1} + \frac{a_2}{N_2} + \frac{a_3}{n} + \frac{b_1}{N_1^2} + \frac{b_2}{N_2^2} + \frac{b_{12}}{N_1N_2} + \frac{b_{13}}{N_1n} + \frac{b_{23}}{N_2n} + \frac{b_3}{n^2} + 0_2 \tag{2.19}$$

with

$$a_1 = \frac{1}{2\delta^2}\left[-d_o^4 + (p-4)d_o^2\right]$$

$$a_2 = \frac{1}{2\delta^2}\left[3d_o^4 + (p+8)d_o^2\right] \tag{2.20}$$

$$a_3 = \frac{1}{2}(p-1)d_o^2 ,$$

$$d_o^i = (d^i/dy^i)\Phi(y)\Big|_{y=-\frac{\delta}{2}} \quad i = 2,4,6,8$$

and the b's defined in Memon and Okamoto's paper.

Using (2.20), we can rewrite (2.18) as

$$P(Z \leq 0 | X\varepsilon\pi_2) = \Phi(-\frac{\delta}{2}) + A_1 + A_2 + O_3 \tag{2.21}$$

where,

$$A_1 = \frac{1}{16}\phi(-\frac{\delta}{2})\{\frac{1}{N_1}(3\delta+4(p-1)\delta^{-1})+\frac{1}{N_2}(-\delta+4(p-1)\delta^{-1})+\frac{4\delta(p-1)}{n}\} , \tag{2.22}$$

A_2 contains the second order terms and $\phi(\cdot)$ is the standard normal density function.

Exact expressions for the unconditional probabilities of misclassification, $P_2^*(Z)$ and $P_1^*(Z)$, will be derived in section E of Chapter III.

C. Selection of a Cutoff Point

In classifying observations into one of two populations, it is commonplace to assume equal costs of misclassification and equal prior probabilities, so that the value for k in (1.1) is one and hence the cutoff point in (1.2) is zero.

Sedransk (1969) considered the problem of selecting an optimum cutoff point when Anderson's W statistic is used. She proposed a cutoff point k^*, such that the error rate is minimized.

In this section, we will consider the problem of selecting a desirable cutoff point k_o, when the Z classification statistic is used.

Let the overall probability of misclassification be

$$P(k) = P(Z \leq k|X\varepsilon\pi_2] + P[Z \geq k|X\varepsilon\pi_1]. \tag{2.23}$$

The problem is to find k_o so that

$$P(k_o) = \min_k P(k). \tag{2.24}$$

Note that minimizing P(k) with respect to k is equivalent to minimizing the error rate when the prior population probabilities are both 1/2.

Since the exact distribution of Z has not been derived yet, we will make use of the Memon-Okamoto expansion for the cumulative distribution of Z when $X\varepsilon\pi_1$ or $X\varepsilon\pi_2$. This leads to

$$P(k) = 1-\Phi(a) - \phi(a)F_1 + 1 - \Phi(c) - \phi(c)F_2 + O_2, \tag{2.25}$$

where

$$a = \frac{k}{2\delta} + \frac{\delta}{2},$$

$$c = -\frac{k}{2\delta} + \frac{\delta}{2} \tag{2.26}$$

$$F_1 = (2N_1\delta^2)^{-1}[3a-a^3+\delta(a^2-1)-pa]$$

$$+ (2N_2\delta^2)^{-1}[3a-a^3+\delta(a^2-1)-pa+\delta^2a-\delta^3] \tag{2.27}$$

$$+ n^{-1}[3a-a^3+\delta(a^2-1)-\tfrac{1}{4}(\delta^2+6p+6)a+(\tfrac{p+1}{2})\delta]$$

and

$$F_2 = (2N_1\delta^2)^{-1}[3c-c^3+\delta(c^2-1)-pc+\delta^2c-\delta^3]$$

$$+ (2N_2\delta^2)^{-1}[3c-c^3+\delta(c^2-1)-pc]$$

$$+ n^{-1}[3c-c^3+\delta(c^2-1)-\tfrac{1}{4}(\delta^2+6p+6)c+(\tfrac{p+1}{2})\delta] \tag{2.28}$$

and O_2 stands for second order terms of N_1^{-1}, N_2^{-1} and n^{-1}.

Taking the first derivative with respect to k, we obtain,

$$P'(k) = -\frac{1}{2\delta}\phi(a) - \phi(a)[(-\frac{1}{2\delta})aF_1 + \frac{dF_1}{dk}]$$

$$+ \frac{1}{2\delta}\phi(c) - \phi(c)[(\frac{1}{2\delta})cF_2 + \frac{dF_2}{dk}] \tag{2.29}$$

where,

$$\frac{dF_1}{dk} = (2\delta)^{-1}\{(2N_1\delta^2)^{-1}[3-3a^2+2\delta a-p]+(2N_2\delta^2)^{-1}[3-3a^2+2\delta a-p+\delta^2]$$

$$+ (n^{-1})[3-3a^2+2\delta a-\tfrac{1}{4}(\delta^2+6p+6)]\} \tag{2.30}$$

and

$$\frac{dF_2}{dk} = (-2\delta)^{-1}\{(2N_1\delta^2)^{-1}[3-3c^2+2\delta c-p+\delta^2]+(2N_2\delta^2)^{-1}[3-3c^2+2\delta c-p]$$

$$+ (n^{-1})[3-3c^2+2\delta c-\tfrac{1}{4}(\delta^2+6p+6)]\}. \tag{2.31}$$

From these expressions, it is not clear how we can solve for k after Equation (2.29) is set equal to zero, since P'(k) involves $\phi(\frac{k}{2\delta}+\frac{\delta}{2})$ and $\phi(-\frac{k}{2\delta}+\frac{\delta}{2})$ and each of these factors is multiplied by a third degree polynomial in k.

Although it is expected that zero is not the optimal cutoff point, it is anticipated that the optimum will lie close to zero since this is the case when all the parameters are known.

Expanding P'(k) as a Taylor's series about zero gives

$$P'(k_o) = 0 \doteq P'(0) + k_oP''(0) + O_2 \tag{2.32}$$

This can be further simplified by noting that P(k) can be written as $f(k) + g(k) + O_2$, with

$$f(k) = 1 - \Phi(a) + 1 - \Phi(c)$$

and

$$g(k) = -[\phi(a)F_1 + \phi(c)F_2]. \tag{2.33}$$

Since g(k) involves only first order terms with respect to N_1^{-1}, N_2^{-1} and n^{-1}, and k_o is assumed to be close to zero, it follows that $k_og''(0)$ is of higher order than the other terms in (2.32) and may be omitted. Hence, solving for k_o, we have

$$k_o = - \frac{f'(0) + g'(0)}{f''(0)}$$

$$= - \frac{g'(0)}{f''(0)} \quad . \tag{2.34}$$

From Equations (2.30), (2.31) and (2.33), we obtain

$$k_o = \frac{\phi(\frac{\delta}{2})(\frac{1}{N_2} - \frac{1}{N_1})(\frac{\delta^2+4}{16\delta})}{(4\delta)^{-1}\phi(\delta/2)}$$

$$= (\frac{1}{N_2} - \frac{1}{N_1})\ (1 + \frac{\delta^2}{4}) \ . \tag{2.35}$$

This implies that for $N_1 > N_2$, k_o is positive and $P[Z \geq k_o|X\varepsilon\pi_1] < P[Z \geq 0|X\varepsilon\pi_1]$. For $N_1 < N_2$, k_o is negative and $P[Z \leq k_o|X\varepsilon\pi_2] < P[Z \leq 0|X\varepsilon\pi_2]$. This means that if we have more information, i.e., larger sample size, on population π_i than on population π_j, $i,j = 1,2$, the probability of misclassifying an observation from π_i is smaller when we use cutoff point k_o, than when we use zero as cutoff point.

Also, as δ^2 increases, k_o moves away from zero making it easier to classify correctly an observation that comes from the population for which we have more information.

In order to get an idea of the improvement accomplished by using k_o instead of zero as cutoff point, we must look at the difference $P(0) - P(k_o)$. The expression for this difference is so complicated that it is very difficult to reach general conclusions. Hence, we will study this difference in overall probabilities of misclassification by substituting some specific values for the population parameter δ^2, the sample sizes N_1, N_2 and the dimension p.

A numerical evaluation of $P(0) - P(k_o)$ was done for several combinations of p, N_1, N_2 and δ^2 values. Six different values of the population Mahalanobis distance were used. They were selected so that the overall optimum probability of misclassification attained a fixed value. They are given in Table 2.1.

Table 2.1. Mahalanobis distance between the two populations and corresponding overall probability of misclassification

δ^2	$2\,\Phi(-\delta/2)$
1.098	0.6
1.817	0.5
2.836	0.4
4.293	0.3
6.574	0.2
11.842	0.1

For each δ^2, three different values of p were used and for each p value, eight pairs of values for N_1 and N_2 were examined. These are given in Table 2.2.

Table 2.2. Sample sizes and dimension

p = 3		p = 4		p = 8	
N_1	N_2	N_1	N_2	N_1	N_2
4	5	8	9	8	9
4	8	8	16	8	16
4	16	8	32	8	32
4	24	8	48	8	48
10	11	20	21	20	21
10	20	20	40	20	40
10	40	20	80	20	80
10	60	20	120	20	120

The values of $P(0) - P(k_o)$ obtained for each combination of p, N_1, N_2 and δ^2 are given in Tables 2.3 to 2.5.

Several conclusions can be obtained by looking at these tables. First, for any given pair of sample sizes N_1 and N_2, the difference $P(0) - P(k_o)$ decreases as δ^2 increases. This means that for large δ^2 the advantages of using k_o instead of zero as cutoff point become smaller. Second, for any given p and δ^2, $P(0) - P(k_o)$ increases as the ratio N_2/N_1 increases. This indicates that it is better to use k_o than zero as cutoff point when the sample sizes are different. Third, for given N_1, N_2 and δ^2, the difference in overall probabilities of misclassification increases as p increases. This suggests that it is better to use cutoff point k_o when dealing with higher dimensions.

Table 2.3. Values of $P(0) - P(k_o)$ when $p = 3$

N_1	N_2	$\delta^2 = 1.098$	$\delta^2 = 1.817$	$\delta^2 = 2.836$	$\delta^2 = 4.293$	$\delta^2 = 6.574$	$\delta^2 = 11.482$
4	5	0.0002129078	0.0001917481	0.0001789927	0.0001673102	0.0001482368	0.0000983477
4	8	0.0011911980	0.0010766380	0.0010125040	0.0009590387	0.0008699894	0.0006174445
4	16	0.0023151040	0.0021045200	0.0019981260	0.0019205210	0.0017874240	0.0013632170
4	24	0.0026750560	0.0024388430	0.0023254150	0.0022497170	0.0021170370	0.0016615390
10	11	0.0000059009	0.0000059009	0.0000055432	0.0000054240	0.0000045896	0.0000038147
10	20	0.0001735687	0.0001592636	0.0001530051	0.0001491904	0.0001427531	0.0001170039
10	30	0.0003727674	0.0003429055	0.0003308654	0.0003249049	0.0003142953	0.0002636313
10	40	0.0004503727	0.0004143119	0.0004005432	0.0003947020	0.0003833175	0.0003239512

Table 2.4. Values of P(0) - P(k) when p = 4

N_1	N_2	$\delta^2 = 1.098$	$\delta^2 = 1.817$	$\delta^2 = 2.836$	$\delta^2 = 4.293$	$\delta^2 = 6.574$	$\delta^2 = 11.482$
8	9	0.0000212193	0.0000179410	0.0000163317	0.0000147223	0.0000132322	0.0000101328
8	16	0.0003762841	0.0003197789	0.0002921820	0.0002754331	0.0002544522	0.0001969934
8	32	0.0007484555	0.0006406903	0.0005906820	0.0005631447	0.0005286336	0.0004245043
8	48	0.0008745193	0.0007512569	0.0006948113	0.0006642938	0.0006284714	0.0005128384
20	21	0.0000000596	0.0000008345	0.0000007749	0.0000002980	0.0000002980	0.0000005960
20	40	0.0000488162	0.0000434518	0.0000414252	0.0000397563	0.0000382066	0.0000325441
20	80	0.0001039505	0.0000932813	0.0000888109	0.0000866055	0.0000842214	0.0000712276
20	120	0.0001257658	0.0001133680	0.0001075864	0.0001050830	0.0001016855	0.0000874400

Table 2.5. Values of $P(0) - P(k_o)$ when $p = 8$

N_1	N_2	$\delta^2 = 1.098$	$\delta^2 = 1.817$	$\delta^2 = 2.836$	$\delta^2 = 4.293$	$\delta^2 = 6.574$	$\delta^2 = 11.482$
8	9	0.0000479817	0.0000364184	0.0000305176	0.0000265241	0.0000225306	0.0000158548
8	16	0.0007818341	0.0005925894	0.0004994273	0.0004424453	0.0003862381	0.0002748966
8	32	0.0014165040	0.0010690090	0.0009016395	0.0008033514	0.0007105470	0.0005264282
8	48	0.0015848270	0.0011911980	0.0010030260	0.0008941889	0.0007960200	0.0006018281
20	21	0.0000002980	0.0000010729	0.0000009537	0.0000004768	0.0000003576	0.0000007153
20	40	0.0000741482	0.0000604391	0.0000542998	0.0000500679	0.0000462532	0.0000373125
20	80	0.0001463294	0.0001204610	0.0001083612	0.0001016855	0.0000955462	0.0000776052
20	120	0.0001710653	0.0001414418	0.0001270175	0.0001196861	0.0001122355	0.0000930429

Finally, relative to the overall optimal probability of misclassification, and hence to error rate, the differences $P(0) - P(k_o)$ in all cases are small and negligible.

Summarizing, it has been demonstrated that the usual cutoff point, zero, is not always an optimal choice; an alternative cutoff point $k_o = (\frac{1}{N_2} - \frac{1}{N_1})\ (1 + \frac{\delta^2}{4})$ has been found. However, it appears that the improvement of using k_o over zero is very small and the two procedures may be regarded as equivalent. This can be taken as a justification of the use of zero as cutoff point. For this reason we will use zero in the remaining part of this thesis.

III. ESTIMATORS FOR THE UNCONDITIONAL PROBABILITY OF MISCLASSIFICATION

A. Introduction

The unconditional probability of misclassification with the use of the Z statistic and zero cutoff point is given by

$$P_2^*(Z) = P[Z \leq 0 | X \varepsilon \pi_2]. \tag{3.1}$$

Since the exact distribution of Z is not known, we will use the Memon-Okamoto asymptotic expansion for (3.1) given in Equation (2.18) in Chapter II. Since $P_2^*(Z)$ depends on the unknown parameter δ^2, the unconditional probability of misclassification must be estimated from samples.

Let independent samples of sizes N_1 and N_2 be available from the two populations. Denote by $\bar{X}_1$, $\bar{X}_2$ and S the sample means and sample covariance matrix respectively and let $D^2 = (\bar{X}_1 - \bar{X}_2)'S^{-1}(\bar{X}_1 - \bar{X}_2)$ be the Mahalanobis squared distance. Five different estimators of (3.1) with the following general form will be considered in this chapter:

$$Q = \Phi(-\frac{D}{2}) + q, \tag{3.2}$$

where q is a function of D^2 and consists of first order terms of N_1^{-1}, N_2^{-1} and n^{-1}, $n = N_1 + N_2 - 2$.

Other estimators, not of this form, will also be considered in this thesis. The Lachenbruch jackknife estimator will be studied in Section E of this chapter and the apparent error rate will be considered in Chapter IV.

The estimators for the unconditional probability of misclassification are proposed in Section B of this chapter. The asymptotic bias and asymptotic mean square error, for each of the proposed estimators, are derived in Sections C and D, respectively. To obtain these results, we use a theorem by McLachlan (1972 and 1974a). This theorem and its assumptions are given at the end of this section.

In Section E of this chapter, we will obtain the uniformly minimum variance unbiased estimator (UMVUE) of the expectation of the estimator P_L^* given in Equation (3.64). We will also derive expressions for the first two moments of this estimator; the first moment of this estimator will then be used to obtain an exact expression for the unconditional probability of misclassification.

We now state the theorem by McLachlan:

Theorem 1. Consider a function of $\bar{X}_1$, $\bar{X}_2$ and S, $H(\bar{X}_1, \bar{X}_2, S)$, which may also be an explicit function of the parameters μ_1, μ_2 and Σ, assumed to be the identity matrix, and of the sample sizes N_1, N_2. Suppose that H satisfies the following conditions:

a) For all possible values of $\bar{X}_1$, $\bar{X}_2$ and S, $|H| < K$ where K is a nonnegative constant independent of N_1 and N_2.

b) In some neighborhood of the point

$$(\bar{X}_1 = \mu_1, \bar{X}_2 = \mu_2, S = I), \tag{3.3}$$

H is continuous and possesses continuous derivatives of the first to sixth orders inclusive, with respect to the $\bar{X}_{1i}$, $\bar{X}_{2i}$ and S_{ij} $(i \leq j)$, these derivatives all being O(1) with respect to (N_1^{-1}, N_2^{-1} and n^{-1}).

With the notation

$$\partial_{i,j} = \frac{\partial}{\partial(\bar{X}_i)_j} \qquad i = 1,2;\ j = 1,\ldots,p$$

and

$$\partial_{ij} = \partial_{ji} \qquad i,j = 1,\ldots,p$$

$$= \frac{1}{2}(1+\delta_{ij}) \frac{\partial}{\partial(S)_{ij}} \qquad i \leq j = 1,\ldots,p$$

where δ_{ij} is the Kronecker's delta, the second order expansion of the expectation of H over the joint distribution of $\bar{X}_1$, $\bar{X}_2$ and S is given by

$$E(H) = \Theta H(\mu_1, \mu_2, I) + O_3$$

where Θ is the differential operator defined formally by

$$\Theta = 1 + \sum^{p} [\frac{1}{2N_1} \partial^2_{1,i} + \frac{1}{2N_2} \partial^2_{2,i} + \frac{1}{n} \partial^2_{ij} + \frac{1}{8N_1^2} \partial^2_{1,i} \partial^2_{1,j}$$

$$+ \frac{1}{8N_2^2} \partial^2_{2,i} \partial^2_{2,j} + \frac{1}{4N_1N_2} \partial^2_{1,i} \partial^2_{2,j} + \frac{1}{2N_1 n} \partial^2_{1,i} \partial^2_{jk}$$

$$+ \frac{1}{2N_2 n} \partial^2_{2,i} \partial^2_{jk} + \frac{1}{n^2} \{\frac{1}{2} \partial^2_{ij} \partial^2_{km} + \frac{4}{3} \partial_{ij}\partial_{jk}\partial_{ki}\}], \qquad (3.4)$$

and $\Theta H(\mu_1, \mu_2, \Sigma)$ is the value of ΘH at the point (3.3); the single summation symbol Σ^p denotes summation over the range 1 to p of the subindices appearing in the summands.

B. Estimators Based on the Normal Distribution

When the population are assumed to be normally distributed and the parameters μ_1, μ_2 and Σ are known, the optimal probability of misclassification is

$$\Phi(-\delta/2). \tag{3.5}$$

The first two estimators for the unconditional probability of misclassification, $P_2^*(Z)$, denoted by Q_D and Q_{DS}, are defined as

$$Q_D = \Phi(-\frac{D}{2}) \tag{3.6}$$

and

$$Q_{DS} = \Phi(-\frac{D^*}{2}), \tag{3.7}$$

where

$$D^2 = (\bar{X}_2 - \bar{X}_1)'S^{-1}(\bar{X}_2 - \bar{X}_1) \tag{3.8}$$

and

$$D^{*2} = (\frac{n-p-1}{n})D^2. \tag{3.9}$$

The next two estimators for $P_2^*(Z)$, denoted by Q_{MO} and Q_{MOS}, are defined as

$$Q_{MO} = \Phi(-\frac{D}{2}) + \hat{A}_1 \tag{3.10}$$

and

$$Q_{MOS} = \Phi(-\frac{D^*}{2}) + \hat{A}_1^*, \tag{3.11}$$

where $\hat{A}_1$ and $\hat{A}_1^*$ are obtained by replacing δ^2 by D^2 and D^{*2} respectively in the linear term of Memon-Okamoto expansion for $P_2^*(Z)$ given in Equation (2.22).

Note that the estimators Q_D and Q_{DS} can also be obtained from the Memon-Okamoto expansion in a similar way, but using only the leading term.

The last estimator of the form in (3.2), denoted by Q_{MC}, is obtained as in McLachlan (1974a and 1974c), with the linear term in the Okamoto (1963) expansion replaced by the linear term in the Memon-Okamoto (1971) expansion. This estimator is defined as

$$Q_{MC} = \Phi(-\tfrac{D}{2}) + \hat{A}_1 - \hat{B}_1, \tag{3.12}$$

where $\hat{B}_1$ is obtained by replacing δ by D in

$$B_1 = \frac{1}{16}\phi(-\frac{\delta}{2})\{(\frac{1}{N_1}+\frac{1}{N_2})(\delta-4(p-1)\delta^{-1}+\frac{\delta}{2n}(\delta^2-4(2p+1))\}, \tag{3.13}$$

which is obtained from $\Theta_1\ \Phi(-\frac{D}{2})$, where Θ_1 is the operator consisting only of the first order terms with respect to N_1^{-1}, N_2^{-1} and n^{-1} of the operator Θ given in Theorem 1.

We will now show that all these five estimators take the form given in (3.2). It is obvious that Q_D can be written as in (3.2), with q equal to zero.

Since

$$\begin{aligned}(\frac{n-p-1}{n})^{1/2} &= (1 - \frac{p+1}{n})^{1/2}\\ &= (1 - \frac{p+1}{2n} + 0_2),\end{aligned}$$

$$Q_{DS} = \Phi(-\frac{D^*}{2})$$

$$= \Phi(-\frac{D}{2}(\frac{n-p-1}{n})^{1/2})$$

$$= \Phi(-\frac{D}{2}(1 - \frac{1}{2}\frac{p+1}{n} + o_2))$$

$$Q_{DS} = \Phi(-\frac{D}{2} + \frac{p+1}{4n} D + o_2)$$

$$= \Phi(-\frac{D}{2}) + \frac{p+1}{4n} D\phi(-\frac{D}{2}) + o_2. \tag{3.14}$$

with the last expression obtained by Taylor's series expansion.

Similarly, we can show that

$$\hat{A}_1^* = \hat{A}_1 + \frac{p+1}{4n} D\phi(-\frac{D}{2}) + o_2, \tag{3.15}$$

and hence all the proposed estimators take the form given in (3.2), with the q terms as defined in Table 3.1.

Table 3.1. Estimators for $P_2^*(Z)$ and the corresponding first order terms

Estimator	First order term
Q_D	$q_D = 0$
Q_{DS}	$q_{DS} = (\frac{p+1}{4n})D\phi(-\frac{D}{2})$
Q_{MO}	$q_{MO} = \hat{A}_1$
Q_{MOS}	$q_{MOS} = \hat{A}_1 + (\frac{p+1}{4n})D\phi(-\frac{D}{2})$
Q_{MC}	$q_{MC} = \hat{A}_1 - \hat{B}_1$

C. Asymptotic Biases

In this section, we will use Theorem 1 to obtain an expression for the asymptotic bias of each one of the estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} of the unconditional probability of misclassification, $P_2^*(Z)$.

Since $0 \leq P_2^*(Z) \leq 1$ and $0 \leq \Phi(\cdot) \leq 1$, it follows that $|P_2^*(Z)-\Phi(-\frac{D}{2})|$ and $|P_2^*(Z)-\Phi(-\frac{D^*}{2})|$ are bounded by one. In order to show that $|P_2^*(Z)-Q_{MO}|$, $|P_2^*(Z) - Q_{MOS}|$ and $|P_2^*(Z) - Q_{MC}|$ are bounded by a constant which is independent of N_1 and N_2, we follow McLachlan (1974a), and assume that $D^2 > \delta_o$ with δ_o a very small positive constant. This restriction on D^2 has to be imposed since the estimators Q_{MO}, Q_{MOS} and Q_{MC} involve powers of D^{-2}. Then, under the condition that $D^2 > \delta_o$, we can verify that $|P_2^*(Z)-Q_{MO}|$, $|P_2^*(Z)-Q_{MOS}|$ and $|P_2^*(Z)-Q_{MC}|$ indeed satisfy assumption a) of Theorem 1. McLachlan (1974a) showed that Theorem 1 still holds when D^2 is assumed to be greater than δ_o.

Since the five estimators considered here depend on $\bar{X}_1$, $\bar{X}_2$ and S only through $D^2 = (\bar{X}_2-\bar{X}_1)'S^{-1}(\bar{X}_2-\bar{X}_1)$, we can verify, using results on the derivatives of a quadratic form with respect to the elements of a matrix, that condition b of Theorem 1 is also satisfied.

Using Theorem 1 and Equation (2.21), we have,

$$\begin{aligned}
\text{Bias } (Q_D) &= E[P_2^*(Z)-\Phi(-\tfrac{D}{2})] \\
&= \Theta[\Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2})] + 0_3 \\
&= \Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Theta\Phi(-\tfrac{D}{2}) + 0_3 \\
&= A_1 - B_1 + 0_2. \qquad (3.16)
\end{aligned}$$

where A_1 and A_2 are defined in Section B of Chapter II and B_1 is given in Equation (3.13).

Using q_{DS} as defined in Table 3.1 and Theorem 1, it follows that,

$$\begin{aligned}\text{Bias }(Q_{DS}) &= E[P_2^*(Z)-\Phi(-\tfrac{D}{2}) - q_{DS}]\\ &= \Theta[\Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2}) - q_{DS}] + 0_3\\ &= \Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Theta\Phi(-\tfrac{D}{2}) - \Theta q_{DS} + 0_3\\ &= A_1 - B_1 - \frac{p+1}{4n}\,\delta\phi(-\tfrac{\delta}{2}) + 0_2,\end{aligned} \tag{3.17}$$

since $B_1 = \Theta_1\Phi(-\frac{D}{2})$.

Using Equations (2.22) and (3.13), we have,

$$\begin{aligned}\text{Bias }(Q_{DS}) = \frac{1}{16}\,\phi(-\tfrac{\delta}{2})\{&\frac{2}{N_1}(\delta + 4(p-1)\delta^{-1})\\ &+ \frac{2}{N_2}(-\delta + 4(p-1)\delta^{-1})\\ &+ \frac{2\delta}{n}(2p - 3 - \frac{\delta^2}{4})\} + 0_2\end{aligned} \tag{3.18}$$

Using q_{MO}, q_{MOS}, and q_{MC} from Table 3.1 and applying repeatedly Theorem 1, we have,

$$\begin{aligned}\text{Bias }(Q_{MO}) &= E[P_2^*(Z) - Q_{MO}]\\ &= \Theta[\Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2}) - \hat{A}_1] + 0_3\\ &= -B_1 + 0_2\\ &= -\frac{1}{16}\phi(-\tfrac{\delta}{2})\{(\frac{1}{N_1} + \frac{1}{N_2})(\delta-4(p-1)\delta^{-1}) +\end{aligned}$$

$$\frac{\delta}{2n}(\delta^2-4(2p+1))\} + O_2, \tag{3.19}$$

$$\begin{aligned}\text{Bias }(Q_{MOS}) &= E[P_2^*(Z) - Q_{MOS}]\\ &= \Theta[\phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2}) - \hat{A}_1 - q_{DS}] + O_3\\ &= -(B_1 + \tfrac{p+1}{4n}\,\delta\phi(-\tfrac{\delta}{2})) + O_2,\end{aligned} \tag{3.20}$$

and

$$\begin{aligned}\text{Bias }(Q_{MC}) &= E[P_2^*(Z) - Q_{MC}]\\ &= \Theta[\Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2}) - \hat{A}_1 + \hat{B}_1] + O_3\\ &= A_1 - B_1 - (A_1 - B_1) + O_2\\ &= \text{zero} + O_2.\end{aligned} \tag{3.21}$$

D. Asymptotic Mean Square Errors

To obtain an expression for the asymptotic mean square error for each one of the estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC}, we will use Theorem 1 in Section A of this chapter. In doing so, we will first derive an expression for the asymptotic mean square error (AMSE) of an estimator Q of the form given in (3.2). Then, the AMSE for each one of the estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} will be obtained as a particular case of this general result.

The argument used to verify that the assumptions of Theorem 1 were satisfied by the functions $|P_2^*(Z) - Q|$, where Q can be any of the five estimators considered, can again be used to verify that the function $[P_2^*(Z) - Q]^2$ also satisfies assumptions a) and b) of that theorem.

From Theorem 1 and Equation (2.21), we have that,

$$\begin{aligned}
\mathrm{AMSE}(Q) &= E[P_2^*(Z) - Q]^2 \\
&= \Theta[P_2^*(Z) - Q]^2 + 0_3 \\
&= \Theta[\Phi(-\tfrac{\delta}{2}) + A_1 + A_2 - \Phi(-\tfrac{D}{2}) - q + 0_3]^2 + 0_3 \\
&= \Theta[\Phi(-\tfrac{\delta}{2}) - \Phi(-\tfrac{D}{2})]^2 + \Theta[A_1 + A_2 - q]^2 \\
&\quad + \Theta\{2[\Phi(-\tfrac{\delta}{2}) - \Phi(-\tfrac{D}{2})][A_1 + A_2 - q]\} + 0_3 \\
&= m_1(Q_D) + [A_1 - q(\delta)]^2 + 2\Phi(-\tfrac{\delta}{2})[A_1 + A_2 - \Theta[q]] \\
&\quad - \Theta\{2\Phi(-\tfrac{D}{2})(A_1 + A_2 - q)\} + 0_3 \\
&= m_1(Q_D) + [A_1 - q(\delta)]^2 - 2\Phi(-\tfrac{\delta}{2})\Theta[q] \\
&\quad - 2A_1\Theta_1[\Phi(-\tfrac{D}{2})] + 2\Theta[q\Phi(-\tfrac{D}{2})] + 0_3 \\
&= m_1(Q_D) + [A_1 - q(\delta)]^2 - 2A_1B_1 + T + 0_3,
\end{aligned} \tag{3.22}$$

where

$$\begin{aligned}
T &= 2\Theta[q\Phi(-\tfrac{D}{2})] - 2\Phi(-\tfrac{\delta}{2})\Theta[q] \\
&= 2\Theta_1[q\Phi(-\tfrac{D}{2})] - 2\Phi(-\tfrac{\delta}{2})\Theta_1[q],
\end{aligned} \tag{3.23}$$

Θ_1 is as defined in Theorem 1,

$$m_1(Q_D) = E[\Phi(-\tfrac{\delta}{2}) - \Phi(-\tfrac{D}{2})]^2$$

and $q(\delta)$ is a function of δ, N_1, N_2 and p obtained by replacing D by δ in the function q defined in Table 3.1.

Since $m_1(Q_D)$ appears in the expression for the AMSE of each one of the estimators, we now obtain its form up to the first order terms. From Theorem 1, we have

$$\begin{aligned} m_1(Q_D) &= E[\Phi(-\tfrac{\delta}{2})-\Phi(-\tfrac{D}{2})]^2 \\ &= \Theta[\Phi(-\tfrac{\delta}{2})-\Phi(-\tfrac{D}{2})]^2 + O_3 \\ &= [\Phi(-\tfrac{\delta}{2})]^2-2\Phi(-\tfrac{\delta}{2})\Theta[\Phi(-\tfrac{D}{2})] \\ &\quad + \Theta[\Phi(-\tfrac{D}{2})]^2 + O_3 \\ &= -2\Phi(-\tfrac{\delta}{2})\Theta_1[\Phi(-\tfrac{D}{2})]+\Theta_1[\Phi(-\tfrac{D}{2})]^2 + O_2 \\ &= -2\Phi(-\tfrac{\delta}{2})B_1 + \Theta_1[\Phi(-\tfrac{D}{2})]^2 + O_2, \end{aligned} \tag{3.24}$$

where B_1 is given in Equation (3.13) and $\Theta_1[\Phi(-\frac{D}{2})]^2$ is obtained as follows: From Theorem 1

$$\Theta_1[\Phi(-\tfrac{D}{2})]^2 = \sum^{p}\{(\frac{1}{2N_1}\,\partial^2_{1,i} + \frac{1}{2N_2}\,\partial^2_{2,i} + \frac{1}{n}\,\partial^2_{ij})\}[\Phi(-\tfrac{D}{2})]^2\Big|_o \tag{3.25}$$

where the notation $\Big|_o$ indicates the value evaluated at the point (3.3). Since the distribution of the Z classification statistic is invariant under any linear transformation on the observations X, we may suppose without loss of generality that, $\mu_1 = \mu_o$, $\mu_2 = -\mu_o$ and S = I, where μ_o denotes a vector with first component equal to $\delta/2$ and the other components equal to zero.

The contribution of the $\frac{1}{N_1}$ term in (3.25) can now be obtained by first setting $\bar{X}_2 = \mu_2$, $S = I$ and hence,

$$D^2 = \sum_{i=1}^{p} (\bar{X}_{1i} - \mu_{2i})^2 \tag{3.26}$$

and then taking derivatives with respect to $\bar{X}_{1i}$ and evaluating at the point (3.3). Proceeding in this way, we have that,

$$\begin{aligned} \partial_{1,i}[\Phi(-\tfrac{D}{2})]^2 &= 2\Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})(-\tfrac{1}{2})(\tfrac{1}{2})D^{-1}(2)(\bar{X}_{1i} - \mu_{2i}) \\ &= -\Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})D^{-1}(\bar{X}_{1i} - \mu_{2i}), \end{aligned}$$

and

$$\begin{aligned} \partial^2_{1,i}[\Phi(-\tfrac{D}{2})]^2 &= -\Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})[D^{-1}-(\frac{D^{-1}}{4} + D^{-3})(\bar{X}_{1i} - \mu_{2i})^2] \\ &\quad + \tfrac{1}{2}[\phi(-\tfrac{D}{2})]^2 D^{-1}(\bar{X}_{1i} - \mu_{2i})^2. \end{aligned}$$

so that,

$$\begin{aligned} \frac{1}{2N_1}\sum_{i=1}^{p} \partial^2_{1,i}\{[\Phi(-\tfrac{D}{2})]^2\}\Big|_o &= \frac{1}{2N_1}\Phi(-\tfrac{\delta}{2})\phi(-\tfrac{\delta}{2})[\tfrac{1}{4}(\delta-4(p-1)\delta^{-1})] \\ &\quad + \frac{1}{4N_1}[\phi(-\tfrac{\delta}{2})]^2. \end{aligned} \tag{3.27}$$

Similarly, the contribution of the $\frac{1}{N_2}$ terms in (3.25) is

$$\begin{aligned} \frac{1}{2N_2}\sum_{i=1}^{p} \partial^2_{2,i}\{[\Phi(-\tfrac{D}{2})]^2\}\Big|_o &= \frac{1}{2N_2}\Phi(-\tfrac{\delta}{2})\phi(-\tfrac{\delta}{2})[\tfrac{1}{4}(\delta-4(p-1)\delta^{-1})] \\ &\quad + \frac{1}{4N_2}[\phi(-\tfrac{D}{2})]^2. \end{aligned} \tag{3.28}$$

To obtain the contribution of the $\frac{1}{n}$ term in (3.25), we take derivatives with respect to the (i,j)th element of the inverse of the sample variance covariance matrix S. After setting $\bar{X}_1 = \mu_1$, $\bar{X}_2 = \mu_2$ and hence,

$$D^2 = \delta^2(S'') \tag{3.29}$$

where (S") denotes the (1,1)th element of S^{-1}, we can obtain $\partial^2_{ij}[\Phi(-\frac{D}{2})]\Big|_o$ by using Lemmas 2 and 3 in Okamoto (1963). Proceeding in this manner, we have,

$$\begin{aligned}\partial_{ij}[\Phi(-\tfrac{D}{2})]^2 &= 2\Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})(-\tfrac{\delta}{2})(\tfrac{1}{2})(S'')^{-1/2}(S'')_{ij}\\ &= -\tfrac{\delta}{2}\ \Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})(S'')^{-1/2}(S'')_{ij}\end{aligned}$$

and

$$\begin{aligned}\partial^2_{ij}[\Phi(-\tfrac{D}{2})]^2 &= -\tfrac{\delta}{2}\Phi(-\tfrac{D}{2})\phi(-\tfrac{D}{2})\{(S'')^{-1/2}(S'')_{ij,ij}-\tfrac{1}{2}(S'')^{-3/2}(S'')^2_{ij}\\ &\quad -\tfrac{\delta^2}{8}(S'')^{-1/2}(S'')^2_{ij}\}\\ &\quad +\tfrac{\delta^2}{8}\ [\phi(-\tfrac{D}{2})]^2(S'')^{-1}(S'')^2_{ij}.\end{aligned}$$

It then follows that,

$$\begin{aligned}\frac{1}{n}\ \Sigma\partial^2_{ij}\{[\Phi(-\tfrac{D}{2})]^2\}\Big|_o &= \tfrac{1}{n}\{[-\tfrac{\delta}{2}\Phi(-\tfrac{\delta}{2})\phi(-\tfrac{\delta}{2})(p+\tfrac{1}{2}-\tfrac{\delta^2}{8})]\\ &\quad +\tfrac{\delta^2}{8}\ (\phi(-\tfrac{\delta}{2}))^2\}.\end{aligned} \tag{3.30}$$

Substituting (3.27), (3.28) and (3.30) in (3.25), we have

$$\Theta_1[\Phi(-\tfrac{D}{2})]^2 = (\frac{1}{8N_1} + \frac{1}{8N_2})\Phi(-\tfrac{\delta}{2})\phi(-\tfrac{\delta}{2})[\delta-4(p-1)\delta^{-1}]$$

$$+ (\frac{1}{N_1} + \frac{1}{N_2})[\phi(-\tfrac{\delta}{2})/2]^2$$

$$+ \frac{1}{n}[\frac{\delta}{16}\,\Phi(-\tfrac{\delta}{2})\phi(-\tfrac{\delta}{2})(\delta^2-4(2p+1))]$$

$$+ \frac{1}{2n}\,\delta^2[\phi(-\tfrac{\delta}{2})/2]^2 \tag{3.31}$$

Finally, substituting (3.31) in (3.24), we obtain that,

$$m_1(Q_D) = [\tfrac{1}{2}\phi(-\tfrac{\delta}{2})]^2\{\frac{1}{N_1} + \frac{1}{N_2} + \frac{\delta^2}{2n}\} + O_2 \tag{3.32}$$

Using (3.22) we can now derive the expressions for the AMSE of each one of the estimators.

$$\mathrm{AMSE}(Q_D) = m_1(Q_D) + A_1^2 - 2A_1B_1 + O_3$$

$$m_1(Q_D) + A_1(A_1 - 2B_1) + O_3. \tag{3.33}$$

$$\mathrm{AMSE}(Q_{DS}) = m_1(Q_D) + [A_1 - q_{DS}(\delta)]^2 - 2A_1B_1 + T_{DS} + O_3 \tag{3.34}$$

$$= \mathrm{AMSE}(Q_D) + q_{DS}(\delta)[q_{DS}(\delta) - 2A_1] + T_{DS}, \tag{3.35}$$

where

$$T_{DS} = 2\Theta_1[q_{DS}\Phi(-\tfrac{D}{2})] - 2\Phi(-\tfrac{\delta}{2})\Theta_1[q_{DS}]. \tag{3.36}$$

In order to obtain T_{DS}, we will proceed in a similar way in obtaining (3.31). Denote by T_{DS1}, the contribution of the $\frac{1}{N_1}$ term of T_{DS}, then

$$T_{DS1} = -2\Phi(-\frac{\delta}{2})(\frac{p+1}{4n})\sum_{i=1}^{p}(\frac{1}{2N_1}\partial^2_{1,i})\{D\phi(-\frac{D}{2})\}\Big|_o$$

$$+2(\frac{p+1}{4n})\sum_{i=1}^{p}(\frac{1}{2N_1}\partial^2_{1,i})\{D\Phi(-\frac{D}{2})\phi(-\frac{D}{2})\}\Big|_o. \qquad (3.37)$$

When we examine the contribution of the derivatives

$$\partial^2_{1,i}\{D\phi(-\frac{D}{2})\}$$

and

$$\partial^2_{1,i}[D\Phi(-\frac{D}{2})\phi(-\frac{D}{2})] = \partial_{1,i}\{\partial_{1,i}[D\Phi(-\frac{D}{2})\phi(-\frac{D}{2})]\} -$$

$$= \partial^2_{1,i}[D\phi(-\frac{D}{2})]\Phi(-\frac{D}{2})+D\phi(-\frac{D}{2})\partial^2_{1,i}[\Phi(-\frac{D}{2})]$$

$$+ 2\partial_{1,i}[D\phi(-\frac{D}{2})]\partial_{1,i}[\Phi(-\frac{D}{2})], \qquad (3.38)$$

we observe that, after taking the derivatives in (3.37) and evaluating the resulting expression at the point (3.3), the first term on the right hand side of (3.38) will cancel with the first term on the right hand side of (3.37). After this simplification, T_{DS1} reduces to,

$$T_{DS1} = (\frac{p+1}{2n})\sum_{i=1}^{p}(\frac{1}{2N_1}\{2\partial_{1,i}[D\phi(-\frac{D}{2})]\partial_{1,i}[\Phi(-\frac{D}{2})]$$

$$+ D\phi(-\frac{D}{2})\partial^2_{1,i}[\Phi(-\frac{D}{2})]\})\Big|_o. \qquad (3.39)$$

Using $D^2 = \sum_{i=1}^{p} (\bar{X}_{1i} - \mu_{2i})^2$, we have that

$$\partial_{1,i}[\Phi(-\tfrac{D}{2})] = \phi(-\tfrac{D}{2})(-\tfrac{1}{2})(+\tfrac{1}{2})D^{-1}(2)(\bar{X}_{1i} - \mu_{2i})$$

$$= -\tfrac{1}{2}\phi(-\tfrac{D}{2})D^{-1}(\bar{X}_{1i} - \mu_{2i}),$$

$$\partial^2_{1,i}[\Phi(-\tfrac{D}{2})] = -\tfrac{1}{2}\phi(-\tfrac{D}{2})[D^{-1}-(D^{-3} + \tfrac{D^{-1}}{4})(\bar{X}_{1i} - \mu_{2i})^2]$$

and

$$\partial_{1,i}[D\phi(-\tfrac{D}{2})] = \phi(-\tfrac{D}{2})\partial_{1,i}(D)+D\partial_{1,i}[\phi(-\tfrac{D}{2})]$$

$$= \phi(-\tfrac{D}{2})(D^{-1} - \tfrac{D}{4})(\bar{X}_{1i} - \mu_{2i}).$$

Substituting these in (3.39), it follows that

$$T_{DS1} = (\tfrac{p+1}{4N_1 n})\sum_{i=1}^{p}\{2\phi(-\tfrac{D}{2})(\bar{X}_{1i}-\mu_{2i})(D^{-1}-\tfrac{D}{4})[-\tfrac{1}{2}\phi(-\tfrac{D}{2})D^{-1}(\bar{X}_{1i}-\mu_{2i})]$$

$$+ D\phi(-\tfrac{D}{2})[-\tfrac{1}{2}\phi(-\tfrac{D}{2})[D^{-1}-(D^{-3}+\tfrac{D^{-1}}{4})(\bar{X}_{1i}-\mu_{2i})^2]]\}\Big|_o$$

$$= \frac{p+1}{4N_1 n}[\phi(-\tfrac{\delta}{2})]^2(\frac{3\delta^2}{8} - \frac{p+1}{2})$$

$$= \frac{p+1}{32N_1 n}[\phi(-\tfrac{\delta}{2})]^2[3\delta^2-4(p+1)]. \tag{3.40}$$

Similarly, it can be shown that the contribution to the $\frac{1}{N_2}$ term in T_{DS} is given by

$$T_{DS2} = \frac{p+1}{32N_2 n} [\phi(-\frac{\delta}{2})]^2 [3\delta^2 - 4(p+1)]. \tag{3.41}$$

Now let T_{DS3} be the contribution of the $\frac{1}{n}$ term of T_{DS}. Then we have that,

$$T_{DS3} = -2\Phi(-\frac{\delta}{2})(\frac{p+1}{4n}) \sum_{i,j}^{p} (\frac{1}{n} \partial_{ij}^2)\{D\phi(-\frac{D}{2})\}\Big|_o$$

$$+ 2(\frac{p+1}{4n}) \sum_{i,j}^{p} (\frac{1}{n} \partial_{ij}^2)\{D\Phi(-\frac{D}{2})\phi(-\frac{D}{2})\}\Big|_o . \tag{3.42}$$

It is not difficult to show that after taking derivatives and evaluating the resulting expression at (3.3), T_{DS3} reduces to

$$T_{DS3} = \frac{p+1}{2n^2} \sum_{i,j}^{p} \{2\partial_{ij}[D\phi(-\frac{D}{2})]\partial_{ij}[\Phi(-\frac{D}{2})]$$

$$+ D\phi(-\frac{D}{2})\partial_{ij}^2[\Phi(-\frac{D}{2})\}\Big|_o . \tag{3.43}$$

Using $D^2 = \delta^2(S'')$, we have that

$$\partial_{ij}[\Phi(-\frac{D}{2})] = -\frac{\delta}{4}\, \phi(-\frac{D}{2})[(S'')^{-1/2}(S'')_{ij}],$$

$$\partial_{ij}^2[\Phi(-\frac{D}{2})] = \frac{\delta}{32}\phi(-\frac{D}{2})[\delta^2(S'')^{-1/2}(S'')_{ij}^2 + 4(S'')_{ij}^2(S'')^{-3/2}$$

$$-8(S'')^{-1/2}(S'')_{ij,ij}]$$

and

$$\partial_{ij}[D\phi(-\frac{D}{2})] = \phi(-\frac{D}{2})\partial_{ij}[D] + D\partial_{ij}[\phi(-\frac{D}{2})]$$

$$= \phi(-\frac{D}{2})[\frac{\delta}{2}(S'')^{-1/2}(S'')_{ij} - \frac{\delta^3}{8}(S'')^{1/2}(S'')_{ij}].$$

It follows that,

$$T_{DS3}=(\frac{p+1}{2n^2})\sum_{ij}\{2\phi(-\frac{D}{2})[\frac{\delta}{2}(S")^{-1/2}(S")_{ij}-\frac{\delta^3}{8}(S")^{1/2}(S")_{ij}][-\frac{\delta}{4}\phi(-\frac{D}{2})(S")^{-1/2}(S")_{ij}]$$

$$+\delta(S")^{1/2}\phi(-\frac{D}{2})(\frac{\delta}{32}\phi(-\frac{D}{2})[\delta^2(S")^{-1/2}(S")^2_{ij}+4(S")^2_{ij}(S")^{-3/2}$$

$$-8(S")^{-1/2}(S")_{ij,ij}])\}\Big|_o$$

$$=\frac{p+1}{2n^2}[\phi(-\frac{\delta}{2})]^2\{(\frac{\delta^4}{16}-\frac{\delta^2}{4})+\frac{\delta^2}{32}(\delta^2-4(2p+1))$$

$$=\frac{p+1}{64n^2}[\phi(-\frac{\delta}{2})]^2[\delta^2(3\delta^2-12-8p)]. \qquad (3.44)$$

Substituting Equations (3.40), (3.41) and (3.44) in Equation (3.36) we obtain

$$T_{DS}=\frac{p+1}{32n}[\phi(-\frac{\delta}{2})]^2\{(\frac{1}{N_1}+\frac{1}{N_2})(3\delta^2-4(p+1))+\frac{\delta^2}{2n}(3\delta^2-12-8p)\}. \qquad (3.45)$$

From Equation (3.22) and using q_{MO} as defined in Table 3.1, we have that

$$AMSE(Q_{MO}) = m_1(Q_D) + [A_1 - q_{MO}(\delta)]^2 - 2A_1B_1 + T_{MO} + O_3$$

$$= m_1(Q_D) - 2A_1B_1 + T_{MO} + O_3, \qquad (3.46)$$

where

$$T_{MO} = 2\Theta_1[q_{MO}\Phi(-\frac{D}{2})]-2\phi(-\frac{\delta}{2})\Theta_1[q_{MO}], \qquad (3.47)$$

which is obtained following the same procedure used to obtain T_{DS}.

Letting $T_{MO} = T_{MO1} + T_{MO2} + T_{MO3}$, we have that

$$T_{MO1} = \frac{\delta}{N_1}[\phi(-\frac{\delta}{2})]^2\{\frac{\delta^{-3}}{16}[(\frac{1}{N_2})(\delta^2+4(p-1))-(\frac{1}{N_1})(3\delta^2-4(p-1))]-\frac{\delta^{-1}}{4n}(p-1)$$

$$+(\frac{1}{4})[\frac{1}{16N_2}(-\delta+4(p-1)\delta^{-1})+\frac{1}{16N_1}(3\delta+4(p-1)\delta^{-1})+\frac{\delta}{4n}(p-1)]$$

$$\cdot(1+\frac{1}{2}(\delta-4(p-1)\delta^{-1}))\}, \qquad (3.48)$$

$$T_{MO2} = \frac{N_1}{N_2}\, T_{MO1} \qquad (3.49)$$

and

$$T_{MO3} = \frac{\delta}{n}[\phi(-\frac{\delta}{2})]^2\{\frac{\delta^{-1}}{32}[\frac{1}{N_2}(\delta^2+4(p-1))-\frac{1}{N_1}(3\delta^2-4(p-1))]-\frac{\delta}{8n}(p-1)$$

$$+\frac{1}{16}[\frac{1}{16N_2}(-\delta+4(p-1)\delta^{-1})+\frac{1}{16N_1}(3\delta+4(p-1)\delta^{-1})+\frac{\delta}{4n}(p-1)]$$

$$\cdot(3\delta^2-4(2p+1))\}. \qquad (3.50)$$

Again, from Equation (3.22) and with q_{MOS} as defined in Table 3.1, we have that

$$\mathrm{AMSE}(Q_{MOS}) = m_1(Q_D) + [A_1 - q_{MOS}(\delta)]^2 - 2A_1B_1 + T_{MOS} + O_3$$

$$= m_1(Q_D) + [q_{DS}(\delta)]^2 - 2A_1B_1 + T_{MO} + T_{DS} + O_3 \qquad (3.51)$$

$$= \mathrm{AMSE}(Q_{MO}) + [q_{DS}(\delta)]^2 + T_{DS}. \qquad (3.52)$$

The last expression for AMSE(Q_{MOS}) will be useful when comparing the relative performance of the estimators with respect to AMSE.

Finally, from Equation (3.22) and with q_{MC} as defined in Table 3.1, we have that,

$$\begin{aligned}
AMSE(Q_{MC}) &= m_1(Q_D) + [A_1 - q_{MC}(\delta)]^2 - 2A_1B_1 + T_{MC} + O_3 \\
&= m_1(Q_D) + B_1(B_1 - 2A_1) + T_{MC} + O_3 \\
&= m_1(Q_D) + B_1(B_1 - 2A_1) + T_{MO} - R + O_3, \qquad (3.53)
\end{aligned}$$

where

$$\begin{aligned}
T_{MC} &= 2\Theta_1[q_{MC}\Phi(-\tfrac{D}{2})] - 2\Phi(-\tfrac{\delta}{2})\Theta_1[q_{MC}] \\
&= T_{MO} - R, \qquad (3.54)
\end{aligned}$$

with

$$R = 2\Theta_1[\hat{B}_1\Phi(-\tfrac{D}{2})]-2\Phi(-\tfrac{\delta}{2})\Theta_1[\hat{B}_1] \qquad (3.55)$$

and

$$\hat{B}_1 = \frac{1}{16}\phi(-\tfrac{D}{2})[(\frac{1}{N_1}+\frac{1}{N_2})(D-4(p-1)D^{-1})+\frac{D}{2n}(D^2-4(2p+1))]. \qquad (3.56)$$

Finally, in order to obtain R, we proceed as we did when we obtained T_{DS} and T_{MO}. Writing (3.55) as $R = R_1 + R_2 + R_3$, we have that

$$\begin{aligned}
R_1 = \frac{\delta}{N_1}[\phi(-\tfrac{\delta}{2})]^2\{&-\delta^{-3}[\frac{1}{16}(\frac{1}{N_1}+\frac{1}{N_2})(\delta^2+4(p-1))+\frac{\delta^3}{32n}(3\delta-4(2p+1)\delta^{-1})] \\
&+(\tfrac{1}{4})[\frac{1}{16}(\frac{1}{N_1}+\frac{1}{N_2})(\delta-4(p-1)\delta^{-1}+\frac{\delta}{32n}(\delta^2-4(2p+1))] \\
&\cdot(1+\tfrac{1}{2}(\delta-4(p-1)\delta^{-1}))\}, \qquad (3.57)
\end{aligned}$$

$$R_2 = \frac{N_1}{N_2} R_1 \tag{3.58}$$

and

$$R_3 = \frac{\delta}{n}[\phi(-\frac{\delta}{2})]^2\{-\frac{\delta^{-1}}{2}[\frac{1}{16}(\frac{1}{N_1}+\frac{1}{N_2})(\delta^2+4(p-1))+\frac{\delta^3}{32n}(3\delta-4(2p+1)\delta^{-1})]$$

$$+(\frac{1}{16})[\frac{1}{16}(\frac{1}{N_1}+\frac{1}{N_2})(\delta-4(p-1)\delta^{-1})+\frac{\delta}{32n}(\delta^2-4(2p+1))]$$

$$\cdot(3\delta^2 - 4(2p+1))\}. \tag{3.59}$$

In summary, the expressions for the AMSE of Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} are given in Equations (3.33), (3.34), (3.46), (3.51) and (3.54) respectively.

E. Jackknife Estimator

In this section we consider a jackknife classification procedure which produces an estimator whose properties can be used to derive an expression for the unconditional probability of misclassification $P_1^*(Z)$. Since this probability depends on the sample size N_1 and N_2, we will write $P_1^*(Z)$ as $\alpha(N_1, N_2)$ whenever this dependence needs to be explicit.

Let us consider the Lachenbruch (1967) jackknife estimator P_L, defined as,

$$P_L = \frac{1}{N_1} \sum_{j=1}^{N_1} K_j, \tag{3.60}$$

where

$$K_j = \begin{cases} 1 & C(X_j, \bar{X}_{1(j)}, \bar{X}_2, S_{(j)}) \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.61}$$

$$\begin{aligned} C(X_j,\bar{X}_{1(j)},\bar{X}_2,S_{(j)}) &= \frac{N_1-1}{N_1}(X_j-\bar{X}_{1(j)})'S_{(j)}^{-1}(X_j-\bar{X}_{1(j)}) \\ &\quad - \frac{N_2}{N_2+1}(X_j-\bar{X}_2)'S_{(j)}^{-1}(X_j-\bar{X}_2), \end{aligned} \tag{3.62}$$

and a statistic with a subscript (j) indicates that the observation X_j has been removed from the calculation of that statistic.

The estimator P_L is obtained by successively omitting an observation from the first sample and calculating the classification statistic based on the remaining $(N_1-1)+N_2$ observations. This modified classification statistic is then used to classify the observation that was left out. For each of the N_1 classifications, the omitted observation and the corresponding classification statistic are independent. Hence, the proportion of misclassified observations is an unbiased estimator of $\alpha(N_1-1,N_2)$. Denoting the expected value of P_L by $\tau(N_1,N_2)$, we have that,

$$\begin{aligned} EP_L &= \tau(N_1,N_2) \\ &= P(C(X_1,\bar{X}_{1(1)},\bar{X}_2,S_{(1)}) \geq 0) \\ &= \alpha(N_1-1,N_2). \end{aligned} \tag{3.63}$$

Note that P_L is defined without any assumptions on the distribution of the observations and, thus, can be used when these distributions are unknown. However, when the populations are normal, P_L can be improved.

Since $\alpha(N_1,N_2)$ is a function of μ_1, μ_2 and Σ, and $(\bar{X}_1,\bar{X}_2,S)$ is a complete sufficient statistic for (μ_1,μ_2,Σ), the estimator, defined by

$$\begin{aligned} P_L^* &= E(P_L|\bar{X}_1,\bar{X}_2,S) \\ &= E(K_1|\bar{X}_1,\bar{X}_2,S) \end{aligned} \tag{3.64}$$

is the UMVUE for τ. Clearly, the estimator P_L^* has smaller variance than the estimator P_L.

The problem then is to obtain this conditional expectation for given values of $\bar{X}_1$, $\bar{X}_2$ and S. From Equations (3.61) and (3.64), it follows that,

$$P_L^* = P[C(X_1,\bar{X}_{1(1)},\bar{X}_2,S_{(1)}) \geq 0|\bar{X}_1,\bar{X}_2,S] \tag{3.65}$$

We will now write $C(X_1,\bar{X}_{1(1)},\bar{X}_2,S_{(1)})$ in terms of $\bar{X}_1$, $\bar{X}_2$ and S. Letting $c = N_1(N_1-1)^{-1}$, $u_j = X_j-\bar{X}_1$, $j = 1,\ldots,N_1$, we have that,

$$\begin{aligned} \bar{X}_{1(j)} &= \frac{1}{N_1-1}\left(\sum_{i=1}^{N_1} X_i-X_j\right) = \frac{N_1\bar{X}_1-X_j}{N_1-1} \\ &= \frac{N_1}{N_1-1}\,\bar{X}_1 - \frac{1}{N_1-1}\,X_j = \bar{X}_1 - \left(\frac{1}{N_1-1}\right)u_j \end{aligned} \tag{3.66}$$

and

$$nS = \sum_{i=1}^{N_1}(X_i-\bar{X}_1)(X_i-\bar{X}_1)' + \sum_{i=1}^{N_2}(X_i-\bar{X}_2)(X_i-\bar{X}_2)'$$

$$= \sum_{i=1}^{N_1}(X_i-\bar{X}_{1(j)}-\frac{u_j}{N_1-1})(X_i-\bar{X}_{1(j)}-\frac{u_j}{N_1-1})'+\sum_{i=1}^{N_2}(X_i-\bar{X}_2)(X_i-\bar{X}_2)'$$

$$= (n-1)S_{(j)}+(X_j-\bar{X}_1+\frac{u_j}{N_1-1})(X_j-\bar{X}_1+\frac{u_j}{N_1-1})' + \frac{N_1}{(N_1-1)^2}u_ju_j'$$

$$-\frac{2}{N_1-1}u_j(X_j-\bar{X}_1+\frac{u_j}{N_1-1})'$$

$$= (n-1)S_{(j)}+u_ju_j'[(1+\frac{1}{N_1-1})^2+\frac{N_1}{(N_1-1)^2}-\frac{2N_1}{(N_1-1)^2}]$$

$$= (n-1)S_{(j)}+c(X_j-\bar{X}_1)(X_j-\bar{X}_1)',\ j = 1,\ldots,N_1 \tag{3.67}$$

Hence, $X_1-\bar{X}_{1(1)} = c(X_1-\bar{X}_1)$ and $C(X_1,\bar{X}_{1(1)},\bar{X}_2,S_{(1)}) = \frac{N_1}{N_1-1}(X_1-\bar{X}_1)'S_{(1)}^{-1}(X_1-\bar{X}_1) - \frac{N_2}{N_2+1}(X_1-\bar{X}_2)'S_{(1)}^{-1}(X_1-\bar{X}_2)$.

Using the alternative expression for L ($\eta = 1$) given in Equation (2.10), we can write $C(X_1,\bar{X}_{1(1)},\bar{X}_2,S_{(1)})$ as

$$b^*[(X_1-\bar{X}_1+\alpha_2(\bar{X}_2-\bar{X}_1))'S_{(1)}^{-1}(X_1-\bar{X}_1+\alpha_2(X_2-\bar{X}_1))-\alpha_2(\alpha_2+1)D_{11}^2],$$

where

$$\begin{aligned} d_1 &= c = N_1(N_1-1)^{-1} \\ d_2 &= c_2 = N_2(N_2+1)^{-1} \\ b^* &= d_1 - d_2 \\ \alpha_2 &= d_2/b^* \end{aligned} \tag{3.68}$$

and

$$D_{11}^2 = (\bar{X}_2-\bar{X}_1)'S_{(1)}^{-1}(\bar{X}_2-\bar{X}_1).$$

From Equation (3.65), it follows that,

$$P_L^* = P[(X_1-\bar{X}_1+\alpha_2(\bar{X}_2-\bar{X}_1))'S_{(1)}^{-1}(X_1-\bar{X}_1+\alpha_2(\bar{X}_2-\bar{X}_1)) \geq \alpha_2(\alpha_2+1)D_{11}^2 | \bar{X}_1, \bar{X}_2, S]$$

$$= P[(X_1-\bar{X}_1)'S_{(1)}^{-1}(X_1-\bar{X}_1)+2\alpha_2(X_1-\bar{X}_1)'S_{(1)}^{-1}(\bar{X}_2-\bar{X}_1) \geq \alpha_2 D_{11}^2 | \bar{X}_1, \bar{X}_2, S] \quad (3.69)$$

where the matrix $S_{(1)}^{-1}$ is obtained using the following result. If A and B are two nonsingular matrices and u_1 and u_2 are two column vectors such that $B = A + u_1u_2'$ then $B^{-1} = A^{-1}-(1+u_2'A^{-1}u_1)^{-1}A^{-1}u_1u_2'A^{-1}$. Hence, using Equation (3.67), we have

$$S_{(1)}^{-1} = (\frac{n-1}{n})S^{-1} + \frac{\frac{c(n-1)}{n^2}(S^{-1}(X_1-\bar{X}_1)(X_1-\bar{X}_1)'S^{-1})}{1 - \frac{c}{n}(X_1-\bar{X}_1)'S^{-1}(X_1-\bar{X}_1)}$$

$$= (n-1)\{(nS)^{-1} + \frac{c(nS)^{-1}(X_1-\bar{X}_1)(X_1-\bar{X}_1)'(nS)^{-1}}{1-c(X_1-\bar{X}_1)'(nS)^{-1}(X_1-\bar{X}_1)}\}. \quad (3.70)$$

We now define the Vectors V and ℓ as

$$V = (c)^{1/2}(nS)^{-1/2}(X_1-\bar{X}_1) \quad (3.71)$$

and

$$\ell = S^{-1/2}(\bar{X}_1-\bar{X}_2)/D. \quad (3.72)$$

Substituting Equations (3.70), (3.71) and (3.72) in Equation (3.69) gives

$$P_L^* = P\{(X_1-\bar{X}_1)'[(nS)^{-1}+(\frac{c}{1-V'V})(nS)^{-1}(X_1-\bar{X}_1)(X_1-\bar{X}_1)'(nS)^{-1}](X_1-\bar{X}_1)$$

$$+2\alpha_2(X_1-\bar{X}_1)'[(nS)^{-1}+(\frac{c}{1-V'V})(nS)^{-1}(X_1-\bar{X}_1)(X_1-\bar{X}_1)'(nS)^{-1}](\bar{X}_2-\bar{X}_1)$$

$$\geq\alpha_2(\bar{X}_2-\bar{X}_1)'[(nS)^{-1}+(\frac{c}{1-V'V})(nS)^{-1}(X_1-\bar{X}_1)(X_1-\bar{X}_1)'(nS)^{-1}](\bar{X}_2-\bar{X}_1)$$

$$|\bar{X}_1,\bar{X}_2,S\}.$$

After multiplication and some simplifications, this reduces to

$$P_L^*=P\{(\frac{1}{c})V'V+(\frac{1}{c})\ \frac{(V'V)^2}{1-V'V}\ -2\alpha_2(cn)^{-1/2}D\ell'V-2\alpha_2(\frac{V'V}{1-V'V})\ell'VD(cn)^{-1/2}$$

$$\geq \frac{\alpha_2 D^2}{n} + (\frac{\alpha_2 c}{1-V'V})\ \frac{(\ell'VD)^2}{nc}\Big|\ \bar{X}_1,\bar{X}_2,S\}.$$

Since $1-V'V > 0$ with probability one,

$$P_L^*=P\{\frac{V'V}{c} - 2\alpha_2 D(cn)^{-1/2}(\ell'V)\geq \frac{\alpha_2 D^2}{n}\ (1-V'V + (\ell'V)^2)|\bar{X}_1,\bar{X}_2,S]$$

$$= P\{V'V\geq\alpha_2 D^2(\frac{c}{n})[1-V'V+(\ell'V)^2]+2\alpha_2(\ell'V)D(\frac{c}{n})^{1/2}|\bar{X}_1,\bar{X}_2,S]$$

$$=P[V'V\geq \frac{\alpha_2 D^2(\frac{c}{n})+\alpha_2 D^2(\frac{c}{n})(\ell'V)^2+2\alpha_2(\ell'V)D(\frac{c}{n})^{1/2}}{1+\alpha_2 D^2(\frac{c}{n})}\ |\bar{X}_1,\bar{X}_2,S]$$

$$=P[V'V\geq \frac{D^2(\ell'V)^2+2(n/c)^{1/2}D(\ell'V)+D^2}{nk^*+D^2}\ |\bar{X}_1,\bar{X}_2,S],$$

where $k^* = \frac{1}{\alpha_2 c} = \frac{1}{N_1} + \frac{1}{N_2}$.

Noting that both ℓ and D are fixed when $\bar{X}_1$, $\bar{X}_2$ and S are fixed, we can rewrite the above equation as

$$P_L^* = P[V'V \geq \frac{\dot{D}^2(\dot{\ell}'V)^2 + 2(n/c)^{1/2}\dot{D}(\dot{\ell}'V) + \dot{D}^2}{nk^* + \dot{D}^2} \mid \bar{X}_1, \bar{X}_2, S],$$

where a period above a variable indicates that the variable is fixed. Since Broffitt and Williams (1973) have shown that V is stochastically independent of $(\bar{X}_1, \bar{X}_2, S)$, it follows that

$$P_L^* = P[V'V \geq \frac{\dot{D}^2(\ell'V)^2 + 2(n/c)^{1/2}\dot{D}(\ell'V) + \dot{D}^2}{nk^* + \dot{D}^2}]$$

$$= P[y \geq \frac{\dot{D}^2 w^2 + 2(n/c)^{1/2}\dot{D}w + \dot{D}^2}{nk^* + \dot{D}^2}] \qquad (3.73)$$

where $w = \ell'V$ and $y = V'V$.

The joint density function of ω and y and the density function of w were obtained by Broffitt and Williams (1971). They are given as

$$f(w,y) = \frac{(1-y)^{1/2(n-p)-1}(y-w^2)^{1/2(p-1)-1}}{\beta(\frac{1}{2}, \frac{p-1}{2})\ \beta(\frac{p}{2}, \frac{1}{2}(n-p))};\ -1<w<1,\ w^2<y<1 \qquad (3.74)$$

and

$$f(w) = \frac{(1-w^2)^{1/2(n-1)-1}}{\beta(\frac{1}{2}, \frac{1}{2}(n-1))};\ -1<w<1 \qquad (3.75)$$

When $p \geq 2$ and under the restrictions $n \geq p + 1$, $N_1 \geq 2$ and $N_2 \geq 1$, the estimator P_L^* is obtained from (3.74). When $p = 1$, $y = w^2$ and P_L^* is obtained from (3.75). Under the same restrictions on N_1, N_2 and n, we have

$$P_L^* = 1-P\{(\frac{c}{n})^{1/2}(\frac{1-(1+k^*c)^{1/2}}{k^*c})\dot{D}\underline{<}w\underline{<}(\frac{c}{n})^{1/2}(\frac{1+(1+k^*c)^{1/2}}{k^*c})\dot{D}\} \qquad (3.76)$$

where the limits on w are the roots of w in the quadratic equation

$$w^2 = \frac{\dot{D}^2w^2+2(n/c)^{1/2}\dot{D}w+\dot{D}^2}{nk^*+\dot{D}^2}$$

The estimator P_L^* depends on $\bar{X}_1, \bar{X}_2$ and S only through the value of D^2; to emphasize this dependence, we let $P_L^* = P_L^*(D^2)$. Notice that $P_L^*(D^2)$ is a continuous function of D^2 with maximum value of one, when $D^2 = 0$ and approaching zero as D^2 goes to infinity.

In order to obtain the distribution function and moments of P_L^*, we will first show that $P_L^*(D^2)$ is a strictly decreasing function of D^2. Then we will apply the following result given by Broffitt and Williams (1973).

Let $\psi(D^2)$ be a positive and strictly decreasing function of D^2 for $0 \leq D_o^2 < D^2 < D_1^2 \leq \infty$, where D_o^2 and D_1^2 are constants, $\psi(D^2) = \psi(D_o^2)$ for $0 \leq D^2 \leq D_o^2$ and $\psi(D^2) = 0$ for $D^2 \geq D_1^2$. Define

$$h(t) = \begin{cases} \psi^{-1}(t) & 0 < t < \psi(D_o^2) \\ D_1^2 & t = 0 \end{cases}$$

and

$$\gamma(t) = \begin{cases} \dfrac{h(t)}{nm+h(t)} & 0 < t < \psi(D_o^2) \\ \dfrac{D_1^2}{nm+D_1^2} & t = 0 \text{ and } D_1^2 \neq \infty \\ 1 & t = 0 \text{ and } D_1^2 = \infty \end{cases}$$

where $m = (N_1 + N_2)(N_1N_2)^{-1}$. Then

$$P\{\psi(D^2) \leq t\} = \begin{cases} 1 & \psi(D_o^2) \leq t \\ 1 - \sum_{i=0}^{\infty} \dfrac{\lambda^i e^{-\lambda}}{i!} \dfrac{\beta(\frac{2i+p}{2}, \frac{n-p+1}{2}; \gamma(t))}{\beta(\frac{2i+p}{2}, \frac{n-p+1}{2})}, & 0 \leq t \leq \psi(D_o^2) \\ 0 & t < 0 \end{cases} \quad (3.77)$$

where $\lambda = \frac{1}{2}m^{-1}\delta^2$ and

$$E[\psi(D^2))^r] = r \sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \int_0^{\psi(D_o^2)} t^{r-1} \frac{\beta(\frac{2i+p}{2}, \frac{n-p+1}{2}; \gamma(t))}{\beta(\frac{2i+p}{2}, \frac{n-p+1}{2})} dt \quad (3.78)$$

We will now show that $P_L^*(D^2)$ is a strictly decreasing function in D^2. From Broffitt and Williams (1973), it is known that the function $T_2^*(D^2)$, given by

$$T_2^*(D^2) = P[y \geq \frac{\dot{D}^2 w^2 + 2(n/c)^{1/2}\dot{D}w + \dot{D}^2}{nk + \dot{D}^2}], \quad (3.79)$$

with $k = \frac{1}{N_1-1} + \frac{1}{N_1}$, is strictly decreasing in D^2. This implies that the function

$$H(D^2) = \frac{D^2w^2 + 2(n/c)^{1/2}Dw + D^2}{nk + D^2} \tag{3.80}$$

is strictly increasing in D^2. Since y is positive with probability one, we can assume without loss of generality that $H(D^2)$ is positive. Also, without loss of generality, we can assume that $N_2 < N_1 - 1$ and hence $k^* = \frac{1}{N_2} + \frac{1}{N_1}$ is greater than k; it follows that the functions $\frac{nk+D^2}{nk^*+D^2}$ and

$$H^*(D^2) = H(D^2)\left(\frac{nk+D^2}{nk^*+D^2}\right) \tag{3.81}$$

are both strictly increasing in D^2. It follows that, for each value D_a^2 of D^2, there is a corresponding value D_a^{2*} such that,

$$H^*(D_a^{2*}) = H(D_a^2). \tag{3.82}$$

If we now write $P_L^*(D^2)$ as

$$P_L^*(D^2) = P[y \geq \frac{\dot{D}^2w^2 + 2(n/c)^{1/2}\dot{D}w + \dot{D}^2}{nk + \dot{D}^2}\left(\frac{nk + \dot{D}^2}{nk^* + \dot{D}^2}\right)], \tag{3.83}$$

and consider two values D_a^2 and D_b^2 of D^2, such that $D_a^2 < D_b^2$, we can establish, using Equations (3.79) to (3.83), that,

$$P[y \geq H^*(D_b^{2*})] = P[y \geq H(D_b^2)] \tag{3.84}$$

and

$$P[y \geq H^*(D_a^{2*})] = P[y \geq H(D_a^2)] \tag{3.85}$$

where D_a^{2*} and D_b^{2*} are the corresponding values of D_a^2 and D_b^2, defined in (3.82).

Since $T_2^*(D^2)$ is strictly decreasing in D^2, we have that, $T_2^*(D_a^2) = P[y \geq H(D_a^2)] > P[y \geq H(D_b^2)] = T_2^*(D_b^2)$. Equations (3.84) and (3.85) imply that $P[y \geq H^*(D_a^{2*})] > P[y \geq H^*(D_b^{2*})]$. Hence $H^*(D_a^{2*}) < H^*(D_b^{2*})$ and since $H^*(D^2)$ is strictly increasing in D^2, it follows that $D_a^{2*} < D_b^{2*}$. Thus, we can conclude that, the function $P_L^*(D^2)$ is strictly decreasing function of D^2.

The distribution function and moments of P_L^* can now be obtained from Equations (3.77) and (3.78), respectively, by setting $\psi(D^2) = P_L^*(D^2)$, with $D_o^2 = 0$, $D_1^2 = \infty$ when $p \geq 2$ and $D_1^2 = \frac{n}{c}\left(\frac{k^*c}{|1-(1+k^*c)^{1/2}|}\right)^2$ when $p = 1$.

The expectation of $P_L^*(D^2)$ as a function of N_1 and N_2 is of special interest since it can be used to obtain an exact expression for the unconditional probabilities of misclassification $P_1^*(Z)$, i.e., $EP_L^*(D^2) = \tau(N_1,N_2)$ and $\tau(N_1 + 1, N_2) = \alpha(N_1,N_2) = P_1^*(Z)$.

Letting $\theta = ((N_1 + 1)N_2/(2(N_1 + N_2 + 1)))\delta^2$, then

$$P_1^*(Z) = \alpha(N_1,N_2) = \sum_{i=0}^{\infty} \frac{\theta^i e^{-\theta}}{i!} \int_0^1 \frac{\beta(\frac{2i+p}{2}, \frac{n-p+2}{2}; \tilde{\gamma}(t))}{\beta(\frac{2i+p}{2}, \frac{n-p+2}{2})} \, dt, \qquad (3.86)$$

where

$$\tilde{\gamma}(t) = \frac{\psi^{-1}(t)}{(n+3)[(N_1+1)N_2]^{-1} + \psi^{-1}(t)}, \quad 0 < t < 1$$

and $\psi(D^2) = P^*_{L(N_1+1,N_2)}(D^2)$ can be obtained from Equations (3.73) or (3.76) with all values of N_1 replaced by N_1+1.

From (3.86) we observe that although the estimator P_L^* was designed as an estimator for $\tau(N_1,N_2)$, it can also be viewed as an estimator of $P_1^*(Z)$.

The expression for $P_2^*(Z)$ can now be obtained by interchanging N_1 and N_2 in the expression for $P_1^*(Z)$.

IV. COMPARISONS AND CONCLUSIONS

A. Comparisons with Respect to Asymptotic Bias

In the previous chapters, several estimators of the unconditional probability of misclassification were given. In this chapter, we will examine the performance of these estimators with respect to their asymptotic bias and asymptotic mean square error. Let us first study the bias.

In practice, the consideration of bias is important. For example, in medical applications, it is highly desirable to have accurate values of P_1^* and P_2^*; estimators which underestimate these probabilities will indicate that the classification procedure is much better than what it actually is. In such a case, serious misclassification may result.

From the expressions of asymptotic bias of the estimators considered in Chapter III, we obtain the relationships of the asymptotic biases as follows.

$$\text{Bias } (Q_D) = A_1 - B_1 + O_2 \tag{4.1}$$

$$\text{Bias } (Q_{DS}) = \text{Bias } (Q_D) - q_{DS}(\delta) \tag{4.2}$$

$$\text{Bias } (Q_{MO}) = \text{Bias } (Q_D) - A_1 \tag{4.3}$$

$$\text{Bias } (Q_{MO}) = \text{Bias } (Q_{DS}) + q_{DS}(\delta) - A_1 \tag{4.4}$$

$$\text{Bias } (Q_{MOS}) = \text{Bias } (Q_{MO}) - q_{DS}(\delta) \tag{4.5}$$

$$\text{Bias } (Q_{MOS}) = \text{Bias } (Q_D) - [A_1 + q_{DS}(\delta)] \tag{4.6}$$

$$\text{Bias } (Q_{MOS}) = \text{Bias } (Q_{DS}) - A_1 \tag{4.7}$$

and

$$\text{Bias } (Q_{MC}) = \text{zero} + O_2 \tag{4.8}$$

where $q_{DS}(\delta)$, A_1 and B_1 are given, respectively, as

$$q_{DS}(\delta) = (\frac{p+1}{4n})\delta\phi(-\frac{\delta}{2}) \tag{4.9}$$

$$A_1 = \frac{1}{16}\phi(-\frac{\delta}{2})[\frac{1}{N_1}(3\delta+4(p-1)\delta^{-1})+\frac{1}{N_2}(-\delta+4(p-1)\delta^{-1})+\frac{4}{n}\delta(p-1)] \tag{4.10}$$

and

$$B_1 = \frac{1}{16}\phi(-\frac{\delta}{2})[\frac{1}{N_1}(\delta-4(p-1)\delta^{-1})+\frac{1}{N_2}(\delta-4(p-1)\delta^{-1})+\frac{\delta}{2n}(\delta^2-4(2p+1))] \tag{4.11}$$

From Equations (4.1) to (4.8) it follows that, except for Q_{MC}, all the other estimators have first order bias with respect to N_1^{-1}, N_2^{-1} and n^{-1}. Hence, of the five estimators for the unconditional probability of misclassification, the estimator Q_{MC} is best when they are compared with respect to their asymptotic baises.

From Equations (4.1) to (4.7), we observe that in order to compare the asymptotic bias of the estimators Q_D, Q_{DS}, Q_{MO} and Q_{MOS}, we have to examine the difference A_1-B_1. However, from this difference, it is difficult to obtain general conclusions about the relative superiority of the estimators. Hence, we must introduce some restrictions on the values of N_1, N_2, p and δ^2.

Assuming that $N_1 \leq N_2$, we can show that $A_1-B_1 \geq 0$ for all p and δ^2 values such that $\delta^2 \leq 4(4p-1)$. Considering that in most practical situations the Mahalanobis distance between the two populations is such that $1 \leq \delta^2 \leq 10$, the restriction $\delta^2 < 4(4p-1)$ will usually be satisfied.

It follows that, among the estimators for $P_2^*(Z)$ considered here, Q_D is the worst in most practical situations when the comparison is made with respect to asymptotic bias. However, the estimator Q_D may be the best when $\delta^2 \geq 4(4p-1)$ and for some values of the sample sizes.

From equations (4.5) and (4.7) and under the above restrictions on N_1, N_2, p and δ^2, it follows that, the estimator Q_{MOS} is consistently the second best among the five estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC}. The bias of Q_{MOS} is only larger than that of Q_{MC}, which was explicitly constructed to have second order bias. However, we should note that when $\delta^2 > 4(4p-1)$ or δ^2 large, the estimator Q_{MOS} may be worst than other estimators of $P_2^*(Z)$.

From (4.4) we observe that, the relative performance of the estimators Q_{DS} and Q_{MO} with respect to asymptotic bias, depends on the sign of the quantity $A_1 - q_{DS}(\delta)$. From (4.9) and (4.10), we have that,

$$A_1-q_{DS}(\delta) = \frac{1}{16}\phi(-\frac{\delta}{2})[\frac{1}{N_1}(3\delta+4(p-1)\delta^{-1})+ \frac{1}{N_2}(-\delta+4(p-1)\delta^{-1})- \frac{8\delta}{n}] \qquad (4.12)$$

which will be positive, except when δ^2 is large, p is small and N_1 is much larger than N_2. Hence, under most practical situations, the estimator Q_{MO} will be better than the estimator Q_{DS}.

B. Comparisons with Respect to Asymptotic Mean Square Error

In this section, we will assess the relative superiority of the estimators of the unconditional probability of misclassification, $P_2^*(Z)$, with respect to Asymptotic Mean Square Error.

The expressions for the AMSE of the estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} were derived in Chapter III, Section D. Since these expressions are rather complicated, the comparison of these estimators with respect to their AMSE, will be quite difficult.

An examination of the expressions of the AMSE of the estimators of $P_2^*(Z)$ considered here, reveals that each estimator has the same leading term of the first order $[\frac{1}{2}\phi(-\frac{\delta}{2})]^2 \{\frac{1}{N_1} + \frac{1}{N_2} + \frac{\delta^2}{2n}\}$ in the AMSE. This indicates that all five estimators of $P_2^*(Z)$ are equivalent when they are compared with respect to AMSE and up to the first order terms. However, the second order terms in the AMSEs are different, and by a comparison of these second order terms, the relative performance of the estimators will be determined. Also, since the term $m_1(Q_D)$ appears in all the AMSE expressions, it is not required in the evaluation of the relative superiority of the estimators.

In comparing the estimators with respect to their asymptotic mean square errors, we will first examine the relative performance of the estimators of $P_2^*(Z)$, which have first order bias, i.e., Q_D, Q_{DS}, Q_{MO} and Q_{MOS}. We will try to obtain, under certain conditions, the best estimator in this class and then compare its AMSE against that of the estimator Q_{MC}.

From Equation (3.35), we have that,

$$\mathrm{AMSE}(Q_{DS})-\mathrm{AMSE}(Q_D) = [q_{DS}(\delta)]^2 - 2A_1 q_{DS}(\delta) + T_{DS}$$

$$= (\frac{p+1}{4n})^2\delta^2[\phi(-\frac{\delta}{2})]^2-(\frac{2(p+1)}{4n})\delta[\phi(-\frac{\delta}{2})]^2$$

$$\cdot\ \frac{1}{16}[\frac{1}{N_1}(3\delta+4(p-1)\delta^{-1})+\frac{1}{N_2}(-\delta+4(p-1)\delta^{-1})+\frac{4\delta(p-1)}{n}]$$

$$+(\frac{p+1}{32n})[\phi(-\frac{\delta}{2})]^2[(\frac{1}{N_1}+\frac{1}{N_2})(3\delta^2-4(p+1))+\frac{\delta^2}{2n}(3\delta^2-12-8p)]$$

$$=(\frac{p+1}{256n})[\phi(-\frac{\delta}{2})]^2\{-8\delta[\frac{1}{N_1}(3\delta+4(p-1)\delta^{-1})+\frac{1}{N_2}(-\delta+4(p-1))$$

$$+\frac{4\delta(p-1)}{n}]$$

$$+8[(\frac{1}{N_1}+\frac{1}{N_2})(3\delta^2-4(p+1))+\frac{\delta^2}{2n}(3\delta^2-12-8p)]$$

$$+\frac{16(p+1)\delta^2}{n}\}$$

After multiplying and rearranging terms, we have,

$$\mathrm{AMSE}(Q_{DS})-\mathrm{AMSE}(Q_D)=(\frac{p+1}{256n})[\phi(-\frac{\delta}{2})]^2\{\frac{8}{N_1}[(3\delta^2-4(p+1))-\delta(3\ +4(p-1)\delta^{-1})]$$

$$+\frac{8}{N_2}[(3\delta^2-4(p+1))-\delta(-\delta+4(p-1)\delta^{-1})]$$

$$+\frac{8}{n}[\frac{\delta^2}{2}\ (3\delta^2-12-8p)+2\delta^2(p+1)-4\delta^2(p-1)]\}$$

$$=(\frac{p+1}{256n})[\phi(-\frac{\delta}{2})]^2\{\frac{-64p}{N_1}+\frac{32}{N_2}(\delta^2-2p)+\frac{12}{n}(\delta^2-4p)\} + O_3 \quad (4.13)$$

From this expression, it follows that, for δ^2 and p values such that $\delta^2 \leq 2p$, $AMSE(Q_{DS}) \leq AMSE(Q_D)$ for any values of N_1 and N_2. Under certain restrictions of the sample sizes, the range of values of δ^2, for which (4.13) remains positive, can be increased. For example, for sample sizes N_1 and N_2 such that $N_1 \leq 2N_2$, the difference $AMSE(Q_{DS}) - AMSE(Q_D)$ will be negative for all values of p and δ^2 such that $\delta^2 \leq 3p$.

Moreover, for fixed values of δ^2, the relative superiority of the estimator Q_{DS} over the estimator Q_D increases as p increases. Also, for fixed values of p, $AMSE(Q_{DS}) - AMSE(Q_D)$ becomes positive when δ^2 is much larger than p. This indicates that the estimator Q_D will be better than the estimator Q_{DS} only in the most unusual situations, i.e., $\delta^2 > 10$, $p \leq 5$ and $\delta^2 > 2p$, $p \geq 6$.

The performance of the estimators Q_{MO} and Q_{MOS} can be compared in a similar manner. Using Equation (3.52), we have that,

$$\begin{aligned}
AMSE(Q_{MOS})-AMSE(Q_{MO}) &= [q_{DS}(\delta)]^2 + T_{DS} \\
&= (p+1)[\phi(-\tfrac{\delta}{2})]^2\{\frac{(p+1)\delta^2}{16n^2} + \\
&\quad \frac{1}{32N}[(\frac{1}{N_1}+\frac{1}{N_2})(3\delta^2-4(p+1))+\frac{\delta^2}{2n}(3\delta^2-12-8p)]\} \\
&= \frac{(p+1)[\phi(-\frac{\delta}{2})]^2}{32n} \cdot \\
&\quad \{(\frac{1}{N_1}+\frac{1}{N_2})(3\delta^2-4(p+1))+\frac{\delta^2}{2n}(3\delta^2-4(p+2))\}+0_3. \qquad (4.14)
\end{aligned}$$

It follows that,

$$\mathrm{AMSE}(Q_{MOS})-\mathrm{AMSE}(Q_{MO}) \leq \left(\frac{p+1}{32n}\right)[\phi(-\frac{\delta}{2})]^2\{(\frac{1}{N_1}+\frac{1}{N_2}+\frac{\delta^2}{2n})(3\delta^2-4(p+1))\}+O_3 \quad (4.15)$$

When p and δ^2 are such that $\delta^2 \leq \frac{4}{3}(p+1)$, the above difference is negative for any N_1 and N_2. Hence, the estimator Q_{MOS} will be better than the estimator Q_{MO} when $\delta^2 \leq \frac{4}{3}(p+1)$. Moreover, from Equation (4.14), we have that for fixed δ^2 and increasing p, $\mathrm{AMSE}(Q_{MOS})$ becomes increasingly smaller in relation to $\mathrm{AMSE}(Q_{MO})$.

Noting that $\frac{4}{3}(p+1) \leq 2p$ for $p \geq 2$ and from the previous remarks, we have that whenever the estimator Q_{MOS} is better than the estimator Q_{MO}, the estimator Q_{DS} will be better than the estimator Q_D.

In order to simplify the analysis of the remaining differences in AMSE of any two estimators, it is necessary to introduce some additional restrictions on the values of N_1, N_2, p and δ^2. First, we will only consider δ^2 values in the interval [1,10] since these are the values for the Mahalanobis distance that are more frequently found in practice. And second, we will assume that the sample sizes are such that $N_1 \leq 3N_2$.

Under these general restrictions and from Equations (3.33) and (3.46), we have that the difference $\mathrm{AMSE}(Q_D) - \mathrm{AMSE}(Q_{MO}) = A_1^2 - T_{MO}$. From Equations (3.47) to (3.50) we can observe that it is very difficult to obtain general conclusions about the sign of this difference. However, when δ^2 is small relative to p, T_{MO} is negative and hence $\mathrm{AMSE}(Q_D) > \mathrm{AMSE}(Q_{MO})$. Since the estimator Q_{MOS} is better than the estimator Q_{MO} when $\delta^2 < \frac{4}{3}(p+1)$, it is expected that the estimator Q_{MOS}

will be better than the estimator Q_D when δ^2 is small relative to p.

We can conclude that among the estimators Q_D, Q_{DS}, Q_{MO} and Q_{MOS} for $P_2^*(Z)$, the estimator Q_D is, in most practical situations, consistently inferior than any of the others. The estimator Q_D will be better than the others when δ^2 is large compared with p.

Since we have already shown that the estimator Q_{MOS} is better than the estimator Q_{MO} when $\delta^2 \leq \frac{4}{3}(p+1)$, it remains to compare the performance of the estimators Q_{MOS} and Q_{DS}. However, the difference in their AMSE, given by,

$$\text{AMSE}(Q_{DS}) - \text{AMSE}(Q_{MOS}) = [A_1 - q_{DS}(\delta)]^2 + T_{DS} - T_{MO} + O_3 \tag{4.16}$$

is very difficult to analyze. In view of this, a numerical evaluation of (4.16) was done for several values of N_1, N_2, p and δ^2. The results are given in Table 4.1.

From Table 4.1, we observe that the difference in AMSE, $\text{AMSE}(Q_{DS})-\text{AMSE}(Q_{MOS})$ is positive for $\delta^2 < 4$ and all p values and is negative when $\delta^2 \geq 4$ and p is small. We also observe that the range of values of p for which $\text{AMSE}(Q_{DS})-\text{AMSE}(Q_{MOS})$ is negative, becomes wider as δ^2 increases. For fixed δ^2 and $p \geq 2$, this difference in AMSE increases as p increases. This indicates that for those values of δ^2 for which Q_{MOS} is better than Q_{DS}, the relative superiority of Q_{MOS} over Q_{DS} increases as p increases. On the other hand, for those values of δ^2 for which Q_{DS} is better than Q_{MOS}, the relative superiority of Q_{DS} over Q_{MOS} decreases with p. Finally, we observe that for fixed $p \geq 2$ and $1 \leq \delta^2 \leq 11$, $\text{AMSE}(Q_{DS})-\text{AMSE}(Q_{MOS})$ decreases as δ^2 increases.

Table 4.1. Values of $AMSE(Q_{DS}) - AMSE(Q_{MOS})$

N_1	N_2	δ^2	p = 1	p = 2	p = 3	p = 5	p = 8	p = 10
8	16	1	0.000337861	0.000367497	0.002591451	0.013622310	0.046625960	0.079599910
		1.69	0.000172481	0.000360067	0.002080837	0.010121940	0.033682430	0.057055380
		2.25	0.000051890	0.000235228	0.001662773	0.008250505	0.027453530	0.046493500
		4	-0.000229036	-0.000183582	0.000618235	0.004490957	0.015972770	0.027409110
		9	-0.000419198	-0.000582726	-0.000506962	0.000362452	0.003461274	0.006723624
		11	-0.000372519	-0.000535774	-0.000545519	-0.000104480	0.001708396	0.003684542
10	30	1	0.000209981	0.000249760	0.001344055	0.006696206	0.022633340	0.038530660
		1.69	0.000114005	0.000241476	0.001088439	0.004940838	0.016115660	0.027163020
		2.25	0.000044445	0.000174977	0.000880298	0.004015297	0.013028700	0.021911590
		4	-0.000116415	-0.000045233	0.000362941	0.002190258	0.007458668	0.012655880
		9	-0.000225657	-0.000263909	-0.000201242	0.000226859	0.001625923	0.003063235
		11	-0.000200558	-0.000245967	-0.000227473	0.000001226	0.000823553	0.001691288

Although no general conclusions can be derived from Table 4.1, about the relative superiority of the estimator Q_{MOS} over the estimator Q_{DS}, the results of the numerical evaluation of the difference $AMSE(Q_{DS})-AMSE(Q_{MOS})$ hint that the estimator Q_{MOS} is a better estimator than the estimator Q_{DS} when δ^2 is small compared with p. The estimator Q_{DS} is better than the estimator Q_{MOS} when δ^2 is large in relation to p.

The AMSE of the estimator Q_{MC} for the unconditional probability of misclassification, $P_2^*(Z)$, will be compared to the AMSE of the estimators Q_{MOS} and Q_{DS}. The differences in AMSE involved in these comparisons are given by

$$AMSE(Q_{MOS})-AMSE(Q_{MC}) = T_{DS} + [q_{DS}(\delta)]^2 + R - B_1^2 + O_3 \tag{4.17}$$

and

$$AMSE(Q_{DS})-AMSE(Q_{MC})=(A_1-q_{DS}(\delta))^2+T_{DS}-T_{MO}+R-B_1^2+O_3 \tag{4.18}$$

respectively.

From these expressions, no direct conclusion about the relative superiority of these estimators can be obtained. Hence, a numerical evaluation of (4.17) and (4.18) was done for several values of N_1, N_2, p and δ^2. The results are given in Tables 4.2 and 4.3.

From Table 4.2, we find that the difference in AMSE, $AMSE(Q_{MOS})-AMSE(Q_{MC})$ is positive and increasing with p for all $p \geq 8$ and δ^2 fixed. For $p < 8$, this difference decreases when p increases and is negative for most δ^2 values. We also observe that, when Q_{MC} is inferior to Q_{MOS} on the AMSE criterion, the difference in their AMSE is small.

Table 4.2. Values of AMSE(Q_{MOS}) - AMSE(Q_{MC})

N_1	N_2	δ^2						
8	16	1	-0.000397945	-0.001871120	-0.002799590	-0.003022425	0.000728577	0.005952779
		1.69	-0.000317114	-0.001230066	-0.001642357	-0.000964937	0.003806137	0.009490207
		2.25	-0.000251867	-0.000973922	-0.001245134	-0.000435027	0.004161440	0.009479981
		4	-0.000058507	-0.000537045	-0.000697858	-0.000066304	0.003263962	0.007072777
		9	0.000270317	0.000027012	-0.000103855	-0.000028276	0.000928375	0.002128326
		11	0.000302808	0.000119028	0.000008366	0.000006398	0.000551835	0.001281053
10	30	1	-0.000180148	-0.000901831	-0.001348069	-0.001414209	0.000552405	0.003240693
		1.69	-0.000137332	-0.000582555	-0.000780870	-0.000436770	0.001931183	0.004744366
		2.25	-0.000104203	-0.000455569	-0.000588033	-0.000196255	0.002033174	0.004613984
		4	-0.000012705	-0.000242967	-0.000324286	-0.000040097	0.001503260	0.003276873
		9	0.000122075	0.000010167	-0.000051600	-0.000024718	0.000391646	0.000919922
		11	0.000131416	0.000048047	-0.000003067	-0.000008531	0.000225182	0.000542264

Table 4.3. Values of AMSE(Q_{DS}) - AMSE(Q_{MC})

N_1	N_2	δ^2	p = 1	p = 2	p = 3	p = 5	p = 8	p = 10
8	16	1	-0.000060085	-0.001503623	-0.000208138	0.010599880	0.047354540	0.085552690
		1.69	-0.000144634	-0.000869999	0.000438480	0.009157013	0.037488570	0.066545540
		2.25	-0.000199977	-0.000738694	0.000417639	0.007815477	0.031625070	0.055973480
		4	-0.000287542	-0.000720627	-0.000079623	0.004424654	0.019236730	0.034481890
		9	-0.000148882	-0.000555715	-0.000610817	0.000334175	0.004389647	0.008851953
		11	-0.000069711	-0.000416746	-0.000537153	-0.000098082	0.002260230	0.004965592
10	30	1	0.000029834	-0.000652072	-0.000003014	0.005281996	0.023185740	0.041771350
		1.69	-0.000023327	-0.000341078	0.000307570	0.004504070	0.018046840	0.031907390
		2.25	-0.000059748	-0.000280592	0.000292265	0.003819042	0.015061870	0.026525570
		4	-0.000129120	-0.000288199	0.000038655	0.002150161	0.008961931	0.015932750
		9	-0.000103582	-0.000253742	-0.000252842	0.000202140	0.002017569	0.003983155
		11	-0.000069142	-0.000197920	-0.000230540	-0.000007305	0.001048735	0.002233552

However, when Q_{MC} is superior to Q_{MOS}, the difference in their AMSE is not necessarily as small.

Although no general conclusions can be given from Table 4.2, the results of this numerical evaluation indicate that the estimator Q_{MOS} is better than the estimator Q_{MC} when $p < 8$. Also, the estimator Q_{MC} is better than the estimator Q_{MOS} when $p \geq 8$.

A comparison of the estimator Q_{DS} and Q_{MC} can be done by examining the results in Table 4.3. From this table we find that, the difference $AMSE(Q_{DS})-AMSE(Q_{MC})$ is negative in almost all cases for which $p \leq 3$. This indicates that the estimator Q_{DS} is better than the estimator Q_{MC} when $p \leq 3$. However, from Table 4.1, we have found that Q_{DS} is inferior to Q_{MOS} when δ^2 is small relative to p. Also, from Equation (4.13), we have that for $\delta^2 \geq 2p$, the estimator Q_D is better than Q_{DS}.

From the results of the previous comparisons, we have found that among the estimators Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} of the unconditional probability of misclassification, $P_2^*(Z)$, there is no estimator which is uniformly better than the others for all values of N_1, N_2, p and δ^2. However, when $1 \leq \delta^2 \leq 10$, $p < 8$ and δ^2 is small relative to p, i.e., $\delta^2 \leq 2p$, the estimator Q_{MOS} is the best and the estimator Q_D is the worst. For $p < 8$ and δ^2 very large with respect to p, the estimator Q_D will be better than the others. For $p > 8$, the numerical evaluation of the difference in the AMSE of the estimators Q_{MC} and Q_{MOS} indicates that the estimator Q_{MC} is better than the estimator Q_{MOS}.

C. A Monte Carlo Study

We will consider two additional estimators of the unconditional probability of misclassification, $P_2^*(Z)$. These two estimators do not require any assumptions on the distribution function of the populations.

The first estimator was originally proposed by C.A.B. Smith (1947). It is defined as,

$$Q_R = \frac{1}{N_2} \sum_{i=1}^{N_2} R_{i1}, \tag{4.19}$$

where

$$R_{i1} = \begin{cases} 1 & \text{if } C(X_i, \bar{X}_1, \bar{X}_2, S) \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

and

$$C(X_i, \bar{X}_1, \bar{X}_2, S) = \frac{N_1}{N_1+1}(X_i - \bar{X}_1)' S^{-1} (X_i - \bar{X}_1) - \frac{N_2}{N_2+1}(X_i - \bar{X}_2)' S^{-1} (X_i - \bar{X}_2),$$

$$i = 1, \ldots, N_2. \tag{4.21}$$

The other estimator is the Lachenbruch (1967) jackknife estimator, defined as

$$Q_L = \frac{1}{N_2} \sum_{i=1}^{N_2} R_{i2},$$

where

$$R_{i2} = \begin{cases} 1 & \text{if } C(X_i, \bar{X}_1, \bar{X}_{2(i)}, S_{(i)}) \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.22}$$

and a statistic with the subscript (i) indicates that the observation X_i has been removed from the calculation of that statistic.

The estimator Q_L is obtained by successively omitting an observation from the second sample and calculating the classification statistic based on the remaining $N_1 + (N_2 - 1)$ observations. This modified classification statistic is then used to classify the observation that was left out. For each one of the N_2 classifications, the omitted observation and the corresponding classification statistic are independent. Following the procedure used in Chapter III, Section E, we can show that,

$$X_i - \bar{X}_{2(i)} = \frac{N_2}{N_2-1} (X_i-\bar{X}_2), \tag{4.23}$$

$$nS = (n-1)S_{(i)} + \frac{N_2}{N_2-1} (X_i-\bar{X}_2)(X_i-\bar{X}_2)' \tag{4.24}$$

and

$$S_{(i)}^{-1} = (n-1)\{(nS)^{-1} + \frac{\frac{N_2}{N_2-1}(nS)^{-1}(X_i-\bar{X}_2)(X_i-\bar{X}_2)'(nS)^{-1}}{1-\frac{N_2}{N_2-1}(X_i-\bar{X}_2)'(nS)^{-1}(X_i-\bar{X}_2)}\} \tag{4.25}$$

for $i = 1,\ldots,N_2$.

Using these equations, the classification statistic $C(X_i,\bar{X}_1,\bar{X}_{2(i)},S_{(i)})$ can be written in terms of the original variables $\bar{X}_1$, $\bar{X}_2$ and S^{-1}, requiring the inversion of only one matrix.

In this section, the performance of the estimators Q_R, Q_L, Q_D, Q_{DS}, Q_{MO}, Q_{MOS} and Q_{MC} will be evaluated using a simulation study and assuming that the population are normally distributed. Since the Z classification statistic is invariant under linear transformation on the observation X, we can assume, without loss of generality, that $\Pi_1 \sim N(\mu_1, I)$ and $\Pi_2 \sim N(\mu_2, I)$, respectively, where $\mu_1 = 0$ and $\mu_2 = (\delta, 0,\ldots,0)$.

A total of 720 random samples was taken from two normal populations Π_1 and Π_2. Several combinations of N_1, N_2, p and δ^2 were used. They are given in Tables 4.4 and 4.5. Forty repetitions of the experiment were done for each combination of N_1, N_2, p and δ^2. Letting t denote any of the suffixes R, L, D, DS, MO, MOS or MC, the quantities $Q_t, |P_2^*(Z) - Q_t|$ and $(P_2^*(Z) - Q_t)^2$ were calculated. The value of $P_2^*(Z)$ was approximated using up to the second order terms in the Memon-Okamoto (1971) asymptotic expansion. Then, the average of the measurements $|P_2^*(Z) - Q_t|$ and $(P_2^*(Z) - Q_t)^2$ were obtained for each estimator. These averages were used to compare the relative performance of the estimator of the unconditional probability of misclassification. Since the conclusions obtained based on the mean of the absolute deviation are identical to the conclusions obtained based on the mean of the square of the deviations, only the results based on $|P_2^*(Z) - Q_t|$ are given in Tables 4.4 and 4.5.

Several observations can be made in these tables. First, we observe that there is no estimator which is uniformly better than any of the others for all values of N_1, N_2, p and δ^2. Second, when the observations that are being classified come from normal populations, the estimators based on normality are consistently better than the estimators that are not based on this assumption. Third, among the two estimators,

Table 4.4. Average values of $|P_2^*(Z) - Q_t|$ when p = 1

δ^2	N_1	N_2	Q_D	Q_{DS}	Q_{MO}	Q_{MOS}	Q_{MC}	Q_R	Q_L
1	8	16	0.06299	0.05925	0.05978	0.05639	0.05969	0.08943	0.08787
	10	30	0.04930	0.04711	0.04656	0.04470	0.04665	0.05794	0.05557
	15	30	0.04707	0.04688	0.04690	0.04670	0.04702	0.06736	0.06736
4	8	16	0.04360	0.04211	0.04230	0.04165	0.04336	0.06975	0.07074
	10	30	0.04317	0.04262	0.04263	0.04290	0.04317	0.06255	0.06423
	15	30	0.03940	0.03949	0.03946	0.03986	0.03988	0.04266	0.04281
9	8	16	0.03192	0.03219	0.03209	0.03359	0.03327	0.05037	0.04979
	10	30	0.03145	0.03170	0.03178	0.03263	0.03244	0.03823	0.03656
	15	30	0.02363	0.02362	0.02362	0.02438	0.02418	0.04102	0.04102

Table 4.5. Average values of $|P_2^*(Z) - Q_t|$ when p = 2

δ^2	N_1	N_2	Q_D	Q_{DS}	Q_{MO}	Q_{MOS}	Q_{MC}	Q_R	Q_L
1	8	16	0.07304	0.06402	0.06404	0.05894	0.06569	0.07684	0.07542
	10	30	0.04993	0.04693	0.04953	0.04816	0.05353	0.07039	0.07282
	15	30	0.05387	0.05076	0.05216	0.05002	0.05333	0.06522	0.06785
4	8	16	0.07251	0.06902	0.07143	0.06952	0.07465	0.08242	0.08672
	10	30	0.03633	0.03568	0.03701	0.03819	0.03934	0.05387	0.05360
	15	30	0.03208	0.02974	0.03024	0.03100	0.03156	0.04384	0.04726
9	8	16	0.04159	0.03552	0.03570	0.03530	0.03681	0.05628	0.05800
	10	30	0.02350	0.02260	0.02343	0.02619	0.02485	0.03965	0.03968
	15	30	0.02059	0.02070	0.02106	0.02307	0.02222	0.03310	0.03667

Q_R and Q_L, the estimator Q_L is consistently superior to Q_R. Fourth, among the estimators using normality, it appears that the estimator Q_{MOS} is the best estimator when δ^2 is small relative to p. On the other hand, if δ^2 is large relative to p, there is no estimator which is consistently better than the others; however, for this case, the estimator Q_D is better than the estimator Q_{DS} and the estimator Q_{MO} is better than the estimator Q_{MOS}. Finally, from the simulation study, we observe that the estimator Q_{DS} and Q_{MO} are very similar in their performance.

D. Conclusions

We have considered seven estimators for the unconditional probability of misclassification, $P_2^*(Z)$. Five of these estimators were proposed in Chapter III and are based on normality assumptions; the other two estimators, proposed in Section C of this chapter, do not require any assumptions on the population distribution functions.

When the observations that are being classified come from normal populations, the simulation study shows that the estimators based on normality are consistently superior than the estimators that do not use this assumption. It also indicates that the estimator Q_R is consistently inferior than the estimator Q_L.

Among the estimators based on the normal distribution, we obtained the following conclusions. In Section A of this chapter, it was found that the estimator denoted as Q_{MC}, has the smallest first order bias. In Section B, Chapter IV, we found that these estimators are equivalent, when the comparison is made using only first order terms in the expressions for the AMSE. If in the comparison of the AMSE we include second order terms, we find that no estimator is uniformly better than any of the others for all values of N_1, N_2, p and δ^2. Also, the expressions involved in these comparisons are so complicated that, general conclusions about the relative superiority of any one estimator over the others can be reached only in few cases. For example, we found that, the estimator Q_{DS} is better than the estimator Q_D when $\delta^2 \leq 2p$; the estimator Q_{MOS} is better than the estimator Q_{MO} when $\delta^2 \leq \frac{4}{3}(p+1)$. When δ^2 is large relative to p, the estimator Q_D is better than the estimator Q_{DS} and the estimator Q_{MO} is better than the estimator Q_{MOS}. In those cases, for which it was difficult to analyze the difference between the AMSE of the estimators that were being compared, a numerical evaluation of that difference was done for several values of N_1, N_2, p and δ^2. Although no general conclusions can be obtained from these particular cases, they indicate that in most practical situations, i.e., $1 \leq \delta^2 \leq 10$, the estimator Q_{MOS} is the best estimator for $P_2^*(Z)$ when δ^2 is small relative to p and $p < 8$. When $p \geq 8$, it seems that the best estimator of $P_2^*(Z)$ is the estimator Q_{MC}. The results of the simulation study presented in Section C of this chapter seem to be in agreement with these conclusions.

Finally, it should be noted that most of the conclusions obtained in this thesis are similar to the conclusions obtained by McLachlan (1974b, 1974c) where he considered the problem of estimating various types of probabilities of misclassification using Anderson's W statistic and zero cutoff point.

V. BIBLIOGRAPHY

Aitchison, J. 1975. Goodness of prediction fit. Biometrika 62:547-554.

Aitchison, J. and Dunsmore, I. R. 1975. Statistical Prediction Analysis. Cambridge University Press, Cambridge, U.K.

Aitchison, J., Habbema, J.D.F. and Kay, J. W. 1977. A critical comparison of two methods of statistical discrimination. J. Appl. Stat. 26:15-25.

Anderson, T. W. 1958. An Introduction to Multivariate Statistical Analysis. 1st ed. John Wiley & Sons, Inc., New York.

Anderson, T. W. 1973. Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions. In T. Cacoullos, ed. Discriminant Analysis and Applications. Academic Press, New York.

Broffitt, J. D. and Williams, J. S. 1971. Distributions of functions of $A = (ZZ')^{-1/2}Z$ where the density of Z satisfies certain symmetry conditions. Technical Report No. 7, Dept. of Statistic, Univ. of Iowa, Iowa.

Broffitt, J. D. and Williams, J. S. 1973. Minimum variance estimators for misclassification probabilities in discriminant analysis. J. Multivariate Anal. 3:311-327.

DasGupta, S. 1965. Optimum classification rules for classification into two multivariate normal populations. Ann. Math. Stat. 36:1174-1184.

DasGupta, S. 1973. Theories and methods in classification: A review. In T. Cacoullos, ed. Discriminant Analysis and Applications. Academic Press, New York.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7:179-188.

Geisser, S. 1964. Posterior odds for multivariate normal classification. J. Royal Stat. Soc., Series B, 26:69-76.

Han, C. P. 1979. Alternative methods of estimating the likelihood ratio in classification of multivariate normal observations. The Amer. Statistician 33:204-206.

Hills, M. 1966. Allocation rules and their error rates. J. Royal Stat. Soc., Series B 28:1-32.

John, S. 1960. On some classification problems - I. Sankhya 22:301-308.

John, S. 1963. On classification by the statistics R and Z. Ann. Inst. Stat. Math. 14:237-246.

Johnson, N. L. and Kotz, S. 1972. Distributions in Statistics: Continuous Univariate Distributions. Vol. 2. 1st ed. John Wiley & Sons, Inc., New York.

Kudo, A. 1959. The classification problem viewed as a two decision problem - I. Mem. Fac. Sci. Kyushu Univ., Japan, Series A 13:96-125.

Lachenbruch, P. A. 1967. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics 23:639-645.

Lachenbruch, P. A. 1975. Discriminant Analysis. 1st ed. Hafner Press, New York.

Lachenbruch, P. A. and Mickey, M. R. 1968. Estimation of error rates in discriminant analysis. Technometrics 10:1-11.

Memon, A. Z. and Okamoto, M. 1971. Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. J. Multivariate Anal. 1:294-307.

McLachlan, G. J. 1972. Asymptotic results for discriminant analysis when the initial samples are misclassified. Technometrics 14:415-422.

McLachlan, G. J. 1974a. An asymptotic unbiased technique for estimating the error rates in discriminant analysis. Biometrics 30:239-249.

McLachlan, G. J. 1974b. Estimation of the errors of misclassification on the criterion of asymptotic mean square error. Technometrics 16:255-260.

McLachlan, G. J. 1974c. The relationship in terms of asymptotic mean square error between the separate problem of estimating each of the three types of error rate of the linear discriminant function. Technometrics 16:569-575.

Moran, M. A. and Murphy, B. J. 1979. A closer look at two alternative methods of statistical discrimination. J. Appl. Statistics 28:223-232.

Okamoto, M. 1963. An asymptotic expansion for the distribution of the linear discriminant function. Ann. Math. Stat. 34:1286-1301.

Press, S. J. 1972. Applied Multivariate Analysis. 1st ed. Holt, Rinehart and Winston, Inc., New York.

Sedransk, N. 1969. Contributions to discriminant analysis. Unpublished Ph.D. Thesis. Library, Iowa State University of Science and Technology, Ames, Iowa.

Siotani, M. and Wang, R. H. 1977. Asymptotic expansion for error rates and comparison of the W-procedure and the Z-procedure in discriminant analysis. In P. R. Krishnaiah, ed. Multivariate Analysis IV. North Holland, New York.

Sitgreaves, R. 1952. On the distribution of two random matrices used in classification procedures. Ann. Math. Stat. 23:263-270.

Sitgreaves, R. 1961. Some results of the distribution of the W-classification statistic. Chapter 15 in Salomon, ed. Studies in Item Analysis and Prediction. Stanford University Press, Stanford, Calif.

Smith, C.A.B. 1947. Some examples of discrimination. Annals of Eugenics 13:272-282.

Toussaint, G. T. 1974. Bibliography on estimation of misclassification. IEEE Trans. Inform. Theory IT-20:472-479.

Wald, A. 1944. On a statistical problem arising in the classification of an individual into one of two groups. Ann. Math. Stat. 15:145-162.

VI. ACKNOWLEDGEMENTS

The author wishes to thank Dr. Chien-Pai Han for his continued encouragement and guidance throughout the course of this research.

Special thanks are given to my wife Lucia. Without her love and support, little can be accomplished.

Thanks to Mrs. Maxine Bogue for her excellent typing.

Finally, the author wants to thank CONSEJO NACIONAL DE CIENCIA Y TECHNOLOGIA for providing the funds for the realization of this project.