

THE UNIVERSITY OF MANITOBA

MISCLASSIFICATION ERRORS AND THEIR DISTRIBUTION: ROBUSTNESS

By

Enock Fabiano Ching'anda

A Thesis

Submitted to the Faculty of Graduate Studies
in Partial Fulfilment of the Requirements for the
Degree of Master of Science

Department of Statistics

Winnipeg, Manitoba
Canada

January, 1976



"MISCLASSIFICATION ERRORS AND THEIR DISTRIBUTION: ROBUSTNESS"

by

Enock Fabiano Ching'anda

A dissertation submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

MASTER OF SCIENCE

© 1976

Permission has been granted to the LIBRARY OF THE UNIVERSITY OF MANITOBA to lend or sell copies of this dissertation, to the NATIONAL LIBRARY OF CANADA to microfilm this dissertation and to lend or sell copies of the film, and UNIVERSITY MICROFILMS to publish an abstract of this dissertation.

The author reserves other publication rights, and neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

TABLE OF CONTENTS

	Page
ABSTRACT	i
ACKNOWLEDGEMENT	ii
DEFINITIONS OF TERMS USED IN THE THESIS	iii
CHAPTER 0 INTRODUCTION: SUMMARY	1
CHAPTER 1 REVIEW OF SOME BASIC IDEAS IN	
DISCRIMINANT ANALYSIS	4
1.1 Discrimination, Classification and	
Hypothesis Testing	4
1.2 Errors of Misclassification	5
1.3 Classification Rules	7
1.4 Types of Discriminant Functions	13
a) Fisher's Linear Discriminant Function	
(LDF)	13
b) Anderson's Classification Statistic (W)	16
c) The Quadratic Discriminant Function	
(QDF)	17
d) The Best Linear Discriminant Function	
(BLDF)	18
e) Rao's Classification Statistic (R) . .	19
f) John-Kudo Classification Statistic (Z).	20
1.5 Relationship Among Classification	
Statistics	21
CHAPTER 2 PROBABILITIES OF MISCLASSIFICATION	22
2.1 Introduction: Five Probabilities of Mis-	
classification	22

	Page
a) Optimum Probability of Misclassification .	24
b) Conditional Probability of Misclassification	26
c) Estimated Probability of Misclassification	29
d) Apparent Probability of Misclassification.	30
e) Expected Probability of Misclassification.	31
2.2 Expected Probability of Misclassification: Characteristic Function Approach	36
2.3 Distribution Theory of Probabilities of Misclassification	40
2.4 Asymptotic Expansion of the Probability of Misclassification	44
CHAPTER 3 DISCRIMINATION FOR NON-NORMAL DISTRIBUTIONS:	
JOHNSON'S SYSTEM OF FREQUENCY CURVES	48
3.1 Introduction: Robustness of the Linear Discrimination Function	48
3.2 A General Result	55
3.3 Log Normal Transformation (S)	57
L	
a) Distribution Theory and Rules	57
b) Distribution of the Conditional Probability of Misclassification	60
c) Unconditional Probability of Misclassification	65
3.4 Logit Normal Transformation (S)	68
B	

	Page
a) Distribution Theory and Rules	68
b) Distribution of the Conditional Probability of Misclassification	71
c) Unconditional Probability of Misclassifi- cation	75
3.5 Inverse Hyperbolic Sine Normal (S)	78
a) Distribution Theory and Rules	78
b) Distribution of the Conditional Probability of Misclassification	81
c) Unconditional Probability of Misclassifi- cation	84
CHAPTER 4 DISCRIMINATION FOR NON-NORMAL DISTRIBUTIONS:	
EDGEWORTH SERIES DISTRIBUTION (ESD)	88
4.1 Introduction	88
4.2 Characteristic Function and Moments	91
4.3 Optimum Probability of Misclassification	95
4.4 Conditional Probability of Misclassification.	97
a) Distribution of the Conditional Probability of Misclassification	99
b) Taylor Series Expansion	100
c) Approximate Mean and Variance	104
d) Asymptotic Distribution	106
4.5 Estimated Probability of Misclassification.	107
4.6 Unconditional Probability of Misclassific- ation	108
a) Characteristic Function Approach	109

	Page
b) Conditional Expectation Approach	112
CHAPTER 5 ESTIMATION, COMPARISONS AND RECOMMENDATIONS.	123
5.1 Introduction	123
5.2 Derivation of Classification Rules	124
a) Univariate Normal Distribution (ND)	124
b) Univariate Edgeworth Series Distribution (ESD)	127
5.3 Defining Estimators	129
5.4 Choice of Values for Skewness Factor	130
5.5 Estimation of the Probability of Misclassification	131
a) Optimum Probability of Misclassification	131
b) Apparent Probability of Misclassification	131
5.6 Recommendations and Conclusions	142
APPENDIX A1 COMPUTER SIMULATION OF THE ESD	143
A1.1 Introduction	143
A1.2 Numerical Inverse Interpolation Method	144
A1.3 Newton-Raphson Iteration Technique	145
A1.4 Other algorithms	146
A1.5 Estimation of Parameters and Goodness-Of-Fit Tests	146
a) Estimation of Parameters	146
b) Chi-Square Goodness-Of-Fit Test	147
A1.6 Programs and Flowcharts	153
BIBLIOGRAPHY	167

ABSTRACT

In this thesis the problem of discrimination and classification is studied with a view to assessing the effects of departures from the usual assumptions of normality. Two types of non-normality are considered: (i) Transformable Non-normality where the distribution can be transformed mathematically into a normal form. Three transformations, all belonging to the so-called Johnson's System: Log Normal, Logit Normal and Inverse Hyperbolic Sine Normal, are considered. The distribution theory as well as the classificatory problems associated with these distributions are considered. (ii) Essential or Persistent Non-normality: The distributions of this type cannot be transformed mathematically into the normal form. The Edgeworth Series Distribution is considered here and the type of non-normality is limited to asymmetry. The effects of sampling from a non-normal population is studied. Estimation of probabilities of misclassification using data generated from the Edgeworth Series Distribution are given. Some recommendations based on these studies are also given.

A review of literature concerning probabilities of misclassification is made, the theory is developed for the case of normal distribution. A review is also made of the basic results in discrimination and classification, various types of discriminant functions are discussed.

ACKNOWLEDGEMENT

I would like to thank Dr. Kocherlakota Subrahmaniam for his guidance, the instructive discussions I have had with him and his very useful suggestions during the preparation of the thesis. I would also like to thank the Canadian Development Agency (C.I.D.A.) for sponsoring my study at the University of Manitoba, Canada, and in particular for paying costs incurred during the preparation of the thesis. I would also like to thank the Malawi Government for giving me a study leave to enable me to complete my degree at the University of Manitoba.

The examiners of this thesis provided a great deal of help in revising it. I am grateful to them for this. In particular, Professor H. J. Boom, the external examiner, went through the thesis very carefully.

DEFINITIONS OF TERMS USED IN THE THESIS

TERM	DEFINITION
p	Number of variables.
$\underline{X}$ or $\underline{x}$	$p \times 1$ vector of observations.
Π_i	Population i .
$\underline{\mu}_i$	$p \times 1$ vector in Π_i .
Σ_i	$p \times p$ covariance matrix in Π_i .
$\tilde{\underline{X}}_i$ or $\tilde{\underline{x}}_i$	$p \times 1$ sample mean vector in Π_i .
S_i	$p \times p$ sample covariance matrix in Π_i .
p_i	A priori probability that an observation comes from Π_i .
α_i	Probability of misclassifying an observation from Π_i .
Φ	Cumulative standard normal distribution function.
G or H	Cumulative bivariate normal distribution function.
n_i	Sample size for Π_i .
Δ^2	$(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = \text{Mahalanobis distance squared (parameters known)}.$
$\hat{\Delta}^2$	$(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)' S^{-1} (\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2) = \text{estimated Mahalanobis distance squared (parameters unknown)}.$
k	Number of populations.
$C(2 1)$	Cost of misclassifying an observation from Π_1 .
$C(1 2)$	Cost of misclassifying an observation from Π_2 .

TERM	DEFINITION
R_i	Regions of classification corresponding to the population i .
$f_i(\underline{x})$	Density function of $\underline{X}$ in Π_i .

CHAPTER 0

INTRODUCTION: SUMMARY

In this thesis the problem of discrimination and classification is studied with a view to assessing the effects of departures from the usual assumption of normality. The general procedure for classification of an individual X with p observed characters $(X_1, \dots, X_p)$ consists of determining a function of $(X_1, \dots, X_p)$, approximately, and assigning X to one of the two populations Π_1 or Π_2 depending upon the value of the function. Obviously, since the observation vector is random and since the parameters for determining this function are often not known, this procedure could result in one of two types of errors. The individual X may be classified as belonging to Π_1 when in fact he is from Π_2 or vice versa. These errors are of grave importance in the choice of the procedure. One would require that these errors or, more appropriately, their probabilities be as small as possible. The problems of estimation and optimization of the probabilities have received wide attention.

In Chapter 2 a review is made of the literature concerning probabilities of misclassification. The theory is developed there for the case of a normal distribution. This assumption of normality is made, in part, due to the mathematical intractability of non-normal distributions. Also, in most instances such an assumption is quite realistic. However, it would seem to be of great interest to consider the effects,

if any, of the departures from this simplifying assumption. Some attempts are made in this direction here. There are two types of non-normality one could consider: (i) Transformable non-normality where the distribution can be transformed mathematically into a normal form. The three considered in chapter 3, all belonging to the so-called Johnson's System, are Log Normal, Logit Normal and Inverse Hyperbolic Sine Normal. The distribution theory as well as the classificatory problems associated with these distributions are discussed in this chapter. (ii) Essential or persistent non-normality: The distributions of this type cannot be transformed mathematically into the normal form. This, however, does not preclude the Central Limit Theorem from being applied for large sample situations. The small samples problem is considered in Chapters 4 and 5 for the Edgeworth Series Distribution. The type of non-normality is limited in this thesis to asymmetry.

In chapter 1 a review of the basic results in discrimination and classification is presented. Various types of discriminant functions are discussed. Of these, the one that has been used in this thesis is the Linear Discriminant Function (LDF) introduced by R.A. Fisher (13). The effects of non-normality are studied here when adopting the LDF procedure. Aside from the theoretical implications, it would be of interest to assess the effects of sampling from a non-normal population. This is done in Chapter 5 where the results on the actual estimation of the probabilities of

misclassification using data generated from the Edgeworth Series Distribution are given. Some recommendations based on these studies are also given in this chapter.

Since the generation of this distribution on a computer presents its own difficulties, a discussion of the algorithm along with the actual computer programmes and their flow charts are given in Appendix A1.

CHAPTER 1

REVIEW OF SOME BASIC IDEAS IN DISCRIMINANT ANALYSIS

1.1 Discrimination, Classification and Hypothesis Testing

The concept of hypothesis testing has often been confused with that of discrimination. Rao [49] has discussed this point and has made the distinction clear.

In discrimination we have: (a) a class of alternative hypotheses of which one is chosen; (b) we would like to balance between the wrong and correct decisions; (c) prior probabilities are vital for a satisfactory solution of the problem; and (d) the null hypothesis plays no prominent role in discrimination but is used sometimes to arrive at a decision with a small risk.

In hypothesis testing problems we have: (a) a clearly stated null hypothesis and a comparatively undefined number of alternatives; (b) the emphasis is on the null hypothesis in that we prefer to reject it at a given risk or provisionally accept it; and (c) when we accept the null hypothesis we do not make a decision about the alternative hypothesis.

The concepts of discrimination and classification have been carefully defined by Kendall [25, 26]. In discrimination a sample of members from k populations is taken with p variables, qualitative or quantitative, recorded for each member. A method, based on the recorded variables, is developed for assigning new members to the correct population. In that case we talk of discriminating among the populations. The case

when $k = 2$ is the commonly treated one. In classification we have a sample of n members where each member has p measurable or qualitative characteristics. The question we ask ourselves is whether these members belong to certain groups and, if so, to identify the groups. Anderson [2] defined the problem of classification as that of decision functions where one of a number of hypothesis must be accepted and the rest rejected for a particular observation vector $\underline{X}$.

1.2 Errors of Misclassification

Let $f_i(\underline{x})$ be the probability density function of $\underline{X}$ in Π_i , $i = 1, 2$. We define p_i to be the a priori probability that an observation comes from Π_i , $i = 1, 2$. Let $\underline{X}$ be a point in a p -dimensional region R with R_1 and R_2 as sub-regions. R_1 and R_2 satisfy the property

$$R = R_1 \cup R_2$$

and

$$R_1 \cap R_2 = \phi.$$

Define a classification rule as follows: based on the observation vector $\underline{X}$,

Classify $\underline{X}$ as $\in \Pi_1$ if $\underline{X}$ falls in R_1 ,

Classify $\underline{X}$ as $\in \Pi_2$ if $\underline{X}$ falls in R_2 .

In such a classification procedure we are bound to make two kinds of errors which are defined as errors of misclassification.

If $\underline{X}$ is classified as belonging to Π_1 when in fact $\underline{X}$ is from Π_2 , that is an error; on the other hand, if $\underline{X}$ is classified as belonging to Π_2 when in fact $\underline{X}$ is from Π_1 , that is an error too. In practice such errors could prove costly. A typical realization of the errors and costs involved is as follows:

True State of Nature	Decision	Probability of Misclassification	Cost of Error
$\underline{X} \in \Pi_1$	$\underline{X} \in \Pi_2$	α_1	$C(2 1)$
$\underline{X} \in \Pi_2$	$\underline{X} \in \Pi_1$	α_2	$C(1 2)$

where α_1, α_2 are given by the expressions

$$\alpha_1 = \int_{R_2} f_1(\underline{x}) d\underline{x} \quad (1.1)$$

and

$$\alpha_2 = \int_{R_1} f_2(\underline{x}) d\underline{x}. \quad (1.2)$$

The total probability of misclassification is given by

$$\alpha = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{R_1} f_2(\underline{x}) d\underline{x} \quad (1.3)$$

Details on errors of misclassification are further discussed in Chapter 2.

The costs of errors of misclassification are usually assumed to be greater than zero and are usually known or estimated in some form. For simplicity they are assumed to be unity. The basic question is the choice of R_1 and R_2 . One would obviously choose them so that the total probability of misclassification is minimized. This problem is discussed in the next section.

1.3 Classification Rules

Welch [60] gives a solution for the choice of R_1 and R_2 regions when the total probability of misclassification is minimized and only two populations Π_1 , Π_2 are considered. The total probability of misclassification as given in equation (1.3) can be written as

$$\begin{aligned} \alpha &= p_1 \left[1 - \int_{R_1} f_1(\underline{x}) d\underline{x} \right] + p_2 \int_{R_1} f_2(\underline{x}) d\underline{x} \\ &= p_1 + \int_{R_1} [p_2 f_2(\underline{x}) - p_1 f_1(\underline{x})] d\underline{x}, \end{aligned}$$

which is minimized if R_1 is chosen such that

$$p_2 f_2(\underline{x}) - p_1 f_1(\underline{x}) < 0 \quad \text{for all points in } R_1.$$

Thus the classification rule is: Assign $\underline{x}$ to Π_1 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} > \frac{p_2}{p_1} \tag{1.4}$$

and assign $\underline{X}$ to Π_2 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \frac{p_2}{p_1}. \quad (1.5)$$

Another approach which leads to the same solution as the one in equations (1.4) or (1.5) is through Bayes Theorem. Let the conditional density of $\underline{X}$ in Π_i be $f_i(\underline{x})$, $i = 1, 2$ and a priori probability of $\underline{X}$ from Π_i be p_i , $i = 1, 2$; then the posterior probability of Π_i is given by Bayes Theorem as

$$\Pr(\Pi_i | \underline{x}) = \frac{\Pr(\Pi_i, \underline{x})}{\Pr(\underline{x})} = \frac{p_i f_i(\underline{x})}{p_1 f_1(\underline{x}) + p_2 f_2(\underline{x})} \quad i = 1, 2. \quad (1.6)$$

Assign $\underline{X}$ to the population with the largest posterior probability. Thus the rule becomes: Assign $\underline{X}$ to Π_1

$$\text{if } \Pr(\Pi_1 | \underline{x}) > \Pr(\Pi_2 | \underline{x}), \quad (1.7)$$

that is

$$\text{if } p_1 f_1(\underline{x}) > p_2 f_2(\underline{x})$$

or

$$\text{if } \frac{f_1(\underline{x})}{f_2(\underline{x})} > \frac{p_2}{p_1}. \quad (1.8)$$

The above two procedures are based on equal costs of misclassification in Π_1 and Π_2 . If the costs of misclassification are unequal, by minimizing the total cost of misclassification,

the classification rule is seen to be (see Anderson [1], for example):

Assign $\underline{X}$ to Π_1 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} > \frac{C(1|2)p_2}{C(2|1)p_1} \quad (1.9)$$

and assign $\underline{X}$ to Π_2 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \frac{C(1|2)p_2}{C(2|1)p_1} . \quad (1.10)$$

Another procedure for deriving classification rules is called the minimax procedure. In this case the maximum probability of misclassification is minimized. The procedure has been described by Welch [60] and Lachenbruch [30]. Regions R_1 and R_2 are chosen so that $\alpha_1 = \alpha_2$ and are a minimum under the condition that

$$\int_{R_1} [f_1(\underline{x}) + f_2(\underline{x})] d\underline{x} = 1. \quad (1.11)$$

The minimax rule from the above approach is to classify $\underline{X}$ as a member of Π_1 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} > \beta, \quad (1.12)$$

and to classify $\underline{X}$ as a member of Π_2 if

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \beta .$$

In equations (1.12) and (1.13), β is a constant determined by the distributions in Π_i , $i = 1, 2$.

Another approach which leads to the choice of R_1 and R_2 classification regions is the likelihood ratio (LR) rule. Das Gupta [11] has defined this rule by δ where $\delta_j(x) > 0$ if, for some positive constants $c_1, c_2, \dots, c_m$,

$$C_j f_j(x) < \text{Max}_i \{C_i f_i(x)\} ,$$

where m is the number of groups under consideration. If the C_i 's are all equal, the rule is called the maximum-likelihood (ML) rule. Das Gupta [10] discusses the concept of optimum classification rules in the case of two multivariate normal populations. He shows that the maximum likelihood rule is admissible and minimax in the case when Σ , the dispersion matrix, is known. When Σ is unknown he shows that the maximum likelihood rule is admissible, and, in an invariant class, minimax. To prove the above results Das Gupta uses a loss function which is a function of the Mahalanobis distance $(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$ between the two populations.

If $m = 2$, that is if we have two populations, the classification problem is essentially the problem of testing statistical hypotheses, where a simple null hypothesis against a simple alternative hypothesis are considered. In this case the classification regions are found by use of the Neyman-Pearson Lemma; a proof of this is available in Rao [51]. The likelihood ratio of the observation $\underline{X}$ is given by

$$\ell(\underline{x}) = \frac{f_1(\underline{x})}{f_2(\underline{x})} \quad (1.15)$$

and the likelihood ratio solution is given by

$$\ell(\underline{x}) = \text{a constant.}$$

We note here that the results in equations (1.4, 1.5, 1.8, 1.9, 1.10, 1.12 and 1.13) are of the form given in equation (1.15) except for the difference in the suitable constant. Anderson [2] has derived the likelihood ratio criterion when a composite null hypothesis is being tested against a composite alternative hypothesis. He considers a sample

$$\underline{x}_1^{(1)}, \underline{x}_2^{(1)}, \dots, \underline{x}_{n_1}^{(1)}$$

from $N(\underline{\mu}^{(1)}, \Sigma)$ and a sample

$$\underline{x}_1^{(2)}, \underline{x}_2^{(2)}, \dots, \underline{x}_{n_2}^{(2)}$$

from $N(\underline{\mu}^{(2)}, \Sigma)$ and $\underline{x}$ an observation coming from Π_1 or Π_2 .

Anderson tests the composite hypothesis that

$$\underline{x}, \underline{x}_1^{(1)}, \dots, \underline{x}_{n_1}^{(1)} \text{ are drawn from } N(\underline{\mu}^{(1)}, \Sigma) \text{ and}$$

$$\underline{x}_1^{(2)}, \underline{x}_2^{(2)}, \dots, \underline{x}_{n_2}^{(2)} \text{ are drawn from } N(\underline{\mu}^{(2)}, \Sigma) \text{ against}$$

the alternative composite hypothesis that

$$\underline{x}_1^{(1)}, \underline{x}_2^{(1)}, \dots, \underline{x}_{n_1}^{(1)} \text{ are drawn from } N(\underline{\mu}_1^{(2)}, \Sigma) \text{ and}$$

$\underline{x}, \underline{x}_1^{(2)}, \underline{x}_2^{(2)}, \dots, \underline{x}_{n_2}^{(2)}$ are drawn from $N(\underline{\mu}^{(2)}, \Sigma)$.

By obtaining maximum likelihood estimates of $\underline{\mu}^{(1)}, \underline{\mu}^{(2)}$ and Σ under the two hypotheses, the likelihood ratio is given by

$$D = \frac{1 + \frac{n_2}{n_2+1} (\underline{x} - \tilde{\underline{x}}^{(2)})' C^{-1} (\underline{x} - \tilde{\underline{x}}^{(2)})}{1 + \frac{n_1}{n_1+1} (\underline{x} - \tilde{\underline{x}}^{(1)})' C^{-1} (\underline{x} - \tilde{\underline{x}}^{(1)})} \quad (1.16)$$

where

$$C = \sum_{i=1}^2 \sum_{\alpha=1}^{n_i} (\underline{x}_{\alpha}^{(i)} - \tilde{\underline{x}}^{(i)}) (\underline{x}_{\alpha}^{(i)} - \tilde{\underline{x}}^{(i)})'.$$

The classification region for Π_1 consists of all points for which

$$D > \gamma \quad (1.17)$$

and the classification region for Π_2 consists of all points given by

$$D \leq \gamma \quad (1.18)$$

where γ is a given number.

Hills [18] considers the problem of classification rules as consisting of three categories depending on the information available about the distribution functions in the two populations. If F_1 and F_2 are the distribution functions corresponding to Π_1 and Π_2 respectively, the three categories considered by Hills are as follows:

- (i) F_1, F_2 completely known,
- (ii) F_1, F_2 known to belong to some parametric class of distributions where the parameters are known,

(iii) F_1 , F_2 completely unknown.

If F_1 , F_2 are completely known Hills states that a complete class of admissible rules may be shown to be based on the likelihood ratio rule.

1.4 Types of Discriminant Functions

Since the introduction of the LDF by Fisher several other classification statistics have appeared in the literature. Among reasons for their appearance are (i) violations of the assumptions used in the derivation of the LDF, (ii) attempts to find admissible procedures which minimize the probabilities of misclassification, (iii) attempts to minimize the derivatives of the misclassification errors, (iv) as a problem in hypothesis testing. Some of the classification statistics proposed are:

- (a) Fisher's Linear Discriminant Function (LDF)
- (b) Anderson's Classification Statistic (W)
- (c) The Quadratic Discriminant Function (QDF)
- (d) The Best Linear Classification Statistic (BLDF)
- (e) Rao's Classification Statistic (R)
- (f) John-Kudo Classification Statistic (Z).

A quick review of the basic ideas of each one of these and their properties are given below.

a) Fisher's Linear Discriminant Function (LDF)

The LDF is derived by Fisher [13] while solving a taxonomical problem. Since we refer to this function in most of our discussions, we derive it here.

In the univariate case, when the two populations Π_1, Π_2 are normal with different means μ_1, μ_2 respectively and the same variance σ^2 in both populations, the distance between the means divided by their common standard deviation (S.D) is given by

$$D = \frac{\mu_1 - \mu_2}{\sigma} \quad \text{when} \quad \mu_1 - \mu_2 > 0.$$

Fisher argues that if X is an observation from one of the two populations, X is classified to Π_1 , if it is nearer to μ_1 and X is classified to Π_2 if it is nearer to μ_2 . The risk of misclassifying an observation is small if D is large since the populations are wider apart; if D is small the populations are closer together, hence the errors of misclassification are increased. When the distributions in Π_1, Π_2 are multivariate normal with means $\underline{\mu}_1, \underline{\mu}_2$ respectively and common variance matrix Σ Fisher considers a linear combination of the observations $\underline{X}$

$$Y = \underline{a}'\underline{X}, \quad (1.19)$$

where $\underline{a}' = (a_1, a_2, \dots, a_p)$, $\underline{X}' = (X_1, X_2, \dots, X_p)$ and p is the number of characters. He establishes the discriminant function by maximizing the mean difference in Y for the two populations divided by the common standard deviation of Y within the populations: That is

$$\frac{|E[Y|\Pi_1] - E[Y|\Pi_2]|}{\text{Standard Deviation of } Y}$$

$$\begin{aligned}
&= \frac{|\underline{a}' (\underline{\mu}_1 - \underline{\mu}_2)|}{(\underline{a}' \Sigma \underline{a})^{\frac{1}{2}}} \\
&= \frac{|\underline{a}' \underline{\delta}|}{(\underline{a}' \Sigma \underline{a})^{\frac{1}{2}}} \quad , \quad (1.20)
\end{aligned}$$

where $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$. To maximize equation (1.20) for variations in $\underline{a}$ we differentiate it with respect to $\underline{a}$ and equate to zero, obtaining

$$\underline{\delta} = \left(\frac{\underline{a}' \underline{\delta}}{\underline{a}' \Sigma \underline{a}} \right) \Sigma \underline{a} \quad . \quad (1.21)$$

The above equation has no unique solution; however one solution is given by

$$\underline{a} = \Sigma^{-1} \underline{\delta} \quad (1.22)$$

but any vector proportional to $\underline{a}$ will satisfy equation (1.21).

Substituting the value of $\underline{a}$ into (1.19) we obtain

$$\begin{aligned}
Y &= \underline{\delta}' \Sigma^{-1} \underline{X} \\
&= (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X} \quad (1.23)
\end{aligned}$$

which is the Linear Discriminant Function (LDF).

We formally define the LDF when the parameters are known as

$$D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma) = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X} \quad (1.24)$$

and when the parameters are unknown and, hence, estimated from a sample, as

$$D(\underline{X}; \tilde{\underline{x}}_1, \tilde{\underline{x}}_2; S) = (\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)' S^{-1} \underline{X} \quad (1.25)$$

where $\tilde{\underline{x}}_1, \tilde{\underline{x}}_2, S$ are estimates of $\underline{\mu}_1, \underline{\mu}_2, \Sigma$, respectively.

Welch [60] shows that even if the assumptions which Fisher made in deriving the LDF do not hold, the best function to discriminate between two populations whose density functions are completely specified may be obtained by taking the ratio of the probability densities in the two populations. He deduces the likelihood ratio argument of alternative hypothesis from the Bayes Theorem and shows that for any prior probability the discriminant function is a ratio of the probability densities and the criterion level depends on prior probabilities.

b) Anderson's Classification Statistic (W)

This statistic is derived by Anderson [2] in the case of two populations whose characters are distributed as multivariate normal with different means and the same variance covariance matrix. The likelihood ratio rule [see equation (1.15)] is used in the derivation. The statistic is defined as

$$D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma) = [\underline{X} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)]' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \quad (1.26)$$

when the parameters are known and as

$$D(\underline{X}; \tilde{\underline{x}}_1, \tilde{\underline{x}}_2; S) = [\underline{X} - \frac{1}{2}(\tilde{\underline{x}}_1 + \tilde{\underline{x}}_2)]' S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2) \quad (1.27)$$

when the parameters are unknown, hence estimated from samples. The W statistic differs from the LDF, when the parameters are known, by the constant

$$\frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) . \quad (1.28)$$

If the parameters are estimated then this difference is

$$\frac{1}{2} (\tilde{\underline{x}}_1 + \tilde{\underline{x}}_2)' S^{-1} (\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)$$

The univariate analogue of this statistic has been given by John[21] and by Sedransk and Okamoto [53].

c) The Quadratic Discriminant Function (QDF)

This statistic arises when the assumption of equal dispersion matrices in the two populations no longer holds and the likelihood ratio rule is used in deriving the discriminant function. When the parameters are known, the statistic is defined as;

$$D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma_1, \Sigma_2) = \frac{1}{2} \frac{|\Sigma_2|}{|\Sigma_1|} +$$

$$- \frac{1}{2} (\underline{X} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{X} - \underline{\mu}_1) + \frac{1}{2} (\underline{X} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{X} - \underline{\mu}_2) \quad (1.30)$$

and when the parameters are not known and are estimated from samples, the statistic is defined as

$$D(\underline{X}; \underline{\tilde{x}}_1, \underline{\tilde{x}}_2; S_1, S_2) = \frac{1}{2} \frac{|S_2|}{|S_1|} -$$

$$+ \frac{1}{2}(\underline{X} - \underline{\tilde{x}}_1)' S_1^{-1}(\underline{X} - \underline{\tilde{x}}_1) + \frac{1}{2}(\underline{X} - \underline{\tilde{x}}_2)' S_2^{-1}(\underline{X} - \underline{\tilde{x}}_2)$$

(1.31)

Okamoto [45] introduced this originally. Bartlett and Please [6] independently developed this function for discriminating between monozygotic and dizygotic twins. Gilbert [16] studied the robustness of the QDF when the variance matrices are known and has looked at special forms of these matrices, $\Sigma_1 = I$ and $\Sigma_2 = dI$ where d is a constant. Marks and Dunn [35] have extended Gilbert's results to the more general set up of the parameters in the two distributions being unknown. Lachenbruch, Sneeringer and Revo [32] have considered the robustness of QDF with respect to non-normality. The QDF is optimal in the sense of minimizing the overall probability of misclassification when the parameters are known and the covariance matrices are not equal.

d) The Best Linear Discriminant Function (BLDF)

This statistic is introduced by Clumies-Ross and Riffenburgh [9] and by Anderson and Bahadur [5] under conditions similar to those of the QDF. The main condition is that $\Sigma_1 \neq \Sigma_2$. When the parameters are all known, the statistic is defined as

$$D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma_1, \Sigma_2) = \begin{cases} \underline{X}'\underline{b} - \underline{b}'\underline{\mu}_1 - t_1 \underline{b}'\Sigma_1\underline{b} \\ \underline{X}'\underline{b} - \underline{b}'\underline{\mu}_2 + t_2 \underline{b}'\Sigma_2\underline{b} \end{cases}$$

(1.32)

and when the parameters are not known but are estimated from the samples, the statistic is defined as

$$D(\underline{X}; \tilde{\underline{x}}_1, \tilde{\underline{x}}_2; S_1, S_2) = \begin{cases} \underline{x}'\underline{b} - \underline{b}'\tilde{\underline{x}}_1 - t_1\underline{b}'S_1\underline{b} \\ \underline{x}'\underline{b} - \underline{b}'\tilde{\underline{x}}_2 + t_2\underline{b}'S_2\underline{b} \end{cases} \quad (1.33)$$

In (1.32)

$$\underline{b} = [t_1 S_1 + t_2 S_2]^{-1} (\underline{\mu}_2 - \underline{\mu}_1) \quad (1.34)$$

and in (1.33)

$$\underline{b} = [t_1 S_1 + t_2 S_2]^{-1} (\tilde{\underline{x}}_2 - \tilde{\underline{x}}_1) \quad (1.35)$$

In both these equations t_1, t_2 are scalars chosen by a systematic trial and error method to minimize the probability of misclassification.

When the parameters are unknown and estimated from a sample, Marks and Dunn [35] refer to BLDF as the Best Linear Sample Discriminant Function. This statistic is optimal in the same sense as the QDF.

e) Rao's Classification Statistic (R)

Rao [50] derives the statistic R under the condition of two populations Π_1, Π_2 and the distribution of $\underline{X}$ in the two populations being multivariate normal. He assumes that samples of size n_1, n_2 are obtained from Π_1, Π_2 respectively, Under conditions that the derivative of the misclassification errors are being minimized Rao comes up with the statistic R defined

as follows:

$$D(\underline{X}; \underline{\tilde{x}}_1, \underline{\tilde{x}}_2; n_1, n_2; \Sigma) = \frac{a_1^2 - a_2^2}{b_2^2} U \Sigma^{-1} U' + \frac{2(a_1 - a_2)}{b_1 b_2} U \Sigma^{-1} T' \quad (1.36)$$

where $a_1 = n_2(n_1 + n_2)^{-1}$ $b_1 = n_1^{-1} + n_2^{-2},$

$$a_2 = -n_1(n_1 + n_2)^{-1}, \quad b_2 = 1 + (n_1 + n_2)^{-1},$$

$$U = \underline{X} - (n_1 \underline{\tilde{x}}_1 + n_2 \underline{\tilde{x}}_2)(n_1 + n_2)^{-1}, \quad T = \underline{\tilde{x}}_1 - \underline{\tilde{x}}_2.$$

Rao shows that this statistic is best for discriminating between alternatives that are close to each other.

John [22] gives a form of Rao's Statistic when the variance matrix is estimated by S .

f) John-Kudo Classification Statistic (Z)

The statistic has been suggested under the condition that the distribution of $\underline{X}$ in the two populations is multivariate normal with a common variance matrix and different and unknown means. John [22,23] defines the statistic as

$$D(\underline{X}; \underline{\tilde{x}}_1, \underline{\tilde{x}}_2; n_1, n_2; S) = \frac{n_1}{n_1+1} (\underline{X} - \underline{\tilde{x}}_1)' S^{-1} (\underline{X} - \underline{\tilde{x}}_1) - n \frac{n_2}{n_2+1} (\underline{X} - \underline{\tilde{x}}_2)' S^{-1} (\underline{X} - \underline{\tilde{x}}_2)$$

where S is the estimated dispersion matrix and n is a constant. John [19] suggests the same form with $n = 1$ and the dispersion matrix known. Kudo [27] had discussed this statistic independently in an earlier work. John comments

that the statistic Z with $\eta = 1$ was proposed because it appeared reasonable to give the individual to that population which, in testing, rejects it at a higher level. Presumably, Rao [50, p. 655] had previously stated this in principle.

1.5 Relationships Among Classification Statistics

From the discussion in section 1.4, we note that the LDF is the basic discriminant function among the classification statistics. The QDF reduces to the LDF when the variance matrices Σ_1, Σ_2 are equal. The statistics W, R and Z (with $\eta = 1$) are asymptotically equivalent. R and W statistics are equivalent if sample sizes n_1, n_2 are equal. The Z and W statistics are equivalent if $\frac{n_1}{n_1+1} = \eta \left(\frac{n_2}{n_2+1} \right)$ and hence equivalent to R . In particular, if $\eta = 1$ and $n_1 = n_2 = n$, the Z and W statistics are linearly related in the form $Z = - \left(\frac{2n}{n+1} \right) W$. Siotani [54] states that, if $n_1 = n_2$ and $n = n_1 + n_2 - 2$ are sufficiently large, then $\frac{n_1}{n_i+1} \approx 1$, for $i = 1, 2$, hence the W and Z statistics are equivalent. The statistic W differs from the LDF only by a constant.

CHAPTER 2

PROBABILITIES OF MISCLASSIFICATION

2.1 Introduction: Five Probabilities of Misclassification

In this chapter a review is made of the literature concerning probabilities of misclassification. The assignment of an observation vector $\underline{X}$ to one of two populations Π_1, Π_2 with classification regions R_1, R_2 respectively, depends upon the value of the discriminant function. The discriminant function is a function of the observations and the parameters of the distributions in the two populations. The parameters are usually unknown and hence are estimated from samples. This obviously means loss of information on the parameters which affects the estimation of the probabilities of misclassification, in that they may be underestimated or overestimated. Various techniques have been developed and compared in trying to find the best way of estimating the true probabilities of misclassification when the parameters are unknown and are estimated from samples. The techniques have led to a number of definitions of misclassification errors. We discuss five probabilities of misclassification here. The probabilities of misclassification are important since, after constructing a discriminant function, we are interested in determining how this function performs on future samples. The performance criterion of a suggested procedure is usually the probability of misclassification. The probabilities of misclassification which we consider here are as follows:

- (a) Optimum probability of misclassification,
- (b) Conditional probability of misclassification,
- (c) Estimated probability of misclassification,
- (d) Apparent probability of misclassification,
- (e) Expected probability of misclassification.

Before these probabilities of misclassification are discussed it is useful to introduce the following notation:

- (i) Classification regions corresponding to Π_1 and Π_2 populations are defined as R_1, R_2 when parameters are known
and $\hat{R}_1, \hat{R}_2$ when parameters are unknown,
hence estimated from samples.

As usual $R_1, R_2, \hat{R}_1, \hat{R}_2$ satisfy the conditions that

$$R_1 \cup R_2 = R, \quad R_1 \cap R_2 = \phi$$

$$\hat{R}_1 \cup \hat{R}_2 = R, \quad \hat{R}_1 \cap \hat{R}_2 = \phi,$$

- (ii) The distributions corresponding to populations Π_1 and Π_2 are defined as

f_1, f_2 when the parameters are known

$\hat{f}_1, \hat{f}_2$ when the parameters are estimated.

- (iii) $\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$ is the Mahalanobis distance squared.

$\hat{\Delta}^2 = (\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)' S^{-1} (\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)$ is the estimated Mahalanobis distance squared.

This notation, and the notation used in defining probabilities of misclassification in the remaining part of this chapter,

follows that of Lachenbruch [30].

a) Optimum Probability of Misclassification

This probability of misclassification is said to be optimum in that we cannot improve upon it and assumes the parameters of the distributions in the two populations known. The total optimum probability of misclassification is defined as

$$\alpha[R, f] = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{R_1} f_2(\underline{x}) d\underline{x} \quad (2.1)$$

where the first argument R refers to the entire region of classification and the second argument f refers to the distribution of the observations that will be classified; p_i refers to the a priori probability that an observation comes from Π_i , $i = 1, 2$.

In the univariate situation, when the distributions of X in Π_1, Π_2 are normal, that is

$$X \text{ in } \Pi_1 \text{ is } \sim N(\mu_1, 1)$$

$$X \text{ in } \Pi_2 \text{ is } \sim N(\mu_2, 1)$$

and the classification regions are as follows:

$$\begin{aligned} R_1: \{X; X \leq \frac{1}{2} (\mu_1 + \mu_2)\} \\ R_2: \{X; X > \frac{1}{2} (\mu_1 + \mu_2)\} \end{aligned} \quad (2.2)$$

the optimum probability of misclassification when an observation from Π_1 is misclassified is given by

$$\alpha_1[R, f] = \int_{R_2} f_1(x) dx = 1 - \Phi\left(\frac{1}{2} \delta\right) = \Phi\left(-\frac{\delta}{2}\right) \quad (2.3)$$

where $\delta = \mu_2 - \mu_1 > 0$. When an observation from Π_2 is misclassified we have

$$\alpha_2[R, f] = \int_{R_1} f_2(x) dx = \Phi\left(-\frac{\delta}{2}\right). \quad (2.4)$$

The total probability of misclassification when $p_1 = \frac{1}{2}$ and $p_2 = 1 - p_1$ is given by

$$\begin{aligned} \alpha[R, f] &= \frac{1}{2} \int_{R_2} f_1(x) dx + \frac{1}{2} \int_{R_1} f_2(x) dx \\ &= \frac{1}{2} \Phi\left(-\frac{\delta}{2}\right) + \frac{1}{2} \Phi\left(-\frac{\delta}{2}\right) \\ &= \Phi\left(-\frac{\delta}{2}\right). \end{aligned} \quad (2.5)$$

Sedransk and Okamoto [53] have given similar results on this probability of misclassification when the variance in the two populations Π_1, Π_2 is given by σ^2 .

Let $\underline{x}$ in Π_1 and Π_2 have the density functions

$$f_j(\underline{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu}_j)' \Sigma^{-1}(\underline{x} - \underline{\mu}_j)\right\} \quad j = 1, 2. \quad (2.6)$$

The parameters $\underline{\mu}_j$ and Σ satisfy the conditions $-\infty < \underline{\mu}_j < \infty$ and Σ is a positive definite symmetric matrix of order p . The optimum probabilities of misclassification

based on the classification regions

$$R_1: \{ \underline{X}; D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma) = [\underline{X} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)]' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) > 0 \}$$

$$R_2: \{ \underline{X}; D(\underline{X}; \underline{\mu}_1, \underline{\mu}_2; \Sigma) = [\underline{X} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)]' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \leq 0 \}$$

(2.7)

are given by

$$\alpha_1[R, f] = \Phi\left(-\frac{\Delta}{2}\right), \quad (2.8)$$

$$\alpha_2[R, f] = \Phi\left(-\frac{\Delta}{2}\right). \quad (2.9)$$

see Hills [18], Lachenbruch [28] and Sorum [58], for example.

When $p_1 = p_2 = \frac{1}{2}$ the optimum probability of misclassification is given by

$$\alpha[R, f] = \Phi\left(-\frac{\Delta}{2}\right). \quad (2.10)$$

Results in equations (2.8), (2.9) and (2.10) show that the probability of misclassification is a monotone decreasing function of Δ . As $\Delta \rightarrow \infty$, this probability tends to zero. One would expect this, as large Δ implies that the populations are wide apart.

b) Conditional Probability of Misclassification

The conditional probability of misclassification is usually calculated when a sample discriminant function is involved in the classification rule. Given a discriminant function the probability can be described as the conditional probability that a randomly chosen member of Π_i ($i = 1, 2$) is misallocated.

It is conditional on the observation coming from one of Π_1 or Π_2 and also conditional on estimates of the means of the distribution in Π_1 and Π_2 . The total conditional probability of misclassification is defined as

$$\alpha[\hat{R}, f] = p_1 \int_{\hat{R}_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{\hat{R}_1} f_2(\underline{x}) d\underline{x} \quad (2.11)$$

This probability of misclassification is sometimes called the actual probability of misclassification (see Hills [18] and Moran [44], for example). In the univariate situation, when the distributions of X in Π_1, Π_2 are normal, that is

$$X \text{ in } \Pi_1 \text{ is } \sim N(\mu_1, \sigma_1^2),$$

$$X \text{ in } \Pi_2 \text{ is } \sim N(\mu_2, \sigma_2^2),$$

and the classification regions are

$$\begin{aligned} \hat{R}_1: \{X; X \leq \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \text{ and } \bar{x}_1 < \bar{x}_2 \text{ or if } X > \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \\ \text{and } \bar{x}_1 \geq \bar{x}_2\} \\ \hat{R}_2: \{X; X > \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \text{ and } \bar{x}_1 < \bar{x}_2 \text{ or if } X \leq \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \\ \text{and } \bar{x}_1 \geq \bar{x}_2\}, \end{aligned} \quad (2.12)$$

John [21] has obtained the conditional probabilities of misclassification in this case to be:

$$\alpha_1[\hat{R}, f] = \begin{cases} 1 - \Phi\left[\frac{1}{\sigma_1} \left\{ \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_1 \right\}\right] & \text{if } \bar{x}_1 < \bar{x}_2 \\ \Phi\left[\frac{1}{\sigma_1} \left\{ \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_1 \right\}\right] & \text{if } \bar{x}_1 \geq \bar{x}_2 \end{cases} \quad (2.13)$$

$$\alpha_2[\hat{R}, f] = \begin{cases} \Phi[\sigma_2^{-1}\{\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2\}] & \text{if } \bar{x}_1 < \bar{x}_2 \\ 1 - \Phi[\sigma_2^{-1}\{\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2\}] & \text{if } \bar{x}_1 \geq \bar{x}_2. \end{cases} \quad (2.14)$$

If $\sigma_1^2 = \sigma_2^2 = 1$ in Π_1 and Π_2 , the probabilities of misclassification reduce to

$$\alpha_1[\hat{R}, f] = \begin{cases} \Phi[\mu_1 - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)] & \text{if } \bar{x}_1 < \bar{x}_2 \\ \Phi[\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_1] & \text{if } \bar{x}_1 \geq \bar{x}_2 \end{cases} \quad (2.15)$$

$$\alpha_2[\hat{R}, f] = \begin{cases} \Phi[\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2] & \text{if } \bar{x}_1 < \bar{x}_2 \\ \Phi[\mu_2 - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)] & \text{if } \bar{x}_1 \geq \bar{x}_2 \end{cases} \quad (2.16)$$

(see Hills [18], for example). In the multivariate situation, when $\underline{X}$ in Π_1, Π_2 has density functions as given in equation (2.6) and the classification regions $\hat{R}_1, \hat{R}_2$ are given by

$$\begin{aligned} \hat{R}_1: \{ \underline{X}; D(\underline{X}; \tilde{\underline{x}}_1, \tilde{\underline{x}}_2; S) = [\underline{X} - \frac{1}{2}(\tilde{\underline{x}}_1 + \tilde{\underline{x}}_2)]' S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2) > 0 \} \\ \hat{R}_2: \{ \underline{X}; D(\underline{X}; \tilde{\underline{x}}_1, \tilde{\underline{x}}_2; S) = [\underline{X} - \frac{1}{2}(\tilde{\underline{x}}_1 + \tilde{\underline{x}}_2)]' S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2) \leq 0 \}, \end{aligned} \quad (2.17)$$

the conditional probability of misclassification when an observation from Π_1 is misallocated is given by

$$\alpha_1[\hat{R}, f] = \Phi \left[\frac{-\underline{\mu}_1' S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2) + \frac{1}{2}(\tilde{\underline{x}}_1 + \tilde{\underline{x}}_2)' S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)}{\{(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)' S^{-1} \Sigma S^{-1}(\tilde{\underline{x}}_1 - \tilde{\underline{x}}_2)\}^{1/2}} \right]. \quad (2.18)$$

John [21] discusses the conditional probability of misclassification under various assumptions about $\underline{\mu}_1$, $\underline{\mu}_2$ and Σ in the multivariate normal situation. Details are given in a later section.

c) Estimated Probability of Misclassification

The Estimated Probability of Misclassification is sometimes referred to as the "plug-in estimate" since the maximum likelihood estimates of the parameters are usually plugged in the discriminant function before classification. This method is suggested by Fisher [13]. The total estimated probability of misclassification is defined as

$$\alpha[\hat{R}, \hat{f}] = p_1 \int_{\hat{R}_2} \hat{f}_1(\underline{x}) d\underline{x} + p_2 \int_{\hat{R}_1} \hat{f}_2(\underline{x}) d\underline{x}. \quad (2.19)$$

In the univariate situation when the distributions of X in Π_1 , Π_2 are normal, that is

$$X \text{ in } \Pi_1 \text{ is } \sim N(\mu_1, 1)$$

$$X \text{ in } \Pi_2 \text{ is } \sim N(\mu_2, 1)$$

and the regions of classification are as given in equation (2.12), then the estimated probabilities of misclassification are given by

$$\alpha_1[\hat{R}, \hat{f}] = 1 - \Phi\left[\frac{1}{2} |\hat{\delta}|\right] = \Phi\left[-\frac{1}{2} |\hat{\delta}|\right], \quad (2.20)$$

$$\alpha_2[\hat{R}, \hat{f}] = \Phi\left[-\frac{1}{2} |\hat{\Delta}| \right], \quad (2.21)$$

where $\hat{\Delta} = (\bar{x}_2 - \bar{x}_1)$ (see Hills [18], for example). In the multivariate situation, when $\underline{X}$ in Π_1 and Π_2 has density functions as given in equation (2.6) and the classification regions are given in equation (2.17), the estimated probabilities of misclassification are given by

$$\alpha_1[\hat{R}, \hat{f}] = \Phi\left(-\frac{\hat{\Delta}}{2}\right) \quad (2.22)$$

$$\alpha_2[\hat{R}, \hat{f}] = \Phi\left(-\frac{\hat{\Delta}}{2}\right) \quad (2.23)$$

(see Hills [18] and Lachenbruch [30], for example).

The total estimated probability of misclassification when the distributions in Π_1, Π_2 are as given in equation (2.6) and $p_1 = p_2 = \frac{1}{2}$ is

$$\alpha[\hat{R}, \hat{f}] = \Phi\left(-\frac{\hat{\Delta}}{2}\right). \quad (2.24)$$

Lachenbruch [28] discusses the use of a plug-in procedure to estimate this probability and concludes that, since $\hat{\Delta}^2$ overestimates Δ^2 , $\alpha_1[\hat{R}, \hat{f}]$ or $\alpha_2[\hat{R}, \hat{f}]$ underestimates the true probability of misclassification. He further comments ([30], p. 32) that the total estimated probability of misclassification is a maximum likelihood estimate of error, it is consistent and asymptotically efficient.

d) Apparent Probability of Misclassification

This probability of misclassification is defined as the

proportion of observations in the initial sample which are misclassified by the sample discriminant function. It is a non-parametric measure of the probability of misclassification.

It was suggested by Smith [56] in connection with the QDF. In general, if m_1 is the proportion misclassified by the discriminant function in Π_1 and n_1 is the total sample size in Π_1 , then the apparent probability of misclassification is given by the ratio m_1/n_1 .

e) Expected Probability of Misclassification

Most of the discussion in the literature on this probability of misclassification has been done in the light of Actual, Estimated and Apparent error rates. Mention can be made of John [21], Sorum [59] and Moran [44] and others in this connection. We define this probability as the expected value of the conditional probability of misclassification which we have discussed in section (b) of this chapter. It is sometimes called the unconditional probability of misclassification. The total expected probability of misclassification is defined as

$$E[\alpha(\hat{R}, f)] = E[p_1 \int_{\hat{R}_2} f_1(\underline{x}) d\underline{x} + p_2 \int_{\hat{R}_1} f_2(\underline{x}) d\underline{x}] \quad (2.25)$$

over all samples of sizes n_1, n_2 from Π_1 and Π_2 respectively.

When the distributions in Π_1 and Π_2 of the random variable $\underline{x}$ are given by the density functions as given in equation (2.6), John [21] gives expressions for the expected probability of misclassification and its approximations using

the W discriminant function as the classification statistic. John's result is that

$$E[\alpha_1(\hat{R}, \hat{f})] = \Pr\{W_1/W_2 > (1 - \rho_1)/(1 + \rho_1)\} \quad (2.26)$$

where W_1, W_2 are distributed as noncentral chi-square with p degrees of freedom and noncentrality parameters given by

$$\begin{aligned} \lambda_1 &= \{2(1 + \rho_1)\}^{-1} n_1 n_2 [(n_1 + n_2)^{-\frac{1}{2}} + \\ &\quad - (n_1 + n_2 + 4n_1 n_2)^{-\frac{1}{2}}]^2 \Delta^2, \\ \lambda_2 &= \{2(1 - \rho_1)\}^{-1} n_1 n_2 [(n_1 + n_2)^{-\frac{1}{2}} + \\ &\quad + (n_1 + n_2 + 4n_1 n_2)^{-\frac{1}{2}}]^2 \Delta^2, \end{aligned} \quad (2.27)$$

respectively and

$$\rho_1 = (n_2 - n_1) [(n_1 + n_2)(n_1 + n_2 + 4n_1 n_2)]^{-\frac{1}{2}}. \quad (2.28)$$

In the univariate situation, John [21] considers the expectation of the actual probability of misclassification when an observation from Π_1 is misclassified and the univariate analogue of the W discriminant function is used. The distributions of X in Π_1 and Π_2 are given by

$$X \sim N(\mu_1, \sigma_1^2)$$

and

$$X \sim N(\mu_2, \sigma_2^2),$$

respectively. We note that, although the W function assumes the variances in Π_1, Π_2 to be the same, in finding the

expectations of the probability this assumption is dropped.

The classification regions are as those given in equation (2.12).

Since a wrong assignment of an observation from Π_1 corresponds to two events

$$\bar{x}_1 < \bar{x}_2 \quad \text{and} \quad x \geq \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$$

or

$$\bar{x}_1 \geq \bar{x}_2 \quad \text{and} \quad x \leq \frac{1}{2}(\bar{x}_1 + \bar{x}_2) .$$

(2.29)

John [21] establishes that

$$E[\alpha_1(\hat{R}, f)] = G(a_{11}, a_{21}; \rho) + G(a_{12}, a_{22}; \rho). \quad (2.30)$$

Here

$$a_{11} = -[n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2](\mu_2 - \mu_1) = -a_{12}, \quad (2.31)$$

$$a_{21} = \frac{1}{2}[\sigma_1^2 + \frac{1}{4}(n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)]^{-\frac{1}{2}}(\mu_2 - \mu_1) = -a_{22}, \quad (2.32)$$

$$\rho = \frac{1}{2}[n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2]^{-\frac{1}{2}}[\sigma_1^2 + \frac{1}{4}(n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)]^{-\frac{1}{2}} \\ (n_1^{-1}\sigma_1^2 - n_2^{-1}\sigma_2^2) \quad (2.33)$$

and

$$G(h, k; \rho) = \int_{-\infty}^h \int_{-\infty}^k g[u, v; \rho] du dv \quad (2.34)$$

where $g[u, v; \rho]$ is the standard bivariate normal density function with correlation coefficient ρ .

Further, when $\rho = 0$,

$$E[\alpha_1(\hat{R}, f)] = \phi(a_{11})\phi(a_{21}) + \phi(a_{12})\phi(a_{22}), \quad (2.35)$$

In section (2.2) an alternative proof of (2.30) is given using characteristic functions.

John [22] gives more results on the expectations of the probability of misclassification when the statistics R and Z are used as classification statistics and the covariance matrix is unknown.

Hills [18] shows that the expectation of $\alpha_1[\hat{R}, f]$, $\alpha_1[\hat{R}, \hat{f}]$ and m_1/n_1 , when the distributions of X in Π_1, Π_2 are univariate normal, that is

$$X \sim N(\mu_1, 1) \text{ in } \Pi_1$$

and

$$X \sim N(\mu_2, 1) \text{ in } \Pi_2,$$

are given by

$$E[\alpha_1(\hat{R}, f)] = \phi(-a) + \phi(-b) - 2G(a, b, \rho_1) \quad (2.36)$$

$$E[\alpha_1(\hat{R}, \hat{f})] = \phi(-a) + \phi(-b) - 2G(a, b, \rho_3) \quad (2.37)$$

$$E\left[\frac{m_1}{n_1}\right] = \phi(-c) + \phi(-b) - 2G(a, b, \rho_2) \quad (2.38)$$

where in equations (2.36), (2.37), (2.38) the notation given in equation (2.34) is used and

$$a = \frac{1}{2} \delta / (1 + h)^{\frac{1}{2}}, \quad b = \frac{1}{2} \delta / h^{\frac{1}{2}}, \quad c = \frac{1}{2} \delta / (1 + h - 1/n_1) \quad (2.39)$$

$$\rho_1 = - \frac{\frac{1}{4}(\frac{1}{n_1} - \frac{1}{n_2})}{[h(1 + h)]^{\frac{1}{2}}} \quad (2.40)$$

$$\rho_2 = \left(\frac{h}{1 + h - \frac{1}{n_1}} \right), \quad (2.41)$$

$$\rho_3 = \left(\frac{h}{1 + h} \right)^{\frac{1}{2}} \quad (2.42)$$

and

$$h = \frac{1}{4} \left(\frac{1}{n_1} + \frac{1}{n_2} \right), \quad \delta = \mu_2 - \mu_1 > 0. \quad (2.43)$$

Moran [44] proves the result given by John in equation (2.26) since John gives no explicit justification of the result although he indicates that a proof can be constructed using results contained in an earlier paper (John [20]). Moran has extended these results to $E[\alpha(\hat{R}, \hat{f})]$ and $E[m_1/n_1]$. By a similar method he has given

$$E[\alpha(\hat{R}, \hat{f})] = \Pr \left(\frac{W_5}{W_6} > \frac{1 - \rho_3}{1 + \rho_3} \right) \quad (2.44)$$

where W_5 and W_6 are independent noncentral chi-squares with p degrees of freedom and noncentrality parameters λ_5 , λ_6 given by

$$\lambda_5 = [2(1 + \rho_3)]^{-1} n_1 n_2 \{ (n_1 + n_2)^{-\frac{1}{2}} - (n_1 + n_2 - 4n_1 n_2)^{-\frac{1}{2}} \}^2 \Delta^2, \quad (2.45)$$

$$\lambda_6 = [2(1 - \rho_3)]^{-1} n_1 n_2 \{ (n_1 + n_2)^{-\frac{1}{2}} + (n_1 + n_2 - 4n_1 n_2)^{-\frac{1}{2}} \}^2 \Delta^2 \quad (2.46)$$

and

$$\rho_3 = -(n_1 + n_2)^{\frac{1}{2}}(n_1 + n_2 + 4n_1n_2)^{-\frac{1}{2}}. \quad (2.47)$$

In the case of the apparent probability of misclassification the result is given by

$$E\left[\frac{m_1}{n_1}\right] = \Pr\left(\frac{W_3}{W_4} > \frac{1 - \rho_2}{1 + \rho_2}\right) \quad (2.48)$$

where W_3, W_4 are independent noncentral chisquares with p degrees of freedom and noncentrality parameters given by

$$\lambda_3 = [2(1 + \rho_2)]^{-1} n_1n_2\{(n_1 + n_2)^{-\frac{1}{2}} - (n_1 - 3n_2 + 4n_1n_2)^{-\frac{1}{2}}\}^2 \Delta^2, \quad (2.49)$$

$$\lambda_4 = [2(1 - \rho_2)]^{-1} n_1n_2\{(n_1 + n_2)^{-\frac{1}{2}} + (n_1 - 3n_2 + 4n_1n_2)^{-\frac{1}{2}}\}^2 \Delta^2 \quad (2.50)$$

and

$$\rho_2 = -(n_1 + n_2)^{\frac{1}{2}}(n_1 - 3n_2 + 4n_1n_2)^{-\frac{1}{2}}. \quad (2.51)$$

Moran does not evaluate expressions when Σ is unknown and hence estimated by S ; he has stated that the probability can be expressed in terms of the cumulative distribution of the ratio of two dependent Hotelling's T^2 statistics.

2.2 Expected Probability of Misclassification: Characteristic Function Approach

We present an alternative proof of the result given in equation (2.30). The proof, which seems to be more elegant, is achieved by use of the characteristic function.

Since a wrong assignment of an observation from Π_1 corresponds to the events

$$\bar{x}_1 < \bar{x}_2 \quad \text{and} \quad X \geq \frac{\bar{x}_1 + \bar{x}_2}{2}$$

or

$$\bar{x}_1 \geq \bar{x}_2 \quad \text{and} \quad X \leq \frac{\bar{x}_1 + \bar{x}_2}{2}, \quad (2.52)$$

the expected probability of misclassification is given by

$$\begin{aligned} E[\alpha_1(\hat{R}, \hat{f})] &= \Pr\{\bar{x}_1 < \bar{x}_2 \quad \text{and} \quad X \geq \frac{\bar{x}_1 + \bar{x}_2}{2} | X \in \Pi_1\} + \\ &\quad \Pr\{\bar{x}_1 \geq \bar{x}_2 \quad \text{and} \quad X \leq \frac{\bar{x}_1 + \bar{x}_2}{2} | X \in \Pi_1\} \\ &= \Pr\{\bar{x}_2 - \bar{x}_1 > 0 \quad \text{and} \quad X - \frac{\bar{x}_1 + \bar{x}_2}{2} \geq 0 | X \in \Pi_1\} \\ &\quad + \Pr\{\bar{x}_2 - \bar{x}_1 \leq 0 \quad \text{and} \quad X - \frac{\bar{x}_1 + \bar{x}_2}{2} \leq 0 | X \in \Pi_1\} \end{aligned} \quad (2.53)$$

The joint distribution of $\bar{x}_2 - \bar{x}_1$ and $X - \frac{\bar{x}_1 + \bar{x}_2}{2}$ is required for the determination of this probability. This can be done by use of the characteristic function.

Since X in Π_1 has a $N(\mu_1, \sigma_1^2)$, the characteristic function of X is given by

$$\psi_X(t) = \exp[\mu_1 i t - \frac{1}{2} \sigma_1^2 t^2] \quad (2.54)$$

Similarly the characteristic functions (c.f.) of $\bar{x}_j$ is given by

$$\begin{aligned}\psi_{\bar{x}_j}(t) &= \{\psi_x(\frac{t}{n_j})\}^{n_j} \\ &= \exp[\mu_j it - \frac{1}{2} \frac{\sigma_j^2 t^2}{n_j}] , \quad j = 1, 2 .\end{aligned}\quad (2.55)$$

Let $U = \bar{x}_2 - \bar{x}_1$

$$V = X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \quad (2.56)$$

then the joint characteristic function of U and V is given by

$$\begin{aligned}\psi_{UV}(t_1, t_2) &= E[e^{it_1 U + it_2 V}] \\ &= E[e^{it_1(\bar{x}_2 - \bar{x}_1) + it_2[X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)]]] \\ &= E[e^{(it_1 - \frac{it_2}{2})\bar{x}_2 + (-it_1 - \frac{it_2}{2})\bar{x}_1 + it_2 X}] \\ &\quad (2.57)\end{aligned}$$

Since $X, \bar{x}_1, \bar{x}_2$ are independent, this reduces to

$$\begin{aligned}\psi_{UV}(t_1, t_2) &= E[\exp\{(it_1 - \frac{it_2}{2})\bar{x}_2\}] E[\exp\{(-it_1 - \frac{it_2}{2})\bar{x}_1\}] \\ &\quad E[\exp(it_2 X)] \\ &= \psi_{\bar{x}_2}(t_1 - \frac{t_2}{2}) \psi_{\bar{x}_1}(-t_1 - \frac{t_2}{2}) \psi_X(t_2) .\end{aligned}\quad (2.58)$$

From equations [2.54 and 2.55], we have

$$\begin{aligned}\psi_{UV}(t_1, t_2) &= \exp[\mu_2 i(t_1 - \frac{t_2}{2}) - \frac{1}{2} \frac{\sigma_2^2}{n_2} (t_1 - \frac{t_2}{2})^2] \\ &\quad \cdot \exp[\mu_1 i(-t_1 - \frac{t_2}{2}) - \frac{1}{2} \frac{\sigma_1^2}{n_1} (-t_1 - \frac{t_2}{2})^2] \\ &\quad \cdot \exp[\mu_1 it_2 - \sigma_1^2 t_2^2] .\end{aligned}\quad (2.59)$$

Collecting terms involving t_1, t_2, t_1^2, t_2^2 and $t_1 t_2$:

$$\begin{aligned} \psi_{UV}(t_1 t_2) = & \exp[(\mu_2 - \mu_1)it_1 + (\mu_1 - \mu_2)\frac{it_2}{2} \\ & - \frac{t_1^2}{2}(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \\ & - \frac{t_2^2}{2}(\sigma_1^2 + \frac{\sigma_1^2}{4n_1} + \frac{\sigma_2^2}{4n_2}) - \frac{t_1 t_2}{2}(\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2})]. \end{aligned} \quad (2.60)$$

This shows that

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_2 - \mu_1 \\ \frac{\mu_1 - \mu_2}{2} \end{pmatrix}, \begin{pmatrix} (\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) & \frac{1}{2}(\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}) \\ \frac{1}{2}(\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}) & (\sigma_1^2 + \frac{\sigma_1^2}{4n_1} + \frac{\sigma_2^2}{4n_2}) \end{pmatrix} \right]. \quad (2.61)$$

The expected probability of misclassification can be evaluated directly from (2.61), as

$$E[\alpha_1(\hat{R}, f)] = \Pr[U > 0, V > 0] + \Pr[U < 0, V < 0]. \quad (2.62)$$

By standardizing U and V

$$\begin{aligned} E[\alpha_1(\hat{R}, f)] &= \Pr\left\{\frac{U - \mu_U}{\sigma_U} > -\frac{\mu_U}{\sigma_U}, \frac{V - \mu_V}{\sigma_V} > -\frac{\mu_V}{\sigma_V}\right\} \\ &+ \Pr\left\{\frac{U - \mu_U}{\sigma_U} < -\frac{\mu_U}{\sigma_U}, \frac{V - \mu_V}{\sigma_V} < -\frac{\mu_V}{\sigma_V}\right\} \\ &= \Pr\{Z_1 > -\frac{\mu_U}{\sigma_U}, Z_2 > -\frac{\mu_V}{\sigma_V}\} + \\ &\Pr\{Z_1 < -\frac{\mu_U}{\sigma_U}, Z_2 < -\frac{\mu_V}{\sigma_V}\} \\ &= G\{a_{11}, a_{21}; \rho\} + G\{a_{12}, a_{22}; \rho\} \end{aligned} \quad (2.63)$$

where

$$\begin{aligned}
 z_1 &= (U - \mu_u) / \sigma_u, \quad z_2 = (V - \mu_v) / \sigma_v, \\
 a_{11} &= - \frac{\mu_u}{\sigma_u} = - (\mu_2 - \mu_1) (n_1^{-1} \sigma_1^2 + n_2^{-1} \sigma_2^2)^{-\frac{1}{2}} = -a_{12}, \\
 a_{21} &= - \frac{\mu_v}{\sigma_v} = \frac{1}{2} (\mu_2 - \mu_1) (\sigma_1^2 + \frac{1}{4} [n_1^{-1} \sigma_1^2 + n_2^{-1} \sigma_2^2])^{-\frac{1}{2}} = -a_{22}
 \end{aligned}
 \tag{2.64}$$

and the correlation coefficient is given by

$$\begin{aligned}
 \rho &= \frac{1}{2} [n_1^{-1} \sigma_1^2 + n_2^{-1} \sigma_2^2]^{-\frac{1}{2}} [\sigma_1^2 + \frac{1}{4} n_1^{-1} \sigma_1^2 + \frac{1}{4} n_2^{-1} \sigma_2^2]^{-\frac{1}{2}} \\
 &\quad \cdot (n_1^{-1} \sigma_1^2 - n_2^{-1} \sigma_2^2) .
 \end{aligned}
 \tag{2.64}$$

Equation (2.63) is given in equations (19), (20), (21) of John [21]. A similar expression is obtained for $E[\alpha_2(\hat{R}, f)]$. An extension of this result is presented in section 4.6 when we consider the expected probability of misclassification in the Edgeworth Series distribution [ESD].

2.3 Distribution Theory of Probabilities of Misclassification

The distribution theory of the probabilities of misclassification has proved difficult due to the complicated nature of the forms of the probabilities themselves. Some results are available in John [21], Hills [18] and McLachlan [40], among others.

Consider the univariate normal situation when X in Π_1 is distributed as $N(\mu_1, \sigma_1^2)$ and X in Π_2 is distributed as $N(\mu_2, \sigma_2^2)$ and the classification regions are as given in equation (2.12). The distribution function of the conditional probability of misclassification (see equation (2.13)) is

given by John [21] as

$$\Pr\{\alpha_1(\hat{R}, f) < z\} = G(h_{11}, h_{21}; \rho) + G(h_{12}, h_{22}; \rho) \quad (2.65)$$

$$\left. \begin{aligned} \text{where } h_{11} &= a(\mu_2 - \mu_1) = -h_{12}, \\ h_{21} &= a[2\sigma_1 \Phi^{-1}(z) + \mu_2 - \mu_1], \\ h_{22} &= a[2\sigma_2 \Phi^{-1}(z) + \mu_1 - \mu_2], \\ a &= (n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)^{-\frac{1}{2}}, \\ \rho &= [n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2]^{-1}(n_2^{-1}\sigma_2^2 - n_1^{-1}\sigma_1^2). \end{aligned} \right\} \quad (2.66)$$

John's result assumes the variances in the two populations to be different although the discriminant function used is derived under conditions of equal variances in the two populations.

Hills [18] obtains the distribution of the estimated probability of misclassification by use of the same classification regions as those used by John. His result is that

$$\Pr\{\alpha_1[\hat{R}, \hat{f}] < z\} = \Phi(h_{21}) + \Phi(h_{22}) \quad (2.67)$$

where

$$\left. \begin{aligned} h_{21} &= \frac{1}{2\sqrt{\lambda}} [2\Phi^{-1}(z) + \delta], \quad \delta = \mu_2 - \mu_1 > 0 \\ h_{22} &= \frac{1}{2\sqrt{\lambda}} [2\Phi^{-1}(z) - \delta] \end{aligned} \right\} \quad (2.68a)$$

$$\left. \begin{aligned} \lambda &= \frac{1}{4} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \\ \Phi[\Phi^{-1}(z)] &\equiv z, \quad (0 \leq z \leq 1) \end{aligned} \right\} \quad (2.68b)$$

Hills assumes that the distribution of X in Π_1 is $N(\mu_1, 1)$ and that of X in Π_2 is $N(\mu_2, 1)$.

In the multivariate normal situation when the probability density functions of $\underline{X}$ in Π_1, Π_2 are given by equation (2.6), John [21] obtains the distribution of the conditional probability of misclassification using the classification regions given in equation (2.17). The distributions are given under different assumptions about $\underline{\mu}_1, \underline{\mu}_2$ and Σ in Π_1, Π_2 . The basic results are summarized below:

Unknown Parameters	Probability of Misclassification	Distribution
$\underline{\mu}_2$	$\alpha_1[\hat{R}, f]$	$\Pr\{\alpha_1[\hat{R}, f] < z\} = \Pr\{4n_2 y^2 > 4n_2 [\Phi^{-1}(z)]^2\}, (0 \leq z \leq \frac{1}{2})$ where $y = \frac{1}{2}[(\tilde{x}_2 - \underline{\mu}_1)\Sigma^{-1}(\tilde{x}_2 - \underline{\mu}_1)]^{\frac{1}{2}}$, $4n_2 y^2$ is a noncentral chi-square variable with p degrees of freedom and noncentrality parameter equal to $n_2 \Delta^2/2$. The distribution function is a monotonic function of Δ and therefore can be said to lie between certain bounds provided the upper and lower limits of Δ are known.

Unknown Parameters	Probability of Misclassification	Distribution
	$\alpha_2[\hat{R}, f]$	The exact distribution of $\alpha_2[\hat{R}, f]$ is complicated. The asymptotic distribution is given by $\Pr\{\alpha_2[\hat{R}, f] < z\} \sim \Phi[2n_2^{\frac{1}{2}}(\Phi^{-1}(z) + \frac{1}{2} \Delta)]$.
μ_1, μ_2	$\alpha_1[\hat{R}, f]$	The exact distribution of $\alpha_1[\hat{R}, f]$ is complicated. The asymptotic distribution is given by $\Pr\{\alpha_1[\hat{R}, f] < z\} \sim \Phi[2(n_1^{-1} + n_2^{-1})^{-\frac{1}{2}}\{\Phi^{-1}(z) + \frac{1}{2} \Delta\}]$.
	$\alpha_2[\hat{R}, f]$	Not considered.
μ_1, μ_2, Σ	$\alpha_1[\hat{R}, f]$	The exact distribution of $\alpha_1[\hat{R}, f]$ is complicated. The asymptotic distribution of $\alpha_1[\hat{R}, f]$ is $\Pr\{\alpha_1[\hat{R}, f] < z\} \sim \Phi(2[n_1^{-1} + n_2^{-1}]^{-\frac{1}{2}}[\Phi^{-1}(z) + \frac{1}{2} \Delta])$.
	$\alpha_2[\hat{R}, f]$	Not considered.

McLachlan [40] finds the asymptotic distribution of the conditional probability of misclassification as given by equation (2.18). An asymptotic expansion for the distribution function of $\alpha_1[\hat{R}, f]$ is found by an expansion of the characteristic

function by the Okamoto [46,47] technique. McLachlan finds that $\alpha_1[\hat{R}, f]$ is asymptotically normal if the terms of second and higher order with respect to $[n_1^{-1}, n_2^{-1}, (n_1 + n_2 - 2)^{-1}]$ are ignored.

2.4 Asymptotic Expansion of the Probabilities of Misclassification

Asymptotic expansion of the probability of misclassification using the W and Z classification statistics has received wide attention. In this connection one can refer to Okamoto [46, 47], Memon and Okamoto [42], Anderson [3,4], McLachlan [37, 38]. Some of the basic results of interest are stated here when the parameters of the distribution in Π_1 and Π_2 are unknown and are estimated from samples.

By use of the characteristic function Okamoto [46, 47] obtains an asymptotic expansion of the distribution of the W statistic. From the asymptotic expansion of the distribution of the W statistic he obtains the asymptotic expansion of the probabilities of misclassification as functions of Δ , n_1 , n_2 and p . The expansion is up to terms of the order n_1^{-2} , n_2^{-2} , $(n_1 n_2)^{-1}$, $(n_1 n)^{-1}$, $(n_2 n)^{-1}$ and n^{-2} where $n = n_1 + n_2 - 2$. The probabilities of misclassification are given by

$$P_r[W < 0 | \underline{X} \in \Pi_1] = \Phi\left(-\frac{\Delta}{2}\right) + \frac{a_1}{n_1} + \frac{a_2}{n_2} + \frac{a_3}{n} + \frac{b_{11}}{n_1^2} + \frac{b_{22}}{n_2^2} + \frac{b_{12}}{n_1 n_2} + \frac{b_{13}}{n_1 n} + \frac{b_{23}}{n_2 n} + \frac{b_{33}}{n^2} + o_3$$

(2.69)

$$\begin{aligned} \Pr[W > 0 | \underline{X} \in \Pi_2] = & \Phi\left(-\frac{\Delta}{2}\right) + \frac{a_1}{n_2} + \frac{a_2}{n_1} + \frac{a_3}{n} + \frac{b_{22}}{n_1^2} + \frac{b_{33}}{n^2} + \\ & + \frac{b_{11}}{n_2^2} + \frac{b_{12}}{n_1 n_2} + \frac{b_{23}}{n_1 n} + \frac{b_{13}}{n_2 n} + o_3 \end{aligned} \quad (2.70)$$

where a's and b's are expressions involving the parameters of the two populations.

Using the technique of Okamoto [46], Memon and Okamoto [42], obtain asymptotic expansion of the probabilities of misclassification when the statistic Z is used. Their expansion is up to third order with respect to $(n_1^{-1}, n_2^{-1}, n^{-1})$ and the probabilities of misclassification are given by

$$\begin{aligned} \Pr[Z > 0 | \Pi_1] = & \Phi\left(-\frac{\Delta}{2}\right) + \frac{a_1}{n_1} + \frac{a_2}{n_2} + \frac{a_3}{n} + \frac{b_{11}}{n_1^2} + \frac{b_{22}}{n_2^2} + \\ & + \frac{b_{23}}{n_2 n} + \frac{b_{12}}{n_1 n_2} + \frac{b_{13}}{n_1 n} + \frac{b_{33}}{n^2} + o_3 \end{aligned} \quad (2.71)$$

$$\begin{aligned} \Pr[Z < 0 | \Pi_2] = & \Phi\left(-\frac{\Delta}{2}\right) + \frac{a_2}{n_1} + \frac{a_1}{n_2} + \frac{a_3}{n} + \frac{b_{22}}{n_1^2} + \frac{b_{11}}{n_2^2} + \\ & + \frac{b_{12}}{n_1 n_2} + \frac{b_{23}}{n_1 n} + \frac{b_{13}}{n_2 n} + \frac{b_{33}}{n^2} + o_3 \end{aligned} \quad (2.72)$$

where a's and b's are expressions involving the parameters of the two populations. Exact forms can be found in equations (3.7), (3.8) of Memon and Okamoto [42].

Anderson [3,4] obtains an asymptotic expansion of the distribution of the 'studentized W statistic', where the studentized W statistic is defined as W less the estimate of its limiting mean divided by the estimate of its limiting standard deviation. Anderson's expansion is a Taylor series

expansion. The asymptotic probabilities of misclassification are given by

$$\Pr\left\{\frac{W - \frac{1}{2} \hat{\Delta}}{\hat{\Delta}^{1/2}} \leq u\right\} = \Phi(u) + \frac{1}{n} \phi(u) \left[\frac{p-1}{\hat{\Delta}^{1/2}} (1+k) - \left(p - \frac{1}{4} + \frac{k}{2}\right)u - \frac{1}{4} u^3 \right] + o(n^{-2}), \quad (2.73)$$

when $E(\underline{X}) = \underline{\mu}_1$. Interchanging n_1, n_2 gives

$$\Pr\left\{\frac{W - \frac{1}{2} \hat{\Delta}}{\hat{\Delta}^{1/2}} \leq v\right\} = \Phi(v) - \frac{1}{n} \phi(v) \left[\frac{p-1}{\hat{\Delta}^{1/2}} \left(1 + \frac{1}{k}\right) + \left(p - \frac{1}{4} + \frac{1}{2k}\right)v + \frac{1}{4} v^3 \right] + o(n^{-2}) \quad (2.74)$$

when $E(\underline{X}) = \underline{\mu}_2$.

In (2.73) and (2.74) $k = \text{limit of } n_2/n_1$ when both n_1, n_2 tend to infinity.

For the W statistic, McLachlan [40] obtains an asymptotic expansion of $E[\Phi(-\frac{\hat{\Delta}}{2})]$ over the joint sampling distribution of $\tilde{X}_1, \tilde{X}_2$ and S . The expansion is given by

$$E[\Phi(-\frac{\hat{\Delta}}{2})] = \beta_0 + \beta_1 + \beta_3 + 0_3 \quad (2.75)$$

In equation (2.75) we have used the notation $n = n_1 + n_2 - 2$,

$$\begin{aligned} \beta_0 &= \Phi[-\Delta/2], \\ \beta_1 &= \frac{b_1}{n_1} + \frac{b_2}{n_2} + \frac{b_3}{n_3}, \\ \beta_2 &= \frac{b_{11}}{n_1^2} + \frac{b_{22}}{n_2^2} + \frac{b_{12}}{n_1 n_2} + \frac{b_{13}}{n_1 n} + \frac{b_{23}}{n_2 n} + \frac{b_{33}}{n^2} \end{aligned} \quad (2.76)$$

and 0_3 are terms of third order with respect to $(n_1^{-1}, n_2^{-1}, n^{-1})$.

The expressions for b_i, b_{ij} are given by

$$b_1 = b_2 = (g/16) [\Delta - \frac{4(p-1)}{\Delta}] ,$$

$$b_3 = (\Delta g/32) [\Delta^2 - 4(2p+1)] ,$$

$$b_{11} = b_{22} = \frac{1}{2} b_{12}$$

$$= (g/1024) [\Delta (\Delta^2 - 4[2p+1])] +$$

$$\frac{16(p-1)}{\Delta} \{ (p-1) + \frac{4(p-3)}{\Delta^2} \} ,$$

$$b_{13} = b_{23} = (g/1024) [\Delta^5 - 4(3p+7)\Delta^3 + 16(2p^2 + 8p + 5)\Delta - \frac{64(p-1)(2p+1)}{\Delta}] ,$$

$$b_{33} = (\Delta g/12,288) [3\Delta^6 - 4(12p+35)\Delta^4 + 16(12p^2 + 72p + 71)\Delta^2 - 192(12p^2 + 12p + 1)] ,$$

and $g = (2\pi)^{-\frac{1}{2}} \exp(-\Delta^2/8) .$

McLachlan's results confirm the well known result that $\Phi[-\hat{\Delta}/2]$ gives too favourable an estimate of the true probability of misclassification. McLachlan [37] also determines expressions for the asymptotic variances of the conditional probabilities of misclassification.

Siotani [54] extends the asymptotic expansions of the probabilities of misclassification when the statistics W and Z are used for classification purposes to more terms. He also compares the performance of the two procedures.

CHAPTER 3

DISCRIMINATION FOR NON-NORMAL DISTRIBUTIONS:

JOHNSON'S SYSTEM OF FREQUENCY CURVES

3.1 Introduction: Robustness of the Linear Discriminant Function (LDF)

The Linear Discriminant Function is used as an assignment rule under the following assumptions:

- (i) The density functions of observations from Π_i ($i = 1, 2$) are multivariate normal; $\Pi_1: N_p(\underline{\mu}_1, \Sigma)$, $\Pi_2: N_p(\underline{\mu}_2, \Sigma)$.
- (ii) The variance covariance matrix in Π_1 is the same as that in Π_2 .
- (iii) The a priori probabilities, p_i ($i = 1, 2$) of an observation coming from Π_i ($i = 1, 2$) respectively, are known.
- (iv) The parameters of the density functions in (i), namely $\underline{\mu}_1, \underline{\mu}_2, \Sigma$ are known.

If the above four assumptions hold the LDF gives the optimal assignment rule in that we cannot improve upon it and the errors of misclassification are minimized. However it would be of concern to determine the effects on the procedures using LDF when some or all of the above assumptions do not hold. A study of this type is referred to as the robustness study.

If the parameters $\underline{\mu}_1, \underline{\mu}_2, \Sigma$ are unknown they are usually estimated from samples. Two problems may however arise at the estimation stage, (a) we may have missing values in the data and (b) the initial sample may not be properly assigned due to clerical errors or inaccuracy in the initial assignment.

Lachenbruch [28] has studied the effects on the errors of misclassification due to initial misclassification, when the assignment rule is based on the W criterion. Two assumptions are made in Lachenbruch's study, namely (i) the observations are randomly misclassified and (ii) each observation has the same chance of initial misclassification. He considers a proportion q_1 of n_1 observations from Π_1 that belong to Π_2 and a proportion q_2 of n_2 observations from Π_2 that belong to Π_1 .

He shows that, when sample sizes are large the coefficients of the discriminant function based on misclassified samples differ from the true coefficients by a constant multiple. He also shows that if $q_1 = q_2$ the optimum errors of misclassification are unaffected; in particular if the total proportion misclassified is not large and the same in both populations then there is no increase in the probability of misclassification. When sample sizes are small Lachenbruch conducts sampling experiments. He computes the conditional probability of misclassification. His conclusions from the sampling study are that the large sample results hold to a good approximation.

McLachlan [38] obtains the asymptotic expansion of the expected error under the condition that $q_1 = 0$. His expansion has an error term of second order with respect to $[n_1^{-1}, n_2^{-1}, (n_1 + n_2 - 2)^{-1}]$. Under the assumptions similar to Lachenbruch's he concludes from his study that the results obtained by Lachenbruch for large sample sizes are valid in most cases.

Lachenbruch [29] states that the assumptions used by Lachenbruch and McLachlan in their previous studies are unrealistic. He decides therefore to study the effects of misclassification under what he calls, hopefully, realistic models. The two models are called (i) the complete separation model and (ii) a generalization of the complete separation model. He conducts sampling experiments and concludes from the experimental results that the LDF is only slightly affected by initial misclassification of the samples in the non-random manner.

If the a priori probabilities are unknown, Lachenbruch [30, p. 40] states that these are usually estimated from a sample if the sample is taken from the whole population rather than two separate samples taken from two subgroups. Otherwise a minimax approach to the problem is adopted.

When the assumption of equal covariance matrices in the two populations does not hold, then the optimal classification rule is based on the QDF. Gilbert [16] studies the effect of using the QDF rule when the variance matrix $\Sigma_1 = I$ (unit matrix) and $\Sigma_2 = dI$, where d is a constant. The performance of the QDF rule instead of the LDF rule is studied by varying values of d , Δ^2 and p_1 (a priori probability of an observation coming from Π_1). She assumes the parameters known and hence compares the performances of the two rules using the optimum error rate. The conclusions of Gilbert are that if d is far from unity the LDF is poorer than the QDF while,

when Δ^2 is large, the QDF and LDF behave similarly.

Marks and Dunn [35] generalize Gilbert's work. They compare the effect of the W criterion, BLDF and QDF rules on errors of misclassification. In their case the parameters of the distributions are assumed unknown. They conduct sampling experiments which indicate that, for large sample sizes from the normal distribution, the QDF is better than the LDF. For small sample sizes the QDF is poorer than the LDF. Their conclusions are based on the asymptotic ratios of the errors of misclassification.

Lachenbruch, Sneeringer and Revo [32] study the robustness of the LDF and QDF with respect to certain types of non-normality. They consider Johnson's system of distributions which are transformable to normality. They consider three members of this family:

- (i) S_L : Log normal transformation $y_i = \log(x_i)$, $0 < x_i < \infty$
- (ii) S_B : Logit normal transformation $y_i = \log\left(\frac{x_i}{1-x_i}\right)$, $0 < x_i < 1$
- (iii) S_u : Inverse hyperbolic sine normal transformation

$$y_i = \sinh^{-1}(x_i), -\infty < x_i < \infty.$$

In each case y_i are normally distributed.

On the assumptions that the two populations have different means and the covariance matrix of $\underline{Y}$ corresponding to the $\underline{X}$ is the identity matrix, sampling studies are conducted in order to examine the behaviour of the errors of misclassification. They find that the total error of misclassification is greatly increased as are individual errors distorted for all transformations in

the case of the LDF. The errors are smallest in the case of the Logit normal transformation. The errors are estimated by the resubstitution and Leave-one-out methods which are considered by Lachenbruch and Mickey [31]. Approximate minimax rules are investigated because of the distortion in the errors and are found to reduce the errors of misclassification greatly.

Because of the non-linear transformations, the $\underline{X}$'s do not have the same covariance matrices in the two populations. They investigate the performance of the QDF and find that it performs poorly. Some theoretical results are also obtained.

Our study is based on the three transformations considered by Lachenbruch, Sneeringer and Revo. We study the univariate analogue of the three distributions. Before we present the details of our investigation we give a general result in section 3.2. We then discuss each of the S_L , S_B and S_U distributions under the headings

- (i) Distribution Theory and Rules,
- (ii) Distribution of the Conditional probability of misclassification,
- (iii) Unconditional probability of misclassification.

Before proceeding to examine specific cases, it is of importance to recall the rules of classification as well as the associated errors of misclassification, assuming normality. Let X be an observation from Π_1 or Π_2 and the density function of X in Π_1 given by

$$f(x) = (\sigma_x^2 2\pi)^{-1/2} \exp[-(x - \mu_1)^2 / 2\sigma_x^2], -\infty < x < \infty \quad (3.1)$$

while that of X in Π_2 is given by

$$f(x) = (\sigma_x^2 2\pi)^{-1/2} \exp[-(x - \mu_2)^2 / 2\sigma_x^2], -\infty < x < \infty. \quad (3.2)$$

The parameters μ_1, μ_2, σ_x satisfy the conditions $-\infty < \mu_1, \mu_2 < \infty$ and $\sigma_x > 0$.

The classification rule based on the W discriminant function may be written as follows: For $\bar{x}_1 < \bar{x}_2$

$$\left. \begin{array}{l} \text{classify } X \text{ as belonging to } \Pi_1 \text{ if } X < \frac{\bar{x}_1 + \bar{x}_2}{2} \\ \Pi_2 \text{ if } X \geq \frac{\bar{x}_1 + \bar{x}_2}{2} \end{array} \right\} \quad (3.3)$$

and for $\bar{x}_1 \geq \bar{x}_2$

$$\left. \begin{array}{l} \text{classify } X \text{ as belonging to } \Pi_1 \text{ if } X > \frac{\bar{x}_1 + \bar{x}_2}{2} \\ \Pi_2 \text{ if } X \leq \frac{\bar{x}_1 + \bar{x}_2}{2} \end{array} \right\} \quad (3.4)$$

where $\bar{x}_1, \bar{x}_2$ are sample estimates of μ_1, μ_2 respectively.

Two distinct probabilities of misclassification are relevant here, the conditional (on $\bar{x}_1, \bar{x}_2$) probability of misclassification and the unconditional probability of misclassification. Define $e_{12}(\bar{x}_1, \bar{x}_2)$ to be the conditional probability of assigning an individual from Π_1 to Π_2 and

$e_{21}(\bar{x}_1, \bar{x}_2)$ to be the conditional probability of assigning an individual from Π_2 to Π_1 . Then

$$\left. \begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= \Pr\left\{X \geq \frac{\bar{x}_1 + \bar{x}_2}{2} \mid \bar{x}_1, \bar{x}_2, X \in \Pi_1\right\} \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Pr\left\{X \leq \frac{\bar{x}_1 + \bar{x}_2}{2} \mid \bar{x}_1, \bar{x}_2, X \in \Pi_1\right\} \text{when } \bar{x}_1 \geq \bar{x}_2 \end{aligned} \right\} \quad (3.5)$$

and

$$\left. \begin{aligned} e_{21}(\bar{x}_1, \bar{x}_2) &= \Pr\left\{X < \frac{\bar{x}_1 + \bar{x}_2}{2} \mid \bar{x}_1, \bar{x}_2, X \in \Pi_2\right\} \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Pr\left\{X > \frac{\bar{x}_1 + \bar{x}_2}{2} \mid \bar{x}_1, \bar{x}_2, X \in \Pi_2\right\} \text{when } \bar{x}_1 \geq \bar{x}_2 \end{aligned} \right\} \quad (3.6)$$

The unconditional probability of misclassification is, by definition, the expectation, with respect to $\bar{x}_1, \bar{x}_2$ of the conditional probability of misclassification. It is sometimes called the expected probability of misclassification. Following the notation introduced above the expected probability of misclassification is given by

$$\begin{aligned} E[e_{12}(\bar{x}_1, \bar{x}_2)] &= \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \geq 0 \mid X \in \Pi_1\} \\ &\quad + \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \leq 0 \mid X \in \Pi_1\} \end{aligned} \quad (3.7)$$

and

$$\begin{aligned} E[e_{21}(\bar{x}_1, \bar{x}_2)] &= \Pr\{(\bar{x}_1 - \bar{x}_2) < 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) < 0 \mid X \in \Pi_2\} \\ &\quad + \Pr\{(\bar{x}_1 - \bar{x}_2) \geq 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) > 0 \mid X \in \Pi_2\}. \end{aligned} \quad (3.8)$$

3.2 A General Result

A general result on the evaluation of BVN probabilities which is repeatedly used in the following sections is developed here.

Let

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right]$$

then, for any constant K ,

$$\Pr\{X_2 - X_1 > 0, X_1 + X_2 > K\} = p \quad (3.9)$$

can be evaluated as follows.

Let $U = X_2 - X_1$ and $V = X_1 + X_2$, then

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right]$$

where

$$\left. \begin{aligned} \mu_u &= \mu_2 - \mu_1 \\ \mu_v &= \mu_2 + \mu_1 \\ \sigma_u^2 &= (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) \\ \sigma_v^2 &= (\sigma_1^2 + \sigma_2^2 + 2\sigma_{12}) \\ \sigma_{uv} &= (\sigma_2^2 - \sigma_1^2) \end{aligned} \right\} \quad (3.10)$$

and

Therefore

$$\begin{aligned}
p &= \Pr\{U > 0, V > K\} \\
&= \Pr\left\{\frac{U - \mu_u}{\sigma_u} > -\frac{\mu_u}{\sigma_u}, \frac{V - \mu_v}{\sigma_v} > \frac{K - \mu_v}{\sigma_v}\right\} \\
&= \Pr\left\{Z_1 > -\frac{\mu_u}{\sigma_u}, Z_2 > \frac{K - \mu_v}{\sigma_v}\right\} \quad (3.11)
\end{aligned}$$

Where

$$\begin{aligned}
\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} &\sim \text{BVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix} \right], \\
\rho^* &= \frac{\sigma_{uv}}{\sigma_u \sigma_v} = (\sigma_2^2 - \sigma_1^2) / \{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) \\
&\quad \cdot (\sigma_1^2 + \sigma_2^2 + 2\sigma_{12})\}^{\frac{1}{2}} \\
&= \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) / \{(\sigma_1^2 + \sigma_2^2)^2 - 4\sigma_{12}^2\} / \sigma_1^4 \sigma_2^4\}^{\frac{1}{2}} \\
&= (\sigma_1^{-2} - \sigma_2^{-2}) / \{(\sigma_1^{-2} + \sigma_2^{-2}) - 4\rho^2 / \sigma_1^2 \sigma_2^2\}^{\frac{1}{2}}
\end{aligned}$$

where

(3.12)

$$\rho = \sigma_{12} / \sigma_1 \sigma_2.$$

Writing

$$H_\rho(z_1, z_2) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} f(u_1, u_2; \rho) du_1 du_2$$

for

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \text{Standard Bivariate Normal},$$

we have, since $H_\rho(z_1, z_2) = \Pr(Z_1 > -z_1, Z_2 > -z_2)$,

$$p = H_{\rho} * \left[\frac{\mu_u}{\sigma_u}, \frac{\mu_v - k}{\sigma_v} \right]. \quad (3.13)$$

In a similar way

$$\Pr\{X_2 - X_1 < 0, X_1 + X_2 < K'\} = p' \quad (3.14)$$

is also given by

$$\begin{aligned} p' &= \Pr\{U < 0, V < K'\} \\ &= \Pr\left\{\frac{U - \mu_u}{\sigma_u} < -\frac{\mu_u}{\sigma_u}, \frac{V - \mu_v}{\sigma_v} < \frac{K' - \mu_v}{\sigma_v}\right\} \\ &= \Pr\left\{Z_1 < -\frac{\mu_u}{\sigma_u}, Z_2 < \frac{K' - \mu_v}{\sigma_v}\right\} \\ &= H_{\rho} * \left[-\frac{\mu_u}{\sigma_u}, \frac{K' - \mu_v}{\sigma_v}\right], \end{aligned} \quad (3.15)$$

where $\mu_u, \mu_v; \sigma_u, \sigma_v$ are as defined by (3.10).

3.3 Log Normal Transformation (S_L)

Definition: We define the random variable X to be log normal if $Y = \ln X$ is normal, $0 < X < \infty$.

a) Distribution Theory and Rules

The density function of Y is given by

$$f(y) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(y - \mu_y)^2\right], -\infty < y < \infty. \quad (3.16)$$

The parameters μ_y, σ_y satisfying the conditions $-\infty < \mu_y < \infty, \sigma_y > 0$. By the transformation

$$Y = \ln X \quad (3.17)$$

the density function of X is given by

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln x - \mu_Y)^2\right], \quad 0 < x < \infty. \quad (3.18)$$

We assume that σ_Y^2 is common to Π_1, Π_2 and for

$$X \in \Pi_1 \quad \mu_Y = \mu_{1Y}$$

$$X \in \Pi_2 \quad \mu_Y = \mu_{2Y}$$

Let

$$\delta = \frac{\bar{x}_1 + \bar{x}_2}{2} \quad (3.19)$$

then from (3.5) and (3.18) the conditional probability of misclassification, when an observation from Π_1 is misclassified, is given by

$$e_{12}(\bar{x}_1, \bar{x}_2) = \int_{\delta}^{\infty} \frac{1}{\sigma_Y x \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln x - \mu_{1Y})^2\right] dx$$

if $\bar{x}_1 < \bar{x}_2$

and

$$= \int_0^{\delta} \frac{1}{\sigma_Y x \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln x - \mu_{1Y})^2\right] dx$$

if $\bar{x}_1 \geq \bar{x}_2$.

By the simple transformation

$$t = \frac{\ln x - \mu_{1Y}}{\sigma_Y} \quad (3.20)$$

we obtain

$$e_{12}(\bar{x}_1, \bar{x}_2) = \int_{\frac{\ln \delta - \mu_{1Y}}{\sigma_Y}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt \quad \text{if } \bar{x}_1 < \bar{x}_2$$

and

$$= \int_{-\infty}^{\frac{\ln \delta - \mu_{1Y}}{\sigma_Y}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt \quad \text{if } \bar{x}_1 \geq \bar{x}_2.$$

Finally

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= 1 - \Phi\left[\frac{\ln \delta - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\ln \delta - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned}$$

Substituting for δ from (3.19) we have

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= 1 - \Phi\left[\frac{\ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned}$$

(3.21)

Following the same procedure the conditional probability of misclassification $e_{21}(\bar{x}_1, \bar{x}_2)$ is seen to be

$$\begin{aligned} e_{21}(\bar{x}_1, \bar{x}_2) &= \Phi\left[\frac{\ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{2Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= 1 - \Phi\left[\frac{\ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{2Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned}$$

(3.22)

Note that, for

$$\begin{aligned} X \in \Pi_1 \quad E(X) &= \exp\left[\mu_{1Y} + \frac{1}{2} \sigma_Y^2\right] = \omega^{\frac{1}{2}} \exp(\mu_{1Y}) \\ V(X) &= \omega(\omega - 1) \exp(2\mu_{1Y}) \end{aligned} \tag{3.23}$$

and for

$$X \in \Pi_2 \quad E(X) = \omega^{\frac{1}{2}} \exp(\mu_{2Y}) \quad (3.24)$$

$$V(X) = \omega(\omega - 1) \exp(2\mu_{2Y})$$

where $\omega = \exp(\sigma_Y^2)$ (3.25)

Since the numbering of the populations is arbitrary, the problem treated here is symmetric, and hence we consider the properties of $e_{12}(\bar{x}_1, \bar{x}_2)$ only.

b) Distribution of the Conditional Probability of Misclassification.

From (3.21) we notice that

$$\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \text{ iff } \{\bar{x}_1 < \bar{x}_2 \text{ and } \ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) > \mu_{1Y} - \sigma_Y \phi^{-1}(z)\}$$

or

$$\{\bar{x}_1 \geq \bar{x}_2 \text{ and } \ln\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) < \mu_{1Y} + \sigma_Y \phi^{-1}(z)\}, \quad (3.26)$$

hence

$$\begin{aligned} \Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} &= \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, \\ &(\bar{x}_1, \bar{x}_2) \geq 2\exp[-\sigma_Y \phi^{-1}(z) + \mu_{1Y}]\} + \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, \\ &(\bar{x}_1 + \bar{x}_2) < 2\exp[\mu_{1Y} + \sigma_Y \phi^{-1}(z)]\}. \end{aligned} \quad (3.27)$$

Now the moment generating function (m.g.f) of $\{(\bar{x}_2 - \bar{x}_1), (\bar{x}_1 + \bar{x}_2)\}$ is given by

$$\begin{aligned}
M(t_1, t_2) &= E[\exp\{t_1(\bar{x}_2 - \bar{x}_1) + t_2(\bar{x}_1 + \bar{x}_2)\}] \\
&= E[\exp\{(t_1 + t_2)\bar{x}_2 + (t_2 - t_1)\bar{x}_1\}]
\end{aligned}$$

and since $\bar{x}_1, \bar{x}_2$ are independent

$$M(t_1, t_2) = M_{\bar{x}_1}^-(t_2 - t_1) M_{\bar{x}_2}^-(t_1 + t_2), \quad (3.28)$$

Although we do not know the m.g.f's of $\bar{x}_1, \bar{x}_2$ we approximate these by the normal m.g.f's using the Central Limit Theorem. We therefore write, for large n_1, n_2 :

$$\begin{aligned}
M(t_1, t_2) &= M_{\bar{x}_1}^-(t_1 - t_2) M_{\bar{x}_2}^-(t_1 + t_2) \\
&\approx \exp[\mu_2(t_1 + t_2) + \frac{1}{2}\{(t_1 + t_2)^2 \frac{\sigma_2^2}{n_2}\} \\
&\quad + (t_2 - t_1)\mu_1 + \frac{1}{2}\{(t_2 - t_1)^2 \frac{\sigma_1^2}{n_1}\}] \quad (3.29)
\end{aligned}$$

Here we assume

$$\bar{x}_i \approx N(\mu_i, \frac{\sigma_i^2}{n_i}) \quad i = 1, 2$$

where

$$\mu_1 = E(\bar{X}) = E(X) \quad \text{when } X \in \Pi_1$$

$$\sigma_1^2 = V(X) \quad \text{when } X \in \Pi_1$$

similarly

$$\mu_2 = E(\bar{X}) = E(X) \quad \text{when } X \in \Pi_2$$

$$\sigma_2^2 = V(X) \quad \text{when } X \in \Pi_2$$

where $E(X)$ and $V(X)$, when $X \in \Pi_1$ or Π_2 , are

given in equations (3.23) and (3.24). Simplifying (3.29) we obtain

$$M(t_1, t_2) \approx \exp[t_1(\mu_2 - \mu_1) + t_2(\mu_1 + \mu_2) + \frac{t_1^2}{2}(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) + \frac{t_2^2}{2}(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) + \frac{2t_1 t_2}{2}(\frac{\sigma_2^2}{n_2} - \frac{\sigma_1^2}{n_1})] \quad (3.30)$$

Thus if we write

$$U = \bar{x}_2 - \bar{x}_1, \quad V = \bar{x}_1 + \bar{x}_2,$$

then

$$\begin{pmatrix} U \\ V \end{pmatrix} \approx \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right],$$

where

$$\mu_u = \mu_2 - \mu_1,$$

$$\mu_v = \mu_1 + \mu_2,$$

$$\sigma_u^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

$$\sigma_v^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2},$$

$$\sigma_{uv} = \frac{\sigma_2^2}{n_2} - \frac{\sigma_1^2}{n_1},$$

where

$$\mu_1 = E(\bar{X}_1) = E(X) \quad \text{when } X \in \Pi_1,$$

$$\mu_2 = E(\bar{X}_2) = E(X) \quad \text{when } X \in \Pi_2,$$

also

$$\sigma_1^2 = V(X) \quad \text{when } X \in \Pi_1,$$

$$\sigma_2^2 = V(X) \quad \text{when } X \in \Pi_2.$$

Therefore

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx \Pr\{U > 0, V > k\} + \Pr\{U \leq 0, V \leq k'\} \quad (3.31)$$

where

$$k = 2\exp[-\sigma_Y \Phi^{-1}(z) + \mu_{1Y}]$$

$$k' = 2\exp[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)]$$

Recalling the result of equations (3.13) and (3.15) we have

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx H_{\rho*}\left[\frac{\mu_U}{\sigma_U}, \frac{\mu_V - k}{\sigma_V}\right] + H_{\rho*}\left[-\frac{\mu_U}{\sigma_U}, \frac{k' - \mu_V}{\sigma_V}\right]. \quad (3.32)$$

The parameters on the right hand side of (3.32) are given by

$$\begin{aligned} \mu_U = \mu_2 - \mu_1 &= \exp(\mu_{2Y} + \frac{\sigma_Y^2}{2}) - \exp(\mu_{1Y} + \frac{\sigma_Y^2}{2}) \\ &= \omega^{\frac{1}{2}}[\exp(\mu_{2Y}) - \exp(\mu_{1Y})], \end{aligned}$$

$$\mu_V = \omega^{\frac{1}{2}}[\exp(\mu_{2Y}) + \exp(\mu_{1Y})],$$

$$\sigma_U^2 = \omega(\omega - 1)\left\{\frac{\exp(2\mu_{1Y})}{n_1} + \frac{\exp(2\mu_{2Y})}{n_2}\right\},$$

$$\sigma_V^2 = \omega(\omega - 1)\left\{\frac{\exp(2\mu_{1Y})}{n_1} + \frac{\exp(2\mu_{2Y})}{n_2}\right\},$$

$$\sigma_{UV} = \omega(\omega - 1)\left\{\frac{\exp(2\mu_{2Y})}{n_2} - \frac{\exp(2\mu_{1Y})}{n_1}\right\}.$$

The distribution function of $e_{12}(\bar{x}_1, \bar{x}_2)$ can be written, as follows

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx G[h_{11}, h_{21}; \rho^*] + G[h_{12}, h_{22}; \rho^*] \quad (3.33)$$

with

$$h_{11} = \frac{\sqrt{\omega} [\exp(\mu_{2Y}) - \exp(\mu_{1Y})]}{\{\omega(\omega - 1) [\frac{\exp(2\mu_{2Y})}{n_2} + \frac{\exp(2\mu_{1Y})}{n_1}]\}}$$

$$= -h_{12},$$

$$h_{21} = \frac{\sqrt{\omega} [\exp(\mu_{2Y}) + \exp(\mu_{1Y}) - 2\exp[-\sigma_Y \Phi^{-1}(z) + \mu_{1Y}]]}{\{\omega(\omega - 1) [\frac{\exp(2\mu_{2Y})}{n_2} + \frac{\exp(2\mu_{1Y})}{n_1}]\}^{\frac{1}{2}}},$$

$$h_{22} = \frac{2\exp[\sigma_Y \Phi^{-1}(z) + \mu_{1Y}] - \sqrt{\omega} [\exp(\mu_{2Y}) + \exp(\mu_{1Y})]}{\{\omega(\omega - 1) [\frac{\exp(2\mu_{2Y})}{n_2} + \frac{\exp(2\mu_{1Y})}{n_1}]\}^{\frac{1}{2}}},$$

$$\rho^* = \left\{ \frac{\frac{\exp(2\mu_{2Y})}{n_2} - \frac{\exp(2\mu_{1Y})}{n_1}}{\frac{\exp(2\mu_{2Y})}{n_2} + \frac{\exp(2\mu_{1Y})}{n_1}} \right\}$$

with

$$G(z_1, z_2; \rho^*) = H_{\rho^*}(z_1, z_2) \quad .$$

Representing this distribution function as

$$R(z) = \int_{-\infty}^{h_{21}(z)} \int_{-\infty}^{h_{11}} f(z_1, z_2) dz_1 dz_2 + \int_{-\infty}^{h_{22}(z)} \int_{-\infty}^{-h_{11}} f(z_1, z_2) dz_1 dz_2 \quad (3.34)$$

the probability density function of $e_{12}(\bar{x}_1, \bar{x}_2)$ is given by

$$r(z) = \int_{-\infty}^{h_{11}} f[z_1, h_{21}(z)] h'_{21}(z) dz_1 + \int_{-\infty}^{-h_{11}} f[z_1, h_{22}(z)] h'_{22}(z) dz_1, \quad (3.35)$$

which is a very complicated form. Further study is proposed for examining this probability density function.

c) Unconditional Probability of Misclassification

From (3.7) or (3.8) we notice that the unconditional probability of misclassification is usually obtained conditional on a future observation coming from Π_1 or Π_2 . We obtain this for $e_{12}(\bar{x}_1, \bar{x}_2)$ by observing that a wrong assignment of an individual from Π_1 corresponds to the occurrence of either of the following two events;

$$\bar{x}_1 < \bar{x}_2 \quad \text{and} \quad X \geq \frac{\bar{x}_1 + \bar{x}_2}{2}$$

or

$$\bar{x}_1 \geq \bar{x}_2 \quad \text{and} \quad X \leq \frac{\bar{x}_1 + \bar{x}_2}{2}.$$

Therefore

$$\begin{aligned} E[e_{12}(\bar{x}_1, \bar{x}_2)] &= \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \geq 0 | X \in \Pi_1\} \\ &\quad + \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \leq 0 | X \in \Pi_1\}. \end{aligned} \quad (3.36)$$

The right hand side of (3.36) is difficult to obtain directly in its exact form.

The distribution of $\bar{x}_1$, $\bar{x}_2$ and X are independent with X having the p.d.f. $g(x)$ given by

$$g(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln x - \mu_{1Y})^2\right],$$

and by the Central Limit Theorem,

$$\begin{aligned}\bar{x}_1 &\approx N(\mu_1, \frac{\sigma_1^2}{n_1}) \\ \bar{x}_2 &\approx N(\mu_2, \frac{\sigma_2^2}{n_2}).\end{aligned}$$

Recall the well known result on conditional expectation that if (X,Y) is a two dimensional random variable then

$$E(X) = E[E\{X|Y\}] \quad (3.37)$$

The right hand side of (3.36) can be determined by using this result. Thus the first term on the right hand side of (3.36) is given by

$$\begin{aligned}\Pr\{\bar{x}_2 - \bar{x}_1 > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \geq 0\} \\ = \int_0^\infty \Pr\{\bar{x}_2 - \bar{x}_1 > 0, \bar{x}_1 + \bar{x}_2 \leq 2x|x\}g(x)dx\end{aligned} \quad (3.38)$$

but the probability in the integrand has been approximated earlier [equation (3.32)] in a form involving the BVN distribution function. Writing this probability as $p(x)$,

$$p(x) = \Pr\{U > 0, V < 2x\}, \quad (3.39)$$

where

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right],$$

we have

$$\begin{aligned}
p(x) &= \Pr\{Z_1 > -\frac{\mu_u}{\sigma_u}, Z_2 < \frac{2x - \mu_v}{\sigma_v}\} \\
&= \int_{-\frac{\mu_u}{\sigma_u}}^{\infty} \int_{-\infty}^{\frac{2x - \mu_v}{\sigma_v}} h_{\rho}(z_1, z_2) dz_2 dz_1 \\
&= \int_{-\infty}^{\frac{\mu_u}{\sigma_u}} \int_{-\infty}^{\frac{2x - \mu_v}{\sigma_v}} h_{-\rho}(z_1, z_2) dz_1 dz_2 \\
&= H_{-\rho}\left[\frac{\mu_u}{\sigma_u}, \frac{2x - \mu_v}{\sigma_v}\right].
\end{aligned} \tag{3.40}$$

Similarly the second term on the right hand side of (3.36) can be written as

$$\begin{aligned}
&\Pr\{\bar{x}_2 - \bar{x}_1 \leq 0, \bar{x}_1 + \bar{x}_2 \geq 2x\} \\
&= \int_0^{\infty} p'(x) g(x) dx
\end{aligned} \tag{3.41}$$

where

$$\begin{aligned}
p'(x) &= \Pr\{U \leq 0, V \geq 2x\} \\
&= \Pr\{Z_1 \leq -\frac{\mu_u}{\sigma_u}, Z_2 \geq \frac{2x - \mu_v}{\sigma_v}\} \\
&= \int_{-\infty}^{-\mu_u/\sigma_u} \int_{\frac{2x - \mu_v}{\sigma_v}}^{\infty} h_{\rho}(z_1, z_2) dz_2 dz_1 \\
&= \int_{-\infty}^{-\frac{\mu_u}{\sigma_v}} \int_{-\infty}^{-(2x - \mu_v)/\sigma_v} h_{-\rho}(z_1, z_2) dz_2 dz_1 \\
&= H_{-\rho}\left[-\frac{\mu_u}{\sigma_u}, \frac{\mu_v - 2x}{\sigma_v}\right]
\end{aligned} \tag{3.42}$$

Finally, an approximation for the conditional probability of misclassification is

$$E[e_{12}(\bar{x}_1, \bar{x}_2)] \approx \int_0^\infty g(x) [H_{-\rho}\left(\frac{\mu_u}{\sigma_u}, \frac{2x - \mu_v}{\sigma_v}\right) + H_{-\rho}\left(-\frac{\mu_u}{\sigma_u}, \frac{\mu_v - 2x}{\sigma_v}\right)] dx. \quad (3.43)$$

In equation (3.43) $g(x)$ is the probability density function

$$g(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(\ln x - \mu_{1y})^2\right], \quad 0 < x < \infty.$$

3.4 Logit Normal Transformation (S_B)

Definition: The random variable X is Logit normal if $Y = \ln\left[\frac{X}{1-X}\right]$ is normal, $0 < X < 1$.

a) Distribution Theory and Rules

The density function of Y is given by

$$f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(y - \mu_y)^2\right], \quad -\infty < y < \infty. \quad (3.44)$$

The parameters μ_y, σ_y^2 satisfy the conditions $-\infty < \mu_y < \infty$, $\sigma_y^2 > 0$. By the transformation

$$Y = \ln\left[\frac{X}{1-X}\right] \quad (3.45)$$

the density function of X is obtained as

$$f(x) = \frac{1}{x(1-x)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}\left(\ln\left[\frac{x}{1-x}\right] - \mu_y\right)^2\right], \quad 0 < x < 1. \quad (3.46)$$

We assume that σ_y^2 is common to Π_1, Π_2 and for

$$X \in \Pi_1 \quad \mu_y = \mu_{1y}$$

$$X \in \Pi_2 \quad \mu_y = \mu_{2y}$$

The conditional probability of misclassification when an observation from Π_1 is misclassified is given by

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= \int_{\frac{\bar{x}_1 + \bar{x}_2}{2}}^1 \frac{1}{x(1-x)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(\ln[\frac{x}{1-x}] - \mu_{1y})^2\right] dx \\ &\quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \int_0^{(\bar{x}_1 + \bar{x}_2)/2} \frac{1}{x(1-x)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(\ln[\frac{x}{1-x}] - \mu_{1y})^2\right] dx \\ &\quad \text{if } \bar{x}_1 \geq \bar{x}_2 . \end{aligned}$$

Transforming the variable of integration by

$$t = \frac{\ln(\frac{x}{1-x}) - \mu_{1y}}{\sigma_y} ,$$

we obtain, for $\bar{x}_1 < \bar{x}_2$

$$e_{12}(\bar{x}_1, \bar{x}_2) = \int_{\ln[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}] - \mu_{1y}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt$$

and for $\bar{x}_1 \geq \bar{x}_2$

$$e_{12}(\bar{x}_1, \bar{x}_2) = \int_{-\infty}^{\frac{\ln[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}] - \mu_{1y}}{\sigma_y}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt .$$

This can be written as

$$\begin{aligned}
 e_{12}(\bar{x}_1, \bar{x}_2) &= 1 - \Phi\left(\frac{\ln\left[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}\right] - \mu_{1Y}}{\sigma_Y}\right) \text{ if } \bar{x}_1 < \bar{x}_2, \\
 &= \Phi\left(\frac{\ln\left[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}\right] - \mu_{1Y}}{\sigma_Y}\right) \text{ if } \bar{x}_1 \geq \bar{x}_2.
 \end{aligned}
 \tag{3.47}$$

By the same procedure as above, the conditional probability of misclassification $e_{21}(\bar{x}_1, \bar{x}_2)$ is seen to be

$$\begin{aligned}
 e_{21}(\bar{x}_1, \bar{x}_2) &= \Phi\left(\frac{\ln\left[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}\right] - \mu_{2Y}}{\sigma_Y}\right) \text{ if } \bar{x}_1 < \bar{x}_2, \\
 &= 1 - \Phi\left(\frac{\ln\left[\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2}\right] - \mu_{2Y}}{\sigma_Y}\right) \text{ if } \bar{x}_1 \geq \bar{x}_2.
 \end{aligned}
 \tag{3.48}$$

Note that if $X \in \Pi_1$

$$\mu_r' = E(X^r) = \int_0^1 x^r \frac{1}{x(1-x)\sigma_Y \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_Y^2} \left\{ \ln\left[\frac{x}{1-x}\right] - \mu_{1Y} \right\}^2\right) dx.
 \tag{3.49}$$

Although it is possible to give an explicit expression for μ_r' , i.e. not involving integral signs, it is very complicated. Johnson ([24], equations 56 and 57) gives the expression for μ_1' . The other moments are given as partial derivatives of the first raw moment. The notation used here for the moments is

$$\text{if } X \in \Pi_1 \quad E(X) = \mu_{1B}$$

$$V(X) = \sigma_{1B}^2,$$

$$\text{if } X \in \Pi_2 \quad E(X) = \mu_{2B}$$

$$V(X) = \sigma_{2B}^2.$$

The actual forms are found in Johnson [24, equations 56, 57 and 58].

b) Distribution of the Conditional Probability of Misclassification

From equation (3.47)

$$\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \text{ iff}$$

$$\{\bar{x}_1 < \bar{x}_2 \text{ and } \frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} > \exp[\mu_{1Y} - \sigma_Y \phi^{-1}(z)]\}$$

or

$$\{\bar{x}_1 \geq \bar{x}_2 \text{ and } \frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} \leq \exp[\mu_{1Y} + \sigma_Y \phi^{-1}(z)]\}. \quad (3.50)$$

Hence

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} = \Pr\{\bar{x}_2 - \bar{x}_1 > 0,$$

$$\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} \geq \exp[\mu_{1Y} - \sigma_Y \phi^{-1}(z)]\}$$

$$+ \Pr\{\bar{x}_2 - \bar{x}_1 \leq 0, \frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} \leq \exp[\mu_{1Y} + \sigma_Y \phi^{-1}(z)]\}.$$

(3.51)

Let

$$k_1 = \exp[\mu_{1Y} - \sigma_Y \phi^{-1}(z)],$$

$$k_2 = \exp[\mu_{1Y} + \sigma_Y \phi^{-1}(z)],$$

then, as $\bar{x}_1 + \bar{x}_2 < 2$,

$$\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} \geq k_1 \Rightarrow \bar{x}_1 + \bar{x}_2 \geq 2k_1 - k_1(\bar{x}_1 + \bar{x}_2)$$

$$\Rightarrow (1 + k_1)(\bar{x}_1 + \bar{x}_2) \geq 2k_1$$

$$\Rightarrow \bar{x}_1 + \bar{x}_2 \geq \frac{2k_1}{1 + k_1}.$$

Similarly

$$\frac{\bar{x}_1 + \bar{x}_2}{2 - \bar{x}_1 - \bar{x}_2} \leq k_2 \Rightarrow \bar{x}_1 + \bar{x}_2 \leq \frac{2k_2}{1 + k_2}.$$

Therefore (3.50) becomes

$$\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \text{ iff } \{\bar{x}_1 + \bar{x}_2 \geq \frac{2k_1}{1+k_1} \text{ and } \bar{x}_1 < \bar{x}_2\}$$

or

$$\{\bar{x}_1 + \bar{x}_2 \leq \frac{2k_2}{1+k_2} \text{ and } \bar{x}_1 \geq \bar{x}_2\}. \quad (3.52)$$

Thus the distribution function of $e_{12}(\bar{x}_1, \bar{x}_2)$ can be written as

$$\begin{aligned} \Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} &= \Pr\{\bar{x}_2 - \bar{x}_1 > 0, \bar{x}_1 + \bar{x}_2 \geq \frac{2k_1}{1+k_1}\} \\ &\quad + \Pr\{\bar{x}_2 - \bar{x}_1 \leq 0, \bar{x}_1 + \bar{x}_2 \leq \frac{2k_2}{1+k_2}\} \end{aligned} \quad (3.53)$$

Let $U = \bar{x}_2 - \bar{x}_1$

$V = \bar{x}_1 + \bar{x}_2$

Then

$$\begin{aligned} \Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} &= \Pr\{U > 0, V \geq \frac{2k_1}{1+k_1}\} + \\ &\quad \Pr\{U < 0, V \leq \frac{2k_2}{1+k_2}\} . \end{aligned} \quad (3.54)$$

Standardizing U and V we obtain

$$\begin{aligned} \Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} &= \Pr\left\{\frac{U - \mu_u}{\sigma_u} > -\frac{\mu_u}{\sigma_u}, \right. \\ &\quad \left. \frac{V - \mu_v}{\sigma_v} \geq \frac{\frac{2k_1}{1+k_1} - \mu_v}{\sigma_v}\right\} + \Pr\left\{\frac{U - \mu_u}{\sigma_u} < -\frac{\mu_u}{\sigma_u}, \frac{V - \mu_v}{\sigma_v} \leq \frac{\frac{2k_2}{1+k_2} - \mu_v}{\sigma_v}\right\} \\ &= \Pr\left\{Z_1 > -\frac{\mu_u}{\sigma_u}, Z_2 \geq \frac{\frac{2k_1}{1+k_1} - \mu_v}{\sigma_v}\right\} + \Pr\left\{Z_1 < -\frac{\mu_u}{\sigma_u}, Z_2 \leq \frac{\frac{2k_2}{1+k_2} - \mu_v}{\sigma_v}\right\} \end{aligned} \quad (3.55)$$

Approximating the distributions of $\bar{x}_1, \bar{x}_2$ by the normal distribution for large n_1, n_2 , we see that

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \approx \text{Standard BVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix} \right] .$$

See equation (3.12) for the expression of ρ^* .

The equation (3.55) becomes

$$\begin{aligned} \Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx H_{\rho^*} \left[\frac{\mu_u}{\sigma_u}, \frac{\mu_v - \frac{2k_1}{1+k_1}}{\sigma_v} \right] + \\ H_{\rho^*} \left[-\frac{\mu_u}{\sigma_u}, \frac{\frac{2k_2}{1+k_2} - \mu_v}{\sigma_v} \right]. \quad (3.56) \end{aligned}$$

Substituting

$$\mu_u = \mu_{2B} - \mu_{1B},$$

$$\mu_v = \mu_{2B} + \mu_{1B},$$

$$\sigma_u^2 = \frac{\sigma_{1B}^2}{n_1} + \frac{\sigma_{2B}^2}{n_2} = \sigma_v^2,$$

$$\sigma_{uv} = \frac{\sigma_{2B}^2}{n_2} - \frac{\sigma_{1B}^2}{n_1},$$

the distribution function of $e_{12}(\bar{x}_1, \bar{x}_2)$ can be written as

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx G[h_{11}, h_{21}; \rho^*] + G[h_{12}, h_{22}; \rho^*]. \quad (3.57)$$

In equation (3.57)

$$h_{11} = \frac{(\mu_{2B} - \mu_{1B})}{\{\sigma_u^2\}^{\frac{1}{2}}} = -h_{12},$$

$$h_{21} = \frac{(\mu_{2B} + \mu_{1B}) - 2\exp[\mu_{1y} - \sigma_y \Phi^{-1}(z)] / \{1 + \exp[\mu_{1y} - \sigma_y \Phi^{-1}(z)]\}}{\left\{ \frac{\sigma_{1B}^2}{n_1} + \frac{\sigma_{2B}^2}{n_2} \right\}^{\frac{1}{2}}}$$

$$h_{22} = \frac{2\{\exp[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)]\} / \{1 + \exp[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)]\} - (\mu_{1B} + \mu_{2B})}{\left\{ \frac{\sigma_{1B}^2}{n_1} + \frac{\sigma_{2B}^2}{n_2} \right\}^{1/2}}$$

$$\rho^* = \frac{\frac{\sigma_{2B}^2}{n_2} - \frac{\sigma_{1B}^2}{n_1}}{\frac{\sigma_{2B}^2}{n_2} + \frac{\sigma_{1B}^2}{n_1}}.$$

As before

$$G(z_1, z_2; \rho^*) = H_{\rho^*}(z_1, z_2).$$

If the distribution function of $e_{12}(\bar{x}_1, \bar{x}_2)$ is written as

$$R(z) = \int_{-\infty}^{h_{21}(z)} \int_{-\infty}^{h_{11}} f(z_1, z_2) dz_1 dz_2 + \int_{-\infty}^{h_{22}(z)} \int_{-\infty}^{-h_{11}} f(z_1, z_2) dz_1 dz_2 \quad (3.58)$$

the probability density function of $e_{12}(\bar{x}_1, \bar{x}_2)$ is given by

$$r(z) = R'(z)$$

where

$$r(z) = \int_{-\infty}^{h_{11}} f[z_1, h_{21}(z)] h'_{21}(z) dz_1 + \int_{-\infty}^{-h_{11}} f[z_1, h_{22}(z)] h'_{22}(z) dz_1. \quad (3.59)$$

Again, as in equation (3.35), it is a very complicated function of z .

c) Unconditional Probability of Misclassification

The unconditional probability of misclassification of $e_{12}(\bar{x}_1, \bar{x}_2)$ is given by

$$\begin{aligned}
E\{e_{12}(\bar{x}_1, \bar{x}_2)\} &= \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{(\bar{x}_1 + \bar{x}_2)}{2} \geq 0 | X \in \Pi_1\} \\
&\quad + \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) < 0 | X \in \Pi_1\}.
\end{aligned}
\tag{3.60}$$

The distributions of $\bar{x}_1, \bar{x}_2$ and X are independent with

$$g(x) = \frac{1}{x(1-x)\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\ln\left[\frac{x}{1-x}\right] - \mu_{1Y})^2\right]$$

$$0 < x < 1 \tag{3.61}$$

and, by the Central Limit Theorem,

$$\begin{aligned}
\bar{x}_1 &\approx N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \\
\bar{x}_2 &\approx N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right),
\end{aligned}$$

where

$$\begin{aligned}
\mu_1 &= E(\bar{X}_1) = \mu_{1B}, \\
\mu_2 &= E(\bar{X}_2) = \mu_{2B}, \\
\sigma_1^2 &= V(X | X \in \Pi_1) = \sigma_{1B}^2, \\
\sigma_2^2 &= V(X | X \in \Pi_2) = \sigma_{2B}^2.
\end{aligned}$$

Using the result on conditional expectation given in (3.37), we can determine the right hand side of (3.60) conditionally and remove the conditionality. Therefore the first term on the right hand side of (3.60) becomes

$$\begin{aligned}
& \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \geq 0\} \\
&= \int_0^1 \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, (\bar{x}_1 + \bar{x}_2) \leq 2x | x\} g(x) dx.
\end{aligned} \tag{3.62}$$

The probability in the integrand has been found earlier in a form involving the BVN distribution function and is given in equations (3.56) and (3.57). Let us write this probability as $p(x)$.

Then

$$p(x) = \Pr\{U > 0, V \leq 2x\} \tag{3.63}$$

where

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right].$$

Standardizing the variables,

$$\begin{aligned}
p(x) &= \Pr\{Z_1 > \frac{\mu_u}{\sigma_u}, Z_2 < \frac{2x - \mu_v}{\sigma_v}\} \\
&= \int_{-\frac{\mu_u}{\sigma_u}}^{\infty} \int_{-\infty}^{\frac{2x - \mu_v}{\sigma_v}} h_{\rho}(z_1, z_2) dz_2 dz_1 \\
&= \int_{-\infty}^{\mu_u/\sigma_u} \int_{-\infty}^{(2x - \mu_v)/\sigma_v} h_{-\rho}(z_1, z_2) dz_1 dz_2 \\
&= H_{-\rho} \left[\frac{\mu_u}{\sigma_u}, \frac{2x - \mu_v}{\sigma_v} \right].
\end{aligned} \tag{3.64}$$

Similarly, corresponding to the third term of (3.60) we have:

$$\Pr\{\bar{x}_2 - \bar{x}_1 \leq 0, \bar{x}_1 + \bar{x}_2 \geq 2x\} = \int_0^1 p'(x) g(x) dx$$

where

$$\begin{aligned}
 p'(x) &= \Pr\{U \leq 0, V \geq 2x\} \\
 &= \Pr\left\{Z_1 \leq -\frac{\mu_u}{\sigma_u}, Z_2 \geq \frac{2x - \mu_v}{\sigma_v}\right\} \\
 &= \int_{-\infty}^{-\mu_u/\sigma_u} \int_{\frac{2x - \mu_v}{\sigma_v}}^{\infty} h(z_1, z_2) dz_2 dz_1 \\
 &= \int_{-\infty}^{-\mu_u/\sigma_u} \int_{-\infty}^{-(2x - \mu_v)/\sigma_v} h_{-\rho}(z_1, z_2) dz_2 dz_1 \\
 &= H_{-\rho}\left[-\frac{\mu_u}{\sigma_u}, \frac{\mu_v - 2x}{\sigma_v}\right].
 \end{aligned}$$

Finally

$$\begin{aligned}
 E[e_{12}(\bar{x}_1, \bar{x}_2)] &\approx \int_0^1 g(x) \left\{ H_{-\rho}\left[\frac{\mu_u}{\sigma_u}, \frac{2x - \mu_v}{\sigma_v}\right] \right. \\
 &\quad \left. + H_{-\rho}\left[-\frac{\mu_u}{\sigma_u}, \frac{\mu_v - 2x}{\sigma_v}\right] \right\} dx, \quad (3.65)
 \end{aligned}$$

where $g(x)$ in (3.65) is given by

$$g(x) = \frac{1}{x(1-x)\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(\ln[\frac{x}{1-x}] - \mu_{1Y})^2\right], \quad 0 < x < 1.$$

3.5 Inverse Hyperbolic Sine Normal (S_U)

Definition: The random variable X is Inverse Hyperbolic Sine Normal if $Y = \sinh^{-1}(X)$ is normal, $-\infty < X < \infty$.

a) Distribution Theory and Rules

The density function of Y is given by

$$f(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(y - \mu_Y)^2\right], \quad -\infty < y < \infty \quad (3.66)$$

the parameters μ_Y, σ_Y satisfying the conditions $-\infty < \mu_Y < \infty$, $\sigma_Y > 0$. By the transformation

$$Y = \sinh^{-1}(X),$$

the density function of X is seen to be

$$f(x) = \frac{1}{(x^2+1)^{\frac{1}{2}}\sigma_Y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\sinh^{-1}(x) - \mu_Y)^2\right] \quad -\infty < x < \infty \quad (3.67)$$

We assume that σ_Y^2 is common to the populations Π_1 and Π_2 for

$$X \in \Pi_1 \quad \mu_Y = \mu_{1Y}$$

and

$$X \in \Pi_2 \quad \mu_Y = \mu_{2Y}.$$

Then the conditional probability of misclassification when an observation from Π_1 is misclassified is given by

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= \int_{(\bar{x}_1 + \bar{x}_2)/2}^{\infty} \frac{1}{(x^2+1)^{\frac{1}{2}}\sigma_Y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\sinh^{-1}(x) - \mu_{1Y})^2\right] dx, \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \int_{-\infty}^{(\bar{x}_1 + \bar{x}_2)/2} \frac{1}{(x^2+1)^{\frac{1}{2}}\sigma_Y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_Y^2}(\sinh^{-1}(x) - \mu_{1Y})^2\right] dx, \quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned} \quad (3.68)$$

Transforming the variable of integration by

$$t = \frac{\sinh^{-1}(x) - \mu_{1Y}}{\sigma_Y}$$

and letting

$$\delta = (\bar{x}_1 + \bar{x}_2)/2$$

we obtain

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= \int_{[\sinh^{-1}(\delta) - \mu_{1Y}]/\sigma_Y}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{t^2}{2}\right] dt, \\ &\quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \int_{-\infty}^{[\sinh^{-1}(\delta) - \mu_{1Y}]/\sigma_Y} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt, \\ &\quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned} \quad (3.69)$$

Equation (3.69) can be written as

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= 1 - \Phi\left[\frac{\sinh^{-1}(\delta) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\sinh^{-1}(\delta) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2, \end{aligned} \quad (3.70)$$

and putting back the expression of δ we get

$$\begin{aligned} e_{12}(\bar{x}_1, \bar{x}_2) &= 1 - \Phi\left[\frac{\sinh^{-1}\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\sinh^{-1}\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{1Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2. \end{aligned} \quad (3.71)$$

Using the above procedure the conditional probability of misclassification $e_{21}(\bar{x}_1, \bar{x}_2)$ is given by

$$\begin{aligned} e_{21}(\bar{x}_1, \bar{x}_2) &= \Phi\left[\frac{\sinh^{-1}\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{2Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 < \bar{x}_2 \\ &= 1 - \Phi\left[\frac{\sinh^{-1}\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) - \mu_{2Y}}{\sigma_Y}\right] \quad \text{if } \bar{x}_1 \geq \bar{x}_2 . \end{aligned} \quad (3.72)$$

From Johnson [24, equation (37)], we note:

$$\begin{aligned} \text{if } X \in \Pi_1 \quad E(X) &= -\omega \sinh(\mu_{1Y}) \\ V(X) &= \frac{1}{2}[\omega^2 - 1]\{-\omega^2 \cosh(2\mu_{1Y}) + 1\} , \\ \text{if } X \in \Pi_2 \quad E(X) &= -\omega \sinh(\mu_{2Y}) \\ V(X) &= \frac{1}{2}[\omega^2 - 1]\{-\omega^2 \cosh(2\mu_{2Y}) + 1\} , \end{aligned}$$

where $\omega = \exp[\sigma_Y^2/2]$.

b) Distribution of the Conditional Probability of Misclassification

From equation (3.70) we notice that

$$\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\}$$

if and only if

$$\left\{ \begin{array}{l} x_1 < x_2 \quad \text{and} \quad x_1 + x_2 \geq 2\sinh[\mu_{1Y} - \sigma_Y \Phi^{-1}(z)] \\ \text{or} \quad x_1 \geq x_2 \quad \text{and} \quad x_1 + x_2 \leq 2\sinh[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)] \end{array} \right\} . \quad (3.73)$$

Hence:

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} = \Pr\{(\bar{x}_2 - \bar{x}_1) > 0,$$

$$(\bar{x}_1 + \bar{x}_2) \geq 2\sinh[\mu_{1Y} - \sigma_Y \Phi^{-1}(z)]\}$$

$$+ \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0,$$

$$(\bar{x}_1 + \bar{x}_2) < 2\sinh[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)]\}.$$

(3.74)

Let

$$U = \bar{x}_2 - \bar{x}_1$$

$$V = \bar{x}_1 + \bar{x}_2$$

then

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} = \Pr\{U > 0, V > k\} + \Pr\{U \leq 0, V < k_2\}$$

(3.75)

where

$$k_1 = 2\sinh[\mu_{1Y} - \sigma_Y \Phi^{-1}(z)]$$

$$k_2 = 2\sinh[\mu_{1Y} + \sigma_Y \Phi^{-1}(z)]$$

Standardizing U and V we obtain

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} = \Pr\left\{\frac{U - \mu_U}{\sigma_U} > -\frac{\mu_U}{\sigma_U}, \frac{V - \mu_V}{\sigma_V} > \frac{k_1 - \mu_V}{\sigma_V}\right\}$$

$$+ \Pr\left\{\frac{U - \mu_U}{\sigma_U} < -\frac{\mu_U}{\sigma_U}, \frac{V - \mu_V}{\sigma_V} < \frac{k_2 - \mu_V}{\sigma_V}\right\}$$

$$= \Pr\left\{Z_1 > -\frac{\mu_U}{\sigma_U}, Z_2 > \frac{k_1 - \mu_V}{\sigma_V}\right\} +$$

$$+ \Pr\left\{Z_2 < -\frac{\mu_U}{\sigma_U}, Z_2 < \frac{k_2 - \mu_V}{\sigma_V}\right\}$$

(3.76)

Approximating the distribution of $\bar{x}_1, \bar{x}_2$ for large n_1, n_2 by the normal distribution, we see that

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim \text{Standard BVN} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix} \right].$$

We therefore write equation (3.76) as

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx H_{\rho^*} \left[\frac{\mu_u}{\sigma_u}, \frac{\mu_v - k_1}{\sigma_v} \right] + H_{\rho^*} \left[-\frac{\mu_u}{\sigma_u}, \frac{k_2 - \mu_v}{\sigma_v} \right]. \quad (3.77)$$

In equation (3.77)

$$\mu_u = -\omega [\sinh(\mu_{2Y}) - \sinh(\mu_{1Y})]$$

$$\mu_v = -\omega [\sinh(\mu_{1Y}) + \sinh(\mu_{2Y})]$$

$$\sigma_u^2 = \sigma_v^2 = \frac{1}{2}(\omega^2 - 1) \left[-\frac{\omega^2 \cosh(2\mu_{1Y})}{n_1} - \frac{\omega^2 \cosh(2\mu_{2Y})}{n_2} + \frac{1}{n_1} + \frac{1}{n_2} \right]$$

$$\sigma_{uv} = \frac{1}{2}(\omega^2 - 1) \left[+\frac{\omega^2 \cosh(2\mu_{1Y})}{n_1} - \frac{\omega^2 \cosh(2\mu_{2Y})}{n_2} - \frac{1}{n_1} + \frac{1}{n_2} \right].$$

Alternatively

$$\Pr\{e_{12}(\bar{x}_1, \bar{x}_2) \leq z\} \approx G[h_{11}, h_{21}; \rho^*] + G[h_{12}, h_{22}; \rho^*] \quad (3.78)$$

with

$$\begin{aligned} h_{11} &= \{-\omega [\sinh(\mu_{2Y}) - \sinh(\mu_{1Y})]\} / \sigma_u \\ &= -h_{12} \end{aligned}$$

$$h_{21} = [-\omega[\sinh(\mu_{2y}) + \sinh(\mu_{1y})] - 2\sinh\{\mu_{1y} - \sigma_y \Phi^{-1}(z)\}]/\sigma_u$$

$$h_{22} = [2\sinh\{\sigma_y \Phi^{-1}(z) + \mu_{1y}\} + \omega\{\sinh(\mu_{2y}) + \sinh(\mu_{1y})\}]/\sigma_u$$

$$\rho^* = \frac{\left[\frac{\{-\omega^2 \cosh(2\mu_{2y}) + 1\}}{n_2} - \frac{\{\omega^2 \cosh(2\mu_{1y}) + 1\}}{n_1} \right]}{\left[\frac{\{-\omega^2 \cosh(2\mu_{1y}) + 1\}}{n_1} + \frac{\{-\omega^2 \cosh(2\mu_{2y}) + 1\}}{n_2} \right]}.$$

From the distribution function in (3.78)

$$R(z) = \int_{-\infty}^{h_{21}(z)} \int_{-\infty}^{h_{11}} f(z_1, z_2) dz_1 dz_2 + \int_{-\infty}^{h_{22}(z)} \int_{-\infty}^{h_{11}} f(z_1, z_2) dz_1 dz_2 \quad (3.79)$$

the probability density function of $e_{12}(\bar{x}_1, \bar{x}_2)$ is obtained as

$$r(z) = \int_{-\infty}^{h_{11}} f[z_1, h_{21}(z)] h'_{21}(z) dz_1 + \int_{-\infty}^{-h_{11}} f[z_1, h_{22}(z)] h'_{22}(z) dz_1 \quad (3.80)$$

c) Unconditional Probability of Misclassification

The unconditional probability of misclassification of $e_{12}(\bar{x}_1, \bar{x}_2)$ is given by

$$\begin{aligned} E\{e_{12}(\bar{x}_1, \bar{x}_2)\} &= \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) > 0 | X \in \Pi_1\} \\ &\quad + \Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \leq 0 | X \in \Pi_1\} \end{aligned} \quad (3.81)$$

[see equation (3.36)].

The distributions of $\bar{x}_1, \bar{x}_2$ and X are independent, with

$$g(x) = \frac{1}{(x^2+1)^{\frac{1}{2}} \sigma_y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(\sinh^{-1}(x) - \mu_{1Y})^2\right],$$

$$-\infty < x < \infty \quad (3.82)$$

the parameters μ_{1Y}, σ_y satisfying the conditions $-\infty < \mu_{1Y} < \infty$, $\sigma_y > 0$. By the Central Limit Theorem,

$$\bar{x}_1 \approx N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right),$$

$$\bar{x}_2 \approx N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

where $\mu_1 = E(\bar{X}_1) = -\omega \sinh(\mu_{1Y})$,

$$\mu_2 = E(\bar{X}_2) = -\omega \sinh(\mu_{2Y}),$$

$$\sigma_1^2 = V(X|X \in \Pi_1) = \frac{1}{2}(\omega^2 - 1)[- \omega^2 \cosh(2\mu_{1Y}) + 1],$$

$$\sigma_2^2 = V(X|X \in \Pi_2) = \frac{1}{2}(\omega^2 - 1)[- \omega^2 \cosh(2\mu_{2Y}) + 1],$$

where $\omega = \exp[\sigma_y^2/2]$.

As before, using the well known result on conditional probabilities given in equation (3.37) we can determine the right hand side of (3.81) conditionally and remove the conditionality. The second term of equation (3.81) then becomes

$$\Pr\{(\bar{x}_2 - \bar{x}_1) > 0, X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \geq 0\}$$

$$= \int_{-\infty}^{\infty} \Pr\{(\bar{x}_2 - \bar{x}_1) > 0, \bar{x}_1 + \bar{x}_2 \leq 2x|x\} g(x) dx \quad (3.83)$$

The probability in the integrand has been found earlier in a form involving the BVN distribution function. Let us write the corresponding unconditional probability as $p(x)$, then

$$p(x) = \Pr\{U > 0, V < 2x\} \quad (3.84)$$

Where $U = \bar{x}_2 - \bar{x}_1$, $V = \bar{x}_1 + \bar{x}_2$, and

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right].$$

Thus

$$\begin{aligned} p(x) &= \Pr\{z_1 > -\frac{\mu_u}{\sigma_u}, z_2 \leq \frac{2x - \mu_v}{\sigma_v}\} \\ &= \int_{-\mu_u/\sigma_u}^{\infty} \int_{-\infty}^{(2x - \mu_v)/\sigma_v} h_{\rho}(z_1, z_2) dz_2 dz_1 \\ &= \int_{-\infty}^{\mu_u/\sigma_u} \int_{-\infty}^{(2x - \mu_v)/\sigma_v} h_{-\rho}(z_1, z_2) dz_1 dz_2 \\ &= H_{-\rho} \left[\frac{\mu_u}{\sigma_u}, \frac{2x - \mu_v}{\sigma_v} \right]. \end{aligned} \quad (3.85)$$

Similarly, corresponding to the third term of (3.81) we have:

$$\Pr\{(\bar{x}_2 - \bar{x}_1) \leq 0, (\bar{x}_1 + \bar{x}_2) \geq 2x\} = \int_{-\infty}^{\infty} p'(x) g(x) dx \quad (3.86)$$

where

$$\begin{aligned} p'(x) &= \Pr\{U \leq 0, V \leq 2x\} \\ &= \Pr\{z_1 \leq -\frac{\mu_u}{\sigma_u}, z_2 \geq \frac{2x - \mu_v}{\sigma_v}\} \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{-\mu_u/\sigma_u} \int_{(2x-\mu_v)/\sigma_v}^{\infty} h_{\rho}(z_1, z_2) dz_2 dz_1 \\
&= \int_{-\infty}^{-\mu_u/\sigma_u} \int_{-\infty}^{-(2x-\mu_v)/\sigma_v} h_{-\rho}(z_1, z_2) dz_2 dz_1 \\
&= H_{-\rho} \left[-\frac{\mu_u}{\sigma_u}, \frac{\mu_v - 2x}{\sigma_v} \right]
\end{aligned} \tag{3.87}$$

Finally

$$E\{e_{12}(\bar{x}_1, \bar{x}_2)\} \approx \int_{-\infty}^{\infty} g(x) \left[H_{-\rho} \left(\frac{\mu_u}{\sigma_u}, \frac{2x-\mu_v}{\sigma_v} \right) + H_{-\rho} \left(-\frac{\mu_u}{\sigma_v}, \frac{\mu_v-2x}{\sigma_v} \right) \right] dx, \tag{3.88}$$

with $g(x)$ in (3.88) given by

$$g(x) = \frac{1}{(x^2+1)^{\frac{1}{2}} \sigma_y \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma_y^2} (\sinh^{-1}(x) - \mu_{1Y})^2 \right],$$

$$-\infty < x < \infty.$$

CHAPTER 4

DISCRIMINATION FOR NON-NORMAL DISTRIBUTIONS:

EDGEWORTH SERIES DISTRIBUTION (ESD).

4.1 Introduction

The effect of non-normality in a two population discrimination problem is studied. The distributions in the two populations are assumed to be the univariate Edgeworth Series with different means and the same variance.

Subrahmaniam [55] has studied the distribution of the linear and quadratic forms using the ESD, with a density function introduced by Gayen [14]. The type of non-normality which he considers includes both skewness and kurtosis in their standard forms. Here we consider non-normality due to skewness alone.

Let X_{1j} ($j = 1, 2, \dots, n_1$) and X_{2j} ($j = 1, 2, \dots, n_2$) denote two independent random samples from populations Π_1 and Π_2 respectively. The observations X_{1j} come from the common distribution defined by the density function

$$f(x) = \left(1 - \frac{\lambda_3}{6} D^3\right) \phi\left(\frac{x - \mu_1}{\sigma}\right), \quad -\infty < x < \infty \quad (4.1)$$

and the observations X_{2j} come from the common distribution defined by the density function

$$f(x) = \left(1 - \frac{\lambda_3}{6} D^3\right) \phi\left(\frac{x - \mu_2}{\sigma}\right), \quad -\infty < x < \infty, \quad (4.2)$$

the parameters λ_3, μ_i , ($i = 1, 2$) and σ satisfying the

conditions $-\infty < \lambda_3 < \infty$, $-\infty < \mu_i < \infty$ and $\sigma > 0$. Here D stands for the operator

$$\left(\frac{\partial}{\partial x}\right) \phi\left(\frac{x - \mu_i}{\sigma}\right) \text{ is the density function}$$

$$\sigma^{-1} (2\pi)^{-\frac{1}{2}} \exp[-(x - \mu_i)^2 / 2\sigma^2] \quad (4.3)$$

and λ_3 is the skewness factor.

In what follows all terms involving powers of λ_3 higher than the first are ignored.

Let X be a new observation, obtained independently of X_{1j} and X_{2j} observations, drawn from either Π_1 or Π_2 . We would like to classify X as belonging to Π_1 or Π_2 . To do this a classification rule is needed which in turn means that the discriminant function has to be determined. In practice one could use the univariate analogue of the W discriminant function which is defined as

$$W = D[X; \bar{x}_1, \bar{x}_2; \sigma^2] = [X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)](\bar{x}_1 - \bar{x}_2) / \sigma^2 \quad (4.4)$$

when σ^2 is known and

$$W = D[X; \bar{x}_1, \bar{x}_2; S^2] = [X - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)](\bar{x}_1 - \bar{x}_2) / S^2 \quad (4.5)$$

when σ^2 is estimated by S^2 , the pooled sample variance of the observations in Π_1 and Π_2 .

We note here that the W discriminant function is derived by use of the likelihood ratio rule when the distributions in the populations Π_1 and Π_2 are normal and known, with density functions

$$f_i(x) = \exp[-(\frac{x - \mu_i}{\sigma})^2/2]/\sqrt{2\pi} \sigma, \quad -\infty < x < \infty, \quad (4.6)$$

$$i = 1, 2.$$

Further the discriminant function defined by (4.4) or (4.5) assumes $\mu_1 < \mu_2$ [see Chapter 5 for derivation of the W discriminant function].

The classification regions based on W may be written as

$$\left. \begin{aligned} \hat{R}_1: \{X; X < (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 < \bar{x}_2 \text{ or} \\ X \geq (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 \geq \bar{x}_2\} \\ \hat{R}_2: \{X; X \geq (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 < \bar{x}_2 \text{ or} \\ X < (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 \geq \bar{x}_2\} \end{aligned} \right\} \quad (4.7)$$

where $\hat{R}_1$ and $\hat{R}_2$ correspond to the populations Π_1 and Π_2 respectively.

If all parameters, $\mu_1, \mu_2, (\mu_1 < \mu_2)$ and σ^2 are known the classification regions are defined as

$$\left. \begin{aligned} R_1: \{X; X < (\mu_1 + \mu_2)/2\} \\ R_2: \{X; X \geq (\mu_1 + \mu_2)/2\} \end{aligned} \right\} \quad (4.8)$$

where R_1, R_2 regions correspond to the populations Π_1 , and Π_2 , respectively.

The problem considered in this chapter is one of examining the effects, if any, on the errors of misclassification, when W , defined by means of the Edgeworth Series type distribution, is used for classifying X as arising from Π_1 or Π_2 . The effects are studied for varying values of λ_3 , the skewness factor.

Since the distribution theory is considerably simplified by the use of the characteristic function technique, the characteristic function (c.f.) of the Edgeworth Series distribution is derived in section (4.2). The problems involved with the probabilities of misclassification are studied in later sections.

4.2 Characteristic Function and Moments

Let X have the p.d.f.

$$f(x) = \left(1 - \frac{\lambda_3}{6} D^3\right) \phi\left(\frac{x - \mu}{\sigma}\right), \quad -\infty < x < \infty$$

In this section, the characteristic functions (c.f.) of X , $\bar{X}$ (the sample mean) and $\bar{X}_1 + \bar{X}_2$ are derived. These yield the moments of the distribution. In later sections the results obtained from the c.f. will be found to be of great applicability.

Lemma 4.1 The characteristic function of the r.v. X with p.d.f.

$$f(x) = \left\{1 - \frac{\lambda_3}{6} D^3\right\} \phi\left(\frac{x - \mu}{\sigma}\right)$$

is given by

$$\psi_X(t) = \left\{1 + \frac{\lambda_3}{6} (it)^3\right\} \exp[it\mu - \sigma^2 t^2/2] \quad (4.9)$$

Proof: By definition

$$\psi_x(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx. \quad (4.10)$$

Let us recall that, if we write $z = \frac{x - \mu}{\sigma}$

$$D^3 \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma^3} \left(\frac{d}{dz}\right)^3 \phi(z) \quad (4.11)$$

and that

$$\int_{-\infty}^{\infty} e^{itx} \phi\left(\frac{x - \mu}{\sigma}\right) dx = \exp[it\mu - \sigma^2 t^2/2] \quad (4.12)$$

Using (4.11) and (4.12) in (4.10) we get

$$\begin{aligned} \psi_x(t) &= \exp[it\mu - \sigma^2 t^2/2] + \\ &- \frac{\lambda_3}{6\sigma^3} e^{it\mu} \int_{-\infty}^{\infty} (3z - z^3) \phi(z) \exp[i\sigma z] dz. \end{aligned} \quad (4.13)$$

The integral on the right hand side of (4.13) can also be written as, putting $u = i\sigma$,

$$\begin{aligned} (2\pi)^{-1/2} &[3 \frac{d}{du} \int_{-\infty}^{\infty} \exp(uz - z^2/2) dz - \left(\frac{d}{du}\right)^3 \int_{-\infty}^{\infty} \exp(uz - z^2/2) dz] \\ &= 3 \frac{d}{du} (\exp[u^2/2]) - \left(\frac{d}{du}\right)^3 (\exp[u^2/2]) \\ &= 3[u \exp(u^2/2)] - [3u + u^3] \exp(u^2/2) \\ &= -\sigma^3 (it)^3 \exp(-\sigma^2 t^2/2) \end{aligned}$$

hence

$$\psi_x(t) = \left[1 + \frac{\lambda_3}{6} (it)^3\right] \exp[it\mu - \sigma^2 t^2/2] \quad (4.14)$$

Lemma 4.2 The characteristic function of the mean is given by

$$\psi_{\bar{X}}(t) = [1 + \frac{\lambda 3}{6} \frac{(it)^3}{n^2}] \exp[\mu it - \sigma^2 t^2 / 2n] \quad (4.15)$$

Proof:

The characteristic function of X is

$$\psi_X(t) = [1 + \frac{\lambda 3}{6} (it)^3] \exp[\mu it - t^2 \sigma^2 / 2]$$

therefore

$$\begin{aligned} \psi_{\bar{X}}(t) &= \prod_{i=1}^n \psi_X\left(\frac{t}{n}\right) = \{\psi_X\left(\frac{t}{n}\right)\}^n \\ &= [1 + \frac{\lambda 3}{6} (\frac{it}{n})^3]^n \exp[\mu it - (\sigma t)^2 / 2n] \\ &= [1 + \frac{\lambda 3}{6} \frac{(it)^3}{n^2}] \exp[\mu it - (\sigma t)^2 / 2n] \end{aligned} \quad (4.16)$$

The final result in (4.16) is obtained by use of the binomial expansion, breaking off after the second term.

From the result in (4.16) we can easily show that the first two moments of the random variable $\bar{X}$ when X is from the ESD are given by

$$\left. \begin{aligned} E(\bar{X}) &= \mu \\ E(\bar{X} - \mu)^2 &= V(\bar{X}) = \frac{\sigma^2}{n} \end{aligned} \right\} \quad (4.17)$$

Also from the result in (4.14) we can easily show that

$$\left. \begin{aligned} E(X) &= \mu \\ V(X) &= E(X - \mu)^2 = \sigma^2 \end{aligned} \right\} \quad (4.18)$$

Lemma 4.3 The characteristic function of

$z = (\bar{x}_1 - \mu_1) + (\bar{x}_2 - \mu_2)$ when $\bar{x}_1, \bar{x}_2$ are the means of independent samples of sizes n_1, n_2 respectively, is given by

$$\psi_z(t) = [1 + \frac{\lambda_3}{6n_2} (it)^3 + \frac{\lambda_3}{6n_1} (it)^3] \exp[-t^2 (\frac{\sigma^2}{2n_1} + \frac{\sigma^2}{2n_2})] \quad (4.18)$$

Proof:

We first show that the characteristic function of $\bar{x}_1 - \mu_1$ is given by

$$\begin{aligned} \psi_{\bar{x}_1 - \mu_1}(t) &= E[\exp\{it(\bar{x}_1 - \mu_1)\}] \\ &= E[\exp\{it\bar{x}_1 - it\mu_1\}] \\ &= \exp[-it\mu_1] E[\exp(it\bar{x}_1)] \\ &= [1 + \frac{\lambda_3}{6n_1} (it)^3] \exp\{-it\mu_1 + it\mu_1 - t^2\sigma^2/2n_1\} \\ &= [1 + \frac{\lambda_3}{6n_1} (it)^3] \exp\{-t^2\sigma^2/2n_1\} \quad (4.19) \end{aligned}$$

The result we need follows from a basic result, that the characteristic function of a sum of two independent random variables is the product of the characteristic functions of the individual random variables. Therefore

$$\begin{aligned} \psi_z(t) &= \psi_{\bar{x}_1 - \mu_1}(t) \psi_{\bar{x}_2 - \mu_2}(t) \\ &= [1 + \frac{\lambda_3}{6n_1} (it)^2] [1 + \frac{\lambda_3}{6n_2} (it)^3] \exp\{-t^2\sigma^2/2n_1 \\ &\quad - t^2\sigma^2/2n_2\} \end{aligned}$$

$$= [1 + \frac{\lambda_3}{6n_2} (it)^2 + \frac{\lambda_3}{6n_1} (it)^3] \exp\{-t^2(\sigma^2/2n_1 + \sigma^2/2n_2)\} \quad (4.20)$$

hence the result. From the result in (4.20) it can be seen that

$$\left. \begin{aligned} E(z) &= 0 \\ E(z^2) &= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \\ E(z^3) &= \lambda_3 \left(\frac{1}{n_1^2} + \frac{1}{n_2^2} \right) \\ E(z^4) &= 3\sigma^4 \left(\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right)^2 \end{aligned} \right\} \quad (4.21)$$

4.3 Optimum Probability of Misclassification

When all parameters of the distributions in the populations are known, the probability of misclassification is optimal in a sense that we cannot improve upon it. Using the normal rule, given in equation (4.8), we derive expressions for the optimum probability of misclassification when the observation X is from the ESD. When an observation from Π_1 is misclassified the optimum probability of misclassification is given by

$$\begin{aligned} \alpha_1[R, f] &= \Pr\{X \geq (\mu_1 + \mu_2)/2\} \\ &= \int_{\delta}^{\infty} [1 - \frac{\lambda_3}{6} D^3] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \\ &= \int_{\delta}^{\infty} [1 + \frac{\lambda_3}{6\sigma^3} H_3\left(\frac{x - \mu_1}{\sigma}\right)] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \end{aligned}$$

$$= \int_{\delta}^{\infty} \phi\left(\frac{x - \mu_1}{\sigma}\right) dx + \frac{\lambda_3}{6\sigma^3} \int_{\delta}^{\infty} H_3\left(\frac{x - \mu_1}{\sigma}\right) \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \quad (4.22)$$

where $\delta = (\mu_1 + \mu_2)/2$. and $H_r(x)$ is a Tchebycheff-Hermite polynomial of degree r and is defined by the identity

$$H_r(x) \phi(x) = [-D]^r \phi(x) \quad (4.23)$$

where $H_0 = 1$ by convention. See Kendall and Stuart [25, p. 155, equation 6.21]. Some properties of $H_r(x)$ are also discussed by Subrahmaniam [55]. Using the result (Kendall and Stuart [25, equation 6.43]),

$$\int_{-\infty}^t H_r(x) \phi(x) dx = -H_{r-1}(t) \phi(t) \quad (4.24)$$

and letting

$$z = \frac{x - \mu_1}{\sigma}$$

in (4.22) we obtain

$$\begin{aligned} \alpha_1[R, f] &= \int_{(\delta - \mu_1)/\sigma}^{\infty} \phi(z) dz + \frac{\lambda_3}{6\sigma^2} \int_{(\delta - \mu_1)/\sigma}^{\infty} H_3(z) \phi(z) dz \\ &= 1 - \phi\left[\frac{\delta - \mu_1}{\sigma}\right] + \frac{\lambda_3}{6\sigma^2} H_2\left(\frac{\delta - \mu_1}{\sigma}\right) \phi\left(\frac{\delta - \mu_1}{\sigma}\right) \\ &= 1 - \phi\left[\frac{\mu_2 - \mu_1}{2\sigma}\right] + \frac{\lambda_3}{6\sigma^2} \left[\left\{\frac{\mu_2 - \mu_1}{2\sigma}\right\}^2 - 1\right] \phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right). \end{aligned} \quad (4.25)$$

If an observation from Π_2 is misclassified the optimum

probability of misclassification is given by

$$\begin{aligned}
 \alpha_2[R, f] &= \Pr\{X < (\mu_1 + \mu_2)/2\} \\
 &= \int_{-\infty}^{\delta} [1 - \frac{\lambda_3}{6} D^3] \phi\left(\frac{x - \mu_2}{\sigma}\right) dx \\
 &= \int_{-\infty}^{\delta} \phi\left(\frac{x - \mu_2}{\sigma}\right) + \frac{\lambda_3}{6\sigma^3} \int_{-\infty}^{\delta} H_3\left(\frac{x - \mu_2}{\sigma}\right) \phi\left(\frac{x - \mu_2}{\sigma}\right) dx
 \end{aligned} \tag{4.26}$$

where $\delta = (\mu_1 + \mu_2)/2$

Letting $z = \frac{x - \mu_2}{\sigma}$ and using result in (4.24) we obtain

$$\begin{aligned}
 \alpha_2[R, f] &= \int_{-\infty}^{(\delta - \mu_2)/\sigma} \phi(z) dz + \frac{\lambda_3}{6\sigma^2} \int_{-\infty}^{(\delta - \mu_2)/\sigma} H_3(z) \phi(z) dz \\
 &= \Phi\left[\frac{\delta - \mu_2}{\sigma}\right] - \frac{\lambda_3}{6\sigma^2} H_2\left[\frac{\delta - \mu_2}{\sigma}\right] \phi\left(\frac{\delta - \mu_2}{\sigma}\right) \\
 &= \Phi\left[\frac{\mu_1 - \mu_2}{2\sigma}\right] - \frac{\lambda_3}{6\sigma^2} \left[\left(\frac{\mu_1 - \mu_2}{2\sigma}\right)^2 - 1\right] \phi\left(\frac{\mu_1 - \mu_2}{2\sigma}\right) .
 \end{aligned} \tag{4.27}$$

The optimum probability of misclassification is of interest for comparison purposes.

4.4 Conditional Probability of Misclassification

We discuss the conditional probability of misclassification when the parameters μ_1, μ_2 are unknown and all other parameters are known. The optimum probability of misclassification cannot be found here due to lack of information on μ_1, μ_2 . The classification regions are defined in (4.7) of section (4.1).

When an observation X from Π_1 is misclassified the

conditional probability of misclassification is

$$\begin{aligned}\alpha_1[\hat{R}, f] &= \Pr\{X \geq (\bar{x}_1 + \bar{x}_2)/2 \mid \bar{x}_1, \bar{x}_2, X \in \Pi_1\} \quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Pr\{X < (\bar{x}_1 + \bar{x}_2)/2 \mid \bar{x}_1, \bar{x}_2, X \in \Pi_1\} \quad \text{when } \bar{x}_1 \geq \bar{x}_2\end{aligned}\quad (4.28)$$

When an observation X from Π_2 is misclassified the conditional probability of misclassification is

$$\begin{aligned}\alpha_2[\hat{R}, f] &= \Pr\{X < (\bar{x}_1 + \bar{x}_2)/2 \mid \bar{x}_1, \bar{x}_2, X \in \Pi_2\} \quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Pr\{X \geq (\bar{x}_1 + \bar{x}_2)/2 \mid \bar{x}_1, \bar{x}_2, X \in \Pi_2\} \quad \text{when } \bar{x}_1 \geq \bar{x}_2\end{aligned}\quad (4.29)$$

Since the problem is symmetric in that the numbering of populations is arbitrary, we consider the probability $\alpha_1[\hat{R}, f]$.

$$\begin{aligned}\alpha_1[\hat{R}, f] &= \int_{\hat{\delta}}^{\infty} \left[1 - \frac{\lambda_3}{6} D^3\right] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \int_{-\infty}^{\hat{\delta}} \left[1 - \frac{\lambda_3}{6} D^3\right] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \quad \text{when } \bar{x}_1 \geq \bar{x}_2\end{aligned}$$

where $\hat{\delta} = (\bar{x}_1 + \bar{x}_2)/2$

which can be written as

$$\begin{aligned}\alpha_1[\hat{R}, f] &= \int_{\hat{\delta}}^{\infty} \phi\left(\frac{x - \mu_1}{\sigma}\right) dx + \frac{\lambda_3}{6\sigma^3} \int_{\hat{\delta}}^{\infty} H_3\left(\frac{x - \mu_1}{\sigma}\right) \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \\ &\quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \int_{-\infty}^{\hat{\delta}} \phi\left(\frac{x - \mu_1}{\sigma}\right) dx + \frac{\lambda_3}{6\sigma^3} \int_{-\infty}^{\hat{\delta}} H_3\left(\frac{x - \mu_1}{\sigma}\right) \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \\ &\quad \text{when } \bar{x}_1 \geq \bar{x}_2\end{aligned}$$

Letting $z = \frac{x - \mu_1}{\sigma}$, we obtain

$$\begin{aligned} \alpha_1[\hat{R}, f] &= \int_{(\hat{\delta} - \mu_1)/\sigma}^{\infty} \phi(z) dz + \frac{\lambda_3}{6\sigma^2} \int_{(\hat{\delta} - \mu_1)/\sigma}^{\infty} H_3(z) \phi(z) dz \\ &= \int_{-\infty}^{(\hat{\delta} - \mu_1)/\sigma} \phi(z) + \frac{\lambda_3}{6\sigma^2} \int_{-\infty}^{(\hat{\delta} - \mu_1)/\sigma} H_3(z) \phi(z) dz \\ &\quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &\quad \text{when } \bar{x}_1 \geq \bar{x}_2 \end{aligned} \quad (4.30)$$

and using (4.23)

$$\begin{aligned} \alpha_1[\hat{R}, f] &= 1 - \Phi\left[\frac{\hat{\delta} - \mu_1}{\sigma}\right] + \frac{\lambda_3}{6\sigma^2} H_2\left\{\frac{\hat{\delta} - \mu_1}{\sigma}\right\} \phi\left(\frac{\hat{\delta} - \mu_1}{\sigma}\right) \quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\hat{\delta} - \mu_1}{\sigma}\right] - \frac{\lambda_3}{6\sigma^2} H_2\left\{\frac{\hat{\delta} - \mu_1}{\sigma}\right\} \phi\left(\frac{\hat{\delta} - \mu_1}{\sigma}\right) \quad \text{when } \bar{x}_1 \geq \bar{x}_2 \end{aligned} \quad (4.31)$$

Upon substituting for $\hat{\delta}$ in (4.31)

$$\begin{aligned} \alpha_1[\hat{R}, f] &= 1 - \Phi\left[\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right] + \frac{\lambda_3}{6\sigma^2} H_2\left\{\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right\} \phi\left(\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right) \\ &\quad \text{when } \bar{x}_1 < \bar{x}_2 \\ &= \Phi\left[\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right] - \frac{\lambda_3}{6\sigma^2} H_2\left\{\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right\} \phi\left(\frac{\bar{x}_1 + \bar{x}_2 - 2\mu_1}{2\sigma}\right) \\ &\quad \text{when } \bar{x}_1 \geq \bar{x}_2 \end{aligned} \quad (4.32)$$

a) Distribution of The Conditional Probability of Misclassification

The distribution of the conditional probability of misclass-

ification in (4.32) is difficult to obtain due to the complex nature of the expression. It involves the cumulative normal distribution function, the normal density function and the Tchebycheff-Hermite polynomial. However the asymptotic distribution of the error can be found by the Taylor series expansion and the approximate mean and variance of the error function can be determined.

b) Taylor Series Expansion

The conditional probability of misclassification is a function of two random variables $\bar{X}_1$ and $\bar{X}_2$ when the parameters are fixed. This error function can be expanded in a Taylor series in two variables around $\{E(\bar{X}_1) = \mu_1, E(\bar{X}_2) = \mu_2\}$. For purposes of completeness of presentation the Taylor series theorem in two variables is stated below without proof.

Theorem 4.41: If all the n^{th} partial derivatives of a function of two variables $f(X,Y)$ are continuous in a closed region around (X_0, Y_0) and if the $(n+1)^{\text{st}}$ partial derivations exist in the open region around (X_0, Y_0) , then

$$\begin{aligned} f(X_0 + h, Y_0 + k) &= f(X_0, Y_0) + [h \frac{\partial}{\partial X} + k \frac{\partial}{\partial Y}] f(X_0, Y_0) + \\ &+ \frac{1}{2!} [h \frac{\partial}{\partial X} + k \frac{\partial}{\partial Y}]^2 f(X_0, Y_0) + \dots \\ &+ \frac{1}{n!} [h \frac{\partial}{\partial X} + k \frac{\partial}{\partial Y}]^n f(X_0, Y_0) + R_n \end{aligned} \quad (4.43)$$

where R_n is the remainder after n terms and is given by

$$R_n = \frac{1}{(n+1)!} (h \frac{\partial}{\partial X} + k \frac{\partial}{\partial Y})^{n+1} f(X_0 + \theta h, Y_0 + \theta k),$$

$$0 < \theta < 1$$

and $[h \frac{\partial}{\partial \bar{X}} + k \frac{\partial}{\partial \bar{Y}}]^n$ is formally expanded by the Binomial theorem.

Since $\alpha_1[\hat{R}, f]$ is a function of $\bar{X}_1$ and $\bar{X}_2$ we denote it by $e_{12}(\bar{X}_1, \bar{X}_2)$. Similarly $\alpha_2[\hat{R}, f]$ becomes $e_{21}(\bar{X}_1, \bar{X}_2)$. Then the expansion of $e_{12}(\bar{X}_1, \bar{X}_2)$ around (μ_1, μ_2) is given by

$$\begin{aligned} e_{12}(\bar{X}_1, \bar{X}_2) &= e_{12}(\mu_1, \mu_2) + [(\bar{X}_1 - \mu_1) \frac{\partial}{\partial \bar{X}_1} + (\bar{X}_2 - \mu_2) \frac{\partial}{\partial \bar{X}_2}]_{\mu_1, \mu_2} e_{12}(\bar{X}_1, \bar{X}_2) \\ &+ \frac{1}{2!} [(\bar{X}_1 - \mu_1) \frac{\partial}{\partial \bar{X}_1} + (\bar{X}_2 - \mu_2) \frac{\partial}{\partial \bar{X}_2}]^2_{\mu_1, \mu_2} e_{12}(\bar{X}_1, \bar{X}_2) + \\ &\dots + \frac{1}{n!} [(\bar{X}_1 - \mu_1) \frac{\partial}{\partial \bar{X}_1} + (\bar{X}_2 - \mu_2) \frac{\partial}{\partial \bar{X}_2}]^n_{\mu_1, \mu_2} e_{12}(\bar{X}_1, \bar{X}_2) + R_n \end{aligned} \quad (4.44)$$

Consider the case when $\bar{X}_1 < \bar{X}_2$ and write equation (4.32) as

$$e_{12}(\bar{X}_1, \bar{X}_2) = \phi\left[\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma}\right] + \frac{\lambda_3}{6\sigma^2} D^2 \phi\left(\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma}\right). \quad (4.45)$$

The partial derivatives with respect to $\bar{X}_1, \bar{X}_2$ of (4.45) are given by

$$\frac{\partial}{\partial \bar{X}_1} \{ e_{12}(\bar{X}_1, \bar{X}_2) \} = -\frac{1}{2\sigma} \phi\left(\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma}\right) + \frac{\lambda_3}{12\sigma^3} D^3 \phi\left(\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma}\right) \quad (4.46)$$

$$\begin{aligned} \frac{\partial^2}{\partial \bar{X}_1^2} \{ e_{12}(\bar{X}_1, \bar{X}_2) \} &= -\frac{1}{4\sigma^2} \left[\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma} \right] \phi\left(\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma}\right) + \\ &\frac{\lambda_3}{24\sigma^4} D^4 \phi\left(\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma}\right) \end{aligned} \quad (4.47)$$

$$\begin{aligned} \frac{\partial^2}{\partial \bar{X}_1 \partial \bar{X}_2} \{ e_{12}(\bar{X}_1, \bar{X}_2) \} = & - \frac{1}{4\sigma^2} \left[\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma} \right] \phi \left(\frac{2\mu_1 - \bar{X}_1 - \bar{X}_2}{2\sigma} \right) \\ & + \frac{\lambda_3}{24\sigma^4} D^4 \phi \left(\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right) \end{aligned} \quad (4.48)$$

Expanding terms involving D^2 , D^3 and D^4 in Tchebycheff-Hermite polynomials we obtain

$$D^2 \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] = \left\{ \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right]^2 - 1 \right\} \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] \quad (4.49)$$

$$D^3 \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] = \left\{ - \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right]^3 + 3 \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] \right\} \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] \quad (4.50)$$

$$\begin{aligned} D^4 \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] = & \left\{ \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right]^4 - 6 \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right]^2 \right. \\ & \left. + 3 \right\} \phi \left[\frac{\bar{X}_1 + \bar{X}_2 - 2\mu_1}{2\sigma} \right] . \end{aligned} \quad (4.51)$$

Evaluating these expressions at the point (μ_1, μ_2) and letting

$$\tau = (\mu_2 - \mu_1)/2\sigma \quad (4.52)$$

We obtain

$$\begin{aligned} e_{12}(\bar{X}_1, \bar{X}_2) \Big|_{\mu_1, \mu_2} &= \phi(-\tau) + \frac{\lambda_3}{6\sigma^2} [\tau^2 - 1] \phi(\tau) = \kappa_1 \\ \frac{\partial}{\partial \bar{X}_1} \{ e_{12}(\bar{X}_1, \bar{X}_2) \} \Big|_{\mu_1, \mu_2} &= - \frac{1}{2\sigma} \phi(-\tau) + \frac{\lambda_3}{12\sigma^3} [3\tau - \tau^3] \phi(\tau) = \kappa_2 \end{aligned} \quad (4.53)$$

$$\frac{\partial^2}{\partial \bar{X}_1^2} \{ e_{12}(\bar{X}_1, \bar{X}_2) \} \Big|_{\mu_1, \mu_2} = \frac{1}{4\sigma^2} \tau \phi(-\tau) + \frac{\lambda_3}{24\sigma^4} [\tau^4 - 6\tau^2 + 3] \phi(\tau) = \kappa_3 \quad (4.54)$$

$$\frac{\partial^2}{\partial \bar{X}_1 \partial \bar{X}_2} \{e_{12}(\bar{X}_1, \bar{X}_2)\} = \frac{1}{4\sigma^2} \tau \phi(-\tau) + \frac{\lambda_3}{24\sigma^4} [\tau^4 - 6\tau^2 + 3] \phi(\tau) = \kappa_3 \quad (4.55)$$

At the point (μ_1, μ_2) the following equalities hold

$$\frac{\partial}{\partial \bar{X}_1} \{e_{12}(\bar{X}_1, \bar{X}_2)\} = \frac{\partial}{\partial \bar{X}_2} \{e_{12}(\bar{X}_1, \bar{X}_2)\}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \bar{X}_1^2} \{e_{12}(\bar{X}_1, \bar{X}_2)\} &= \frac{\partial^2}{\partial \bar{X}_2^2} \{e_{12}(\bar{X}_1, \bar{X}_2)\} = \frac{\partial^2}{\partial \bar{X}_1 \partial \bar{X}_2} \{e_{12}(\bar{X}_1, \bar{X}_2)\} \\ &= \frac{\partial}{\partial \bar{X}_2 \partial \bar{X}_1} \{e_{12}(\bar{X}_1, \bar{X}_2)\} \end{aligned}$$

Neglecting terms containing partial derivatives of order higher than the second

$$\begin{aligned} e_{12}(\bar{X}_1, \bar{X}_2) &\approx \kappa_1 + [(\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2)] \kappa_2 + \frac{1}{2!} [(\bar{X}_1 - \mu_1)^2 + \\ &\quad + 2(\bar{X}_1 - \mu_1)(\bar{X}_2 - \mu_2) + (\bar{X}_2 - \mu_2)^2] \kappa_3 + R_2 \end{aligned} \quad (4.56)$$

When $\bar{X}_1 \geq \bar{X}_2$, the expansion of $e_{12}(\bar{X}_1, \bar{X}_2)$ is obtained in a similar form and the final result is given by

$$\begin{aligned} e_{12}(\bar{X}_1, \bar{X}_2) &\approx \kappa_1' + [(\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2)] \kappa_2' + \frac{1}{2!} [(\bar{X}_1 - \mu_1)^2 + \\ &\quad + 2(\bar{X}_1 - \mu_1)(\bar{X}_2 - \mu_2) + (\bar{X}_2 - \mu_2)^2] \kappa_3' + R_2' \end{aligned} \quad (4.57)$$

where

$$\kappa_1' = \phi(\tau) - \frac{\lambda_3}{6\sigma^2} [\tau^2 - 1] \phi(\tau) \quad (4.58)$$

$$\kappa'_2 = \frac{1}{2\sigma} \phi(\tau) - \frac{\lambda 3}{12\sigma^3} [3\tau - \tau^3] \phi(\tau) \quad (4.59)$$

$$\kappa'_3 = -\frac{1}{4\sigma^2} \tau \phi(\tau) - \frac{\lambda 3}{24\sigma^4} [\tau^4 - 6\tau^2 + 3] \phi(\tau) \quad (4.60)$$

c) Approximate Mean and Variance

The approximate mean and variance of the conditional probability of misclassification are obtained from equations (4.56) and (4.57) by use of a theorem which we state without proof below.

Theorem 4.42: Let (X, Y) be a two-dimensional random variable. Suppose

$$E(X) = \mu_X, \quad E(Y) = \mu_Y$$

and

$$V(X) = \sigma_X^2, \quad V(Y) = \sigma_Y^2$$

Let $Z = H(X, Y)$

and assume that the various derivatives of H exist at the point (μ_X, μ_Y) . Then $E(Z)$ and $V(Z)$ are given by

$$E[Z] \approx H(\mu_X, \mu_Y) + \frac{\partial^2 H}{\partial X^2} \sigma_X^2 + \frac{\partial^2 H}{\partial Y^2} \sigma_Y^2 \quad (4.61)$$

$$V[Z] \approx \left\{ \frac{\partial H}{\partial X} \right\}^2 \sigma_X^2 + \left\{ \frac{\partial H}{\partial Y} \right\}^2 \sigma_Y^2, \quad (4.62)$$

where all partial derivatives are evaluated at the point (μ_X, μ_Y) . Refer to Meyer [43, p. 128], for example.

Let

$$\begin{aligned} T &= (\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2) \\ &= (\bar{X}_1 + \bar{X}_2) - (\mu_1 + \mu_2) \end{aligned} \quad (4.63)$$

and let $\bar{X}_1 < \bar{X}_2$. Then from equation (4.56)

$$e_{12}(\bar{X}_1, \bar{X}_2) \approx \kappa_1 + \kappa_2 T + \kappa_3^{(1)} T^2 + R_2 \quad (4.64)$$

where $\kappa_3^{(1)} = \kappa_3/2$. Now

$$\begin{aligned} E(T) &= 0, \quad V(T) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \\ \text{and} \quad V(T^2) &= E[T^4] - E^2(T^2) \\ &= 3\sigma^4 \left[\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right]^2 - \left[\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right]^2 \end{aligned} \quad (4.65)$$

Thus the approximate mean of $e_{12}(\bar{X}_1, \bar{X}_2)$ is given by

$$\begin{aligned} E\{e_{12}(\bar{X}_1, \bar{X}_2)\} &\approx \kappa_1 + \kappa_2 E(T) + \kappa_3^{(1)} E(T^2) \\ &= \kappa_1 + \kappa_3^{(1)} \left[\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right] \end{aligned} \quad (4.66)$$

and the approximate variance of $e_{12}(\bar{X}_1, \bar{X}_2)$ is given by

$$\begin{aligned} V\{e_{12}(\bar{X}_1, \bar{X}_2)\} &\approx \kappa_2^2 E(T^2) + \kappa_3^{(1)2} E[T^2 - \sigma^2 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}]^2 \\ &\quad + 2\kappa_2 \kappa_3^{(1)} E\left(T [T^2 - \sigma^2 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}]\right) \\ &= \kappa_2^2 E(T^2) + \kappa_3^{(1)2} V(T^2) + 2\kappa_2 \kappa_3^{(1)} E(T^3) \\ &\quad - 2\kappa_2 \kappa_3^{(1)} \left[\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right] E(T) \end{aligned} \quad (4.67)$$

and using the results in (4.65) and (4.21) we get

$$\begin{aligned}
V\{e_{12}(\bar{X}_1, \bar{X}_2)\} &\approx \kappa_2^2 \left[\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right] + \kappa_3^{(1)2} \left[3\sigma^4 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}^2 \right. \\
&\quad \left. - \sigma^4 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}^2 \right] + 2\kappa_2 \kappa_3^{(1)} \lambda_3 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \\
&\quad (4.68)
\end{aligned}$$

A similar procedure yields the expectation and variance of $e_{12}(\bar{X}_1, \bar{X}_2)$ when $\bar{X}_1 \geq \bar{X}_2$. Therefore

$$\begin{aligned}
E\{e_{12}(\bar{X}_1, \bar{X}_2)\} &\approx \kappa_1 + \kappa_3^{(1)} \sigma^2 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \quad \text{when } \bar{X}_1 < \bar{X}_2 \\
&\quad (4.69)
\end{aligned}$$

$$\begin{aligned}
&\approx \kappa_1' + \kappa_3'' \sigma^2 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \quad \text{when } \bar{X}_1 \geq \bar{X}_2, \\
&\quad (4.70)
\end{aligned}$$

where $\kappa_3'' = \kappa_3'/2$

and

$$\begin{aligned}
V\{e_{12}(\bar{X}_1, \bar{X}_2)\} &\approx \kappa_2^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\} + \kappa_3^{(1)2} 2\sigma^4 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}^2 \\
&\quad + 2\kappa_2 \kappa_3^{(1)} \lambda_3 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \quad \text{when } \bar{X}_1 < \bar{X}_2 \\
&\quad (4.71)
\end{aligned}$$

$$\begin{aligned}
&\approx \kappa_2'^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\} + \kappa_3''^2 2\sigma^4 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}^2 \\
&\quad + 2\kappa_2' \kappa_3'' \lambda_3 \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \quad \text{when } \bar{X}_1 \geq \bar{X}_2. \\
&\quad (4.72)
\end{aligned}$$

d) Asymptotic Distribution

We consider the asymptotic distribution of $e_{12}(\bar{X}_1, \bar{X}_2)$, since the problem is symmetric with respect to the two errors.

As n_1, n_2 become large the distribution of

$$y = \frac{e_{12}(\bar{X}_1, \bar{X}_2) - E\{e_{12}(\bar{X}_1, \bar{X}_2)\}}{[V\{e_{12}(\bar{X}_1, \bar{X}_2)\}]^{\frac{1}{2}}} \quad (4.73)$$

is approximately normal with mean 0 and variance unity.

Ignoring terms of order $\frac{1}{n_1^2}$, $\frac{1}{n_2^2}$ and $\frac{1}{n_1 n_2}$ in the variance approximation [see equations (4.69), (4.70)], we obtain

$$\begin{aligned} \Pr\{e_{12}(\bar{X}_1, \bar{X}_2) < u\} &\approx \\ &\approx \Pr\left\{ \frac{e_{12}(\bar{X}_1, \bar{X}_2) - [\kappa_1 + \kappa_3] \frac{(1)^2}{2} \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}}{[\kappa_2^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}]^{\frac{1}{2}}} < \frac{u - [\kappa_1 + \kappa_3] \frac{(1)^2}{2} \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}}{[\kappa_2^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}]^{\frac{1}{2}}} \right\} \\ &= \Phi \left[\frac{u - [\kappa_1 + \kappa_3] \frac{(1)^2}{2} \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}}{[\kappa_2^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}]^{\frac{1}{2}}} \right] \quad \text{when } \bar{X}_1 < \bar{X}_2 \end{aligned}$$

(4.74)

and when $\bar{X}_1 \geq \bar{X}_2$ we obtain

$$\Pr\{e_{12}(\bar{X}_1, \bar{X}_2) < u\} = \Phi \left[\frac{u - [\kappa_1' + \kappa_3] \frac{(1)^2}{2} \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}}{[\kappa_2'^2 \left\{ \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \right\}]^{\frac{1}{2}}} \right] \quad (4.75)$$

4.5 Estimated Probability of Misclassification

When parameters of a discriminant function are unknown the usual technique is to replace them by their maximum likelihood estimates. This procedure is unsatisfactory in that

it leads to the errors being either over - or under-estimated. If $\bar{x}_1$ is an estimate of μ_1 , $\bar{x}_2$ is an estimate of μ_2 and the common variance between the two populations π_1, π_2 is estimated by the pooled variance s^2 , then from (4.25) or (4.32) the plug-in estimate of the error is given by

$$\alpha_1[\hat{R}, \hat{f}] = \phi\left[\frac{\bar{x}_1 - \bar{x}_2}{2s}\right] + \frac{\lambda_3}{6s^2} \left[\left(\frac{\bar{x}_2 - \bar{x}_1}{2s}\right)^2 - 1 \right] \phi\left(\frac{\bar{x}_2 - \bar{x}_1}{2s}\right) \quad (4.76)$$

and from (4.27)

$$\alpha_2[\hat{R}, \hat{f}] = \phi\left[\frac{\bar{x}_1 - \bar{x}_2}{2s}\right] - \frac{\lambda_3}{6s^2} \left[\left(\frac{\bar{x}_1 - \bar{x}_2}{2s}\right)^2 - 1 \right] \phi\left(\frac{\bar{x}_1 - \bar{x}_2}{2s}\right), \quad (4.77)$$

where λ_3 in (4.76) and (4.77) is assumed known.

4.6 Unconditional Probability of Misclassification

a) Characteristic Function Approach

From (3.36) the unconditional probability of misclassification for $e_{12}(\bar{X}_1, \bar{X}_2)$ is given by

$$\begin{aligned} E[e_{12}(\bar{X}_1, \bar{X}_2)] &= \Pr\{\bar{X}_2 - \bar{X}_1 > 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \geq 0 | X \in \pi_1\} \\ &\quad + \Pr\{\bar{X}_2 - \bar{X}_1 \leq 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) < 0 | X \in \pi_1\} \end{aligned} \quad (4.78)$$

To evaluate this probability of misclassification we need the joint distribution of $\bar{X}_2 - \bar{X}_1$ and $X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$, which we derive below using the characteristic function approach as in section (2.2) of Chapter 2. Assume that X is from

population Π_1 and has the ESD with parameters μ_1 , σ_1^2 and λ_3 ; where μ_1 is the population mean, σ_1^2 is the population variance and λ_3 is the population skewness factor.

From (4.4) the characteristic function of X is given by

$$\psi_x(t) = [1 + \frac{\lambda_3}{6}(it)^3] \exp[it\mu_1 - (\sigma_1 t)^2/2] \quad (4.79)$$

and the characteristic function of $\bar{X}_j$ is given by

$$\psi_{\bar{X}_j}(t) = [1 + \frac{\lambda_3}{6n_j}(it)^3] \exp[-\mu_j it - \sigma_j^2 t^2/2n_j], \quad j=1,2 \quad (4.80)$$

[see Lemma 4.2, of chapter 4]

In (4.80) it is assumed that the variances of X in the two populations are different and are given by σ_1^2 in Π_1 and σ_2^2 in Π_2 .

Let

$$U = \bar{X}_2 - \bar{X}_1$$

$$V = X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$$

Then the joint characteristic function of U and V is given by

$$\begin{aligned} \psi_{uv}(t_1 t_2) &= E\{e^{it_1 U} e^{it_2 V}\} \\ &= E\{e^{it_1(\bar{X}_2 - \bar{X}_1)} e^{it_2(X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2))}\} \\ &= E\{e^{it_1 \bar{X}_2 - it_1 \bar{X}_1 + it_2 X - \frac{it_2}{2} \bar{X}_1 - \frac{it_2}{2} \bar{X}_2}\} \\ &= E\{e^{it_2 X + [it_1 - (it_2/2)]\bar{X}_2 + [-it_1 - (it_2/2)]\bar{X}_1}\} \end{aligned}$$

$$= \psi_x(t_2) \psi_{\bar{x}_2}(t_1 - \{t_2/2\}) \psi_{\bar{x}_1}(-t_1 - \{t_2/2\}) \quad (4.81)$$

since $x, \bar{x}_1, \bar{x}_2$ are independent.

Substituting for the various terms

$$\begin{aligned} \psi_{uv}(t_1, t_2) = & \left\{ \left[1 + \frac{\lambda_3}{6} (it_2)^3 \right] \left[1 + \frac{\lambda_3}{6n_2} (it_1 - \{it_2/2\})^3 \right] \right. \\ & \cdot \left. \left[1 + \frac{\lambda_3}{6n_1} (-it_1 - \{it_2/2\})^3 \right] \right\} \\ & \cdot \exp[\mu_1 it_2 - t_2^2 \sigma_1^2 / 2 + \mu_2 it_1 - \mu_2 it_2 / 2 - \\ & \sigma_2^2 (t_1 - t_2/2)^2 / 2n_2 \\ & - \mu_1 it_1 - \mu_1 it_2 / 2 - \sigma_1^2 [t_1 + t_2/2]^2 / 2n_1] \end{aligned} \quad (4.82)$$

Let A be the quantity in $\{ \}$, and B be the exponent then

$$\begin{aligned} B = & \{ \mu_2 - \mu_1 \} it_1 + \left\{ \frac{\mu_1 - \mu_2}{2} \right\} it_2 - \frac{t_1^2}{2} \left\{ \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1} \right\} - \\ & \frac{t_2^2}{2} \left\{ \sigma_1^2 + \frac{\sigma_1^2}{4n_1} + \frac{\sigma_2^2}{4n_2} \right\} \\ & - \frac{t_1 t_2}{2} \left\{ \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2} \right\} \end{aligned} \quad (4.83)$$

where as

$$\begin{aligned} A = & \left[1 + \frac{\lambda_3}{6} (it_2)^3 \right] \left[1 + \frac{\lambda_3}{6n_2} (it_1 - \{it_2/2\})^3 \right] \\ & \left[1 + \frac{\lambda_3}{6n_1} (-it_1 - \{it_2/2\})^3 \right] \end{aligned}$$

$$= 1 + \frac{\lambda^3}{6} [(it_2)^3 + \frac{(it_1 - \{it_2/2\})^3}{n_1^2} + \frac{(-it_1 - \{it_2/2\})^3}{n_2^2}] \quad (4.84)$$

Thus the characteristic function of U and V is written

$$\psi_{uv}(t_1, t_2) = A \exp B \quad (4.85)$$

Now the characteristic function of (4.85) is that of the bivariate ESD, with means, variances and correlation coefficient between U and V given by

Means

$$\left. \begin{aligned} \mu_u &= (\mu_2 - \mu_1) \\ \mu_v &= (\mu_1 - \mu_2)/2 \end{aligned} \right\} \quad (4.86)$$

Variances

$$\left. \begin{aligned} \sigma_u^2 &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \\ \sigma_v^2 &= \sigma_1^2 + \frac{\sigma_2^2}{4n_2} + \frac{\sigma_1^2}{4n_1} \end{aligned} \right\} \quad (4.87)$$

Correlation coefficient

$$\rho_{uv} = \frac{1}{2} \left\{ \frac{\sigma_1^2}{n_1} - \frac{\sigma_1^2}{n_2} \right\} / \left\{ \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right] \left[\sigma_1^2 + \frac{\sigma_1^2}{4n_1} + \frac{\sigma_2^2}{4n_2} \right] \right\}^{1/2}, \quad (4.88)$$

where

$$\mu_1 = E(X | X \in \Pi_1)$$

$$\mu_2 = E(X | X \in \Pi_2)$$

$$\sigma_1^2 = V(X|X \in \Pi_1)$$

$$\sigma_2^2 = V(X|X \in \Pi_2).$$

Mardia ([34], Chapter 3) has an excellent discussion of this distribution. As n_1, n_2 the sample sizes from the populations Π_1 and Π_2 respectively, become large, dropping out terms of order n^{-2} and higher, reduces the characteristic function (4.85) to

$$\begin{aligned} \psi_{uv}(t_1, t_2) = [1 + \frac{\lambda_3}{6}(it_2)^3] \exp \mu_u it_1 + \mu_v it_2 - \\ \frac{t_1^2}{2} \sigma_u^2 - \frac{t_2^2}{2} \sigma_v^2 - t_1 t_2 \sigma_{uv} \end{aligned} \quad (4.89)$$

where

$$\sigma_{uv} = \frac{1}{2} \left\{ \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2} \right\},$$

showing that the characteristic function is still that of the bivariate ESD.

b) Conditional Expectation Approach

If $\lambda_3 = 0$, the characteristic function (4.85) or (4.89) is that of the bivariate normal density function. If $\lambda_3 \neq 0$, the characteristic function is that of the bivariate ESD.

Let

$$U = \bar{X}_2 - \bar{X}_1$$

$$V^* = X - \frac{1}{2}(\bar{X}_1 - \bar{X}_2)$$

An alternative way of finding the unconditional probability of misclassification for $e_{12}(\bar{X}_1, \bar{X}_2)$ is through the asymptotic joint probability density function of U and V , where

$$\begin{aligned} U &= \bar{X}_2 - \bar{X}_1 \\ V &= \{\bar{X}_1 + \bar{X}_2\}/2, \end{aligned} \tag{4.90}$$

and the use of conditional expectation.

Let

$$\begin{aligned} Q &= E\{e_{12}(\bar{X}_1, \bar{X}_2)\} \\ &= \Pr\{\bar{X}_2 - \bar{X}_1 > 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \geq 0 | X \in \Pi_1\} \\ &\quad + \Pr\{\bar{X}_2 - \bar{X}_1 \leq 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) < 0 | X \in \Pi_2\} \end{aligned} \tag{4.91}$$

Consider the term

$$\begin{aligned} p &= \Pr\{\bar{X}_2 - \bar{X}_1 > 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \geq 0 | X \in \Pi_1\} \\ &= \int_{-\infty}^{\infty} \Pr\{\bar{X}_2 - \bar{X}_1 \geq 0, \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \leq x | x\} g(x) dx \end{aligned} \tag{4.92}$$

where

$$g(x) = [1 - \frac{\lambda_3}{6} D^3] \phi\left(\frac{x - \mu_1}{\sigma_1}\right), \quad -\infty < x < \infty.$$

Writing U and V as in (4.90) for large n_1, n_2 , the joint p.d.f of U and V is given by

$$\begin{pmatrix} U \\ V \end{pmatrix} \approx \text{BVN} \left[\begin{pmatrix} \mu_u \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \right]$$

[The justification of this result is given in section 3.3 of chapter 3].

Where

$$\left. \begin{aligned} \mu_u &= \mu_2 - \mu_1, \\ \mu_v &= \frac{1}{2}(\mu_1 + \mu_2), \\ \sigma_u^2 &= \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}, \\ \sigma_v^2 &= \frac{1}{4} \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) = \frac{\sigma_u^2}{4}, \\ \sigma_{uv} &= \frac{1}{2} \left(\frac{\sigma_2^2}{n_2} - \frac{\sigma_1^2}{n_1} \right). \end{aligned} \right\} \quad (4.93)$$

Therefore

$$p = \frac{k}{\sigma_1 \sqrt{2\pi}} \int_{-\infty}^{\infty} \int_0^{\infty} \int_{-\infty}^x \left[1 + \frac{\lambda_3}{6\sigma_1^2} H_3\left(\frac{x-\mu_1}{\sigma}\right) \right] \exp\left\{ -\frac{1}{2(1-\rho^2)} \right. \\ \left. \left[\frac{(u-\mu_u)^2}{\sigma_u^2} + \frac{(v-\mu_v)^2}{\sigma_v^2} - \frac{2\rho\{u-\mu_u\}\{v-\mu_v\}}{\sigma_u\sigma_v} \right] - \frac{(x-\mu_1)^2}{2\sigma_1^2} \right\} du dv dx \quad (4.94)$$

where

$$k = \frac{1}{2\pi \sigma_u \sigma_v (1-\rho^2)^{\frac{1}{2}}} \quad (4.95)$$

Recall here that

$$\begin{aligned}
\left[1 - \frac{\lambda_3}{6} D^3\right] \phi\left(\frac{x-\mu_1}{\sigma_1}\right) &= \phi\left(\frac{x-\mu_1}{\sigma_1}\right) - \frac{\lambda_3}{6\sigma^3} \left[3\left\{\frac{x-\mu_1}{\sigma}\right\} - \left(\frac{x-\mu_1}{\sigma}\right)^3\right] \phi\left(\frac{x-\mu_1}{\sigma}\right) \\
&= \left[1 + \frac{\lambda_3}{6\sigma^3} H_3\left(\frac{x-\mu_1}{\sigma}\right)\right] \phi\left(\frac{x-\mu_1}{\sigma}\right)
\end{aligned}
\tag{4.96}$$

changing variable v to w as $v = x - w$, we have

$$\begin{aligned}
p &= \frac{k}{\sigma_1 \sqrt{2\pi}} \int_{-\infty}^{\infty} \int_0^{\infty} \int_0^{\infty} \left[1 + \frac{\lambda_3}{6\sigma_1^3} H_3\left(\frac{x-\mu_1}{\sigma_1}\right)\right] \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{(u-\mu_u)^2}{\sigma_u^2} \right. \right. \\
&\quad \left. \left. + \frac{(x-w-\mu_v)^2}{\sigma_v^2} - \frac{2\rho\{u-\mu_u\}\{x-w-\mu_v\}}{\sigma_u \sigma_v}\right] - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} dw du dx
\end{aligned}
\tag{4.97}$$

Now by changing the order of integration we get

$$\begin{aligned}
p &= \frac{k}{\sigma_1 \sqrt{2\pi}} \int_0^{\infty} \int_0^{\infty} \int_{-\infty}^{\infty} \left[1 + \frac{\lambda_3}{6\sigma_1^3} H_3\left(\frac{x-\mu_1}{\sigma_1}\right)\right] \exp\left\{-\frac{1}{2(1-\rho^2)} \right. \\
&\quad \left. \left[\frac{(u-\mu_u)^2}{\sigma_u^2} + \frac{(x-w-\mu_v)^2}{\sigma_v^2} - \frac{2\rho\{u-\mu_u\}\{x-w-\mu_v\}}{\sigma_u \sigma_v}\right] - \frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} dx du dw
\end{aligned}
\tag{4.98}$$

Putting

$$t = x - \mu_1,$$

$$\delta = (\mu_1 - \mu_2)/2$$

this becomes

$$p = \frac{k}{\sigma_1 \sqrt{2\pi}} \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \left[1 + \frac{\lambda_3}{6\sigma_1^3} H_3\left(\frac{t}{\sigma_1}\right) \right] \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(u-\mu_u)^2}{\sigma_u^2} + \frac{(t-w+\delta)^2}{\sigma_v^2} - \frac{2\rho\{u-\mu_u\}\{v-\mu_v\}}{\sigma_u\sigma_v} \right] - \frac{t^2}{2\sigma_1^2} \right\} dt du dw \quad (4.99)$$

Putting $x = u - \mu_u$, $y = (w - \delta)$, the exponent of the integrand in (4.99) can be written as $-1/2$ times

$$\begin{aligned} & \frac{1}{1-\rho^2} \left[\frac{x^2}{\sigma_u^2} + \frac{(t-y)^2}{\sigma_v^2} - \frac{2\rho x(t-y)}{\sigma_u\sigma_v} \right] + \frac{t^2}{\sigma_1^2} \\ &= \frac{x^2}{\sigma_u^2(1-\rho^2)} + t^2 \left\{ \frac{1}{\sigma_1^2} + \frac{1}{\sigma_v^2(1-\rho^2)} \right\} + \frac{y^2}{\sigma_v^2(1-\rho^2)} - \\ & \quad 2t \left\{ \frac{y}{\sigma_v^2(1-\rho^2)} + \frac{\rho x}{\sigma_u\sigma_v(1-\rho^2)} + \frac{2\rho xy}{\sigma_u\sigma_v(1-\rho^2)} \right\} \\ &= \frac{x^2}{\sigma_u^2(1-\rho^2)} + \frac{2\rho xy}{\sigma_u\sigma_v(1-\rho^2)} + \frac{y^2}{\sigma_v^2(1-\rho^2)} + a(t - \frac{b}{a})^2 - \frac{b^2}{a} \end{aligned} \quad (4.100)$$

In equation (4.100)

$$a = \left\{ \frac{1}{\sigma_v^2(1-\rho^2)} + \frac{1}{\sigma_1^2} \right\} = \frac{\sigma_1^2 + \sigma_v^2(1-\rho^2)}{\sigma_1^2\{\sigma_v^2(1-\rho^2)\}} \quad (4.101)$$

$$b = \left\{ \frac{y}{\sigma_v^2(1-\rho^2)} + \frac{\rho x}{\sigma_u\sigma_v(1-\rho^2)} \right\} \quad (4.102)$$

Since

$$\begin{aligned}
 \frac{b^2}{a} &= \frac{\left\{ \frac{y}{\sigma_v^2(1-\rho^2)} + \frac{\rho x}{\sigma_u \sigma_v(1-\rho^2)} \right\}^2 \sigma_1^2 \sigma_v^2(1-\rho^2)}{\sigma_1^2 + \sigma_v^2(1-\rho^2)} \\
 &= \frac{y^2 \sigma_1^2}{\sigma_1^2 + \sigma_v^2(1-\rho^2)} + \frac{2\rho xy \sigma_1^2}{\sigma_v \sigma_u(1-\rho^2) \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \\
 &\quad + \frac{\rho^2 x^2 \sigma_1^2}{\sigma_u^2(1-\rho^2) \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \quad (4.103)
 \end{aligned}$$

(4.100) becomes

$$\begin{aligned}
 &= a \left(t - \frac{b}{a} \right)^2 + \frac{y^2}{\sigma_1^2 + \sigma_v^2(1-\rho^2)} + \frac{x^2 \{ (\sigma_1^2 + \sigma_v^2)(1-\rho^2) \}}{\sigma_u^2(1-\rho^2) \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \\
 &\quad + \frac{2xy\rho \sigma_v^2(1-\rho^2)}{\sigma_u \sigma_v(1-\rho^2) \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \quad (4.104)
 \end{aligned}$$

$$\begin{aligned}
 &= a \left(t - \frac{b}{a} \right)^2 + \frac{y^2}{\sigma_1^2 + \sigma_v^2(1-\rho^2)} + \frac{x^2 (\sigma_1^2 + \sigma_v^2)}{\sigma_u^2 \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \\
 &\quad + \frac{2xy \sigma_v \rho}{\sigma_u \{ \sigma_1^2 + \sigma_v^2(1-\rho^2) \}} \quad (4.105)
 \end{aligned}$$

We would like to express the parameters involved in the above expressions in terms of $V(X)$, $V(Y)$ and $\text{Cov}(X,Y)$ where X and Y are defined earlier as

$$\left. \begin{aligned} X &= U - \mu_u \\ Y &= X - V - \delta \end{aligned} \right\} \quad (4.106)$$

Therefore

$$\left. \begin{aligned} V(X) &= \sigma_u^2 \\ V(Y) &= \sigma_1^2 + \sigma_u^2 \\ \text{cov}(X, Y) &= -\rho \sigma_u \sigma_v = \text{cov}(U, V). \end{aligned} \right\} \quad (4.107)$$

Define

$$\rho^* = \frac{-\rho \sigma_u \sigma_v}{\sigma_u [\sigma_1^2 + \sigma_v^2]^{\frac{1}{2}}} = \frac{\text{cov}(X, Y)}{[V(X) V(Y)]^{\frac{1}{2}}} \quad (4.108)$$

$$\rho^{*2} = \frac{\rho^2 \sigma_u^2 \sigma_v^2}{\sigma_u^2 [\sigma_1^2 + \sigma_v^2]} \quad (4.109)$$

and

$$1 - \rho^{*2} = 1 - \frac{\rho^2 \sigma_u^2 \sigma_v^2}{\sigma_u^2 \{\sigma_1^2 + \sigma_v^2\}} = \frac{\sigma_1^2 + \sigma_v^2 (1 - \rho^2)}{\sigma_1^2 + \sigma_v^2} \quad (4.110)$$

Therefore

$$(1 - \rho^{*2}) \{\sigma_1^2 + \sigma_v^2\} = \sigma_1^2 + \sigma_v^2 (1 - \rho^2). \quad (4.111)$$

From (4.105) the coefficients of y^2 , x^2 and xy are respectively given by

$$\frac{1}{(1 - \rho^{*2}) \sigma_y^2} \quad (4.112)$$

$$\frac{1}{\sigma_u^2 (1-\rho^{*2})} \quad (4.113)$$

$$\frac{-\rho^*}{\sigma_x \sigma_y (1-\rho^{*2})} \quad (4.114)$$

Therefore the exponent of the integrand in (4.99) is given by

$$-\frac{1}{2} a(t - \frac{b}{a})^2 - \frac{1}{2(1-\rho^{*2})} \left[\frac{y^2}{\sigma_y^2} + \frac{x^2}{\sigma_x^2} - \frac{2\rho^* xy}{\sigma_x \sigma_y} \right] \quad (4.115)$$

and the triple integral in (4.99) is given by

$$\begin{aligned} p &= \frac{k}{\sigma_1 \sqrt{2\pi}} \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \left[1 + \frac{\lambda_3}{6\sigma_1} H_3\left(\frac{t}{\sigma_1}\right) \right] \exp\left\{-\frac{1}{2} a(t - \frac{b}{a})^2\right. \\ &\quad \left. - \frac{1}{2(1-\rho^{*2})} \left[\frac{y^2}{\sigma_y^2} + \frac{x^2}{\sigma_x^2} - \frac{2\rho^* xy}{\sigma_x \sigma_y} \right] \right\} dt dx dy \\ &= \frac{k}{\sigma_1 \sqrt{2\pi}} \int_0^\infty \int_0^\infty \left\{ \int_{-\infty}^\infty \left[1 + \frac{\lambda_3}{6\sigma_1} H_3\left(\frac{t}{\sigma_1}\right) \right] \exp\left\{-\frac{1}{2} a(t - \frac{b}{a})^2\right\} dt \right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2(1-\rho^{*2})} \left[\frac{y^2}{\sigma_y^2} + \frac{x^2}{\sigma_x^2} - \frac{2\rho^* xy}{\sigma_x \sigma_y} \right] \right\} dx dy \quad (4.116) \end{aligned}$$

Now consider

$$\begin{aligned} p' &= \int_{-\infty}^\infty \left[1 + \frac{\lambda_3}{6\sigma_1} H_3\left(\frac{t}{\sigma_1}\right) \right] \exp\left\{-\frac{1}{2} a(t - \frac{b}{a})^2\right\} dt \\ &= \int_{-\infty}^\infty \exp\left\{-\frac{1}{2} a(t - \frac{b}{a})^2\right\} dt + \int_{-\infty}^\infty \frac{\lambda_3}{6\sigma_1} H_3\left(\frac{t}{\sigma_1}\right) \exp\left\{-\frac{1}{2} a(t - \frac{b}{a})^2\right\} dt \quad (4.117) \end{aligned}$$

It is well known that the second term of (4.117)

$$\int_{-\infty}^{\infty} \exp\{-\frac{1}{2} a(t - \frac{b}{a})^2\} dt = \frac{\sqrt{2\pi}}{\sqrt{a}} \quad (4.118)$$

also setting $k_2 = \lambda_3/6\sigma_1^3$, the third term of (4.117) becomes

$$\begin{aligned} p'' &= k_2 \int_{-\infty}^{\infty} H_3\left(\frac{t}{\sigma_1}\right) \exp\{-\frac{1}{2}(\sqrt{a}t - \frac{b}{\sqrt{a}})^2\} dt \\ &= k_2 \int_{-\infty}^{\infty} \left[\left(\frac{t}{\sigma_1}\right)^3 - 3\left(\frac{t}{\sigma_1}\right)\right] \exp\{-\frac{1}{2}(\sqrt{a}t - \frac{b}{\sqrt{a}})^2\} dt \end{aligned}$$

$$\text{Let } s = \sqrt{a}t - \frac{b}{\sqrt{a}}$$

then

$$\begin{aligned} p'' &= \frac{k_2}{\sqrt{a}} \int_{-\infty}^{\infty} \left\{ \frac{\left(\frac{s}{\sqrt{a}} + \frac{b}{a}\right)^3}{\sigma_1^3} - \frac{3\left(\frac{s}{\sqrt{a}} + \frac{b}{a}\right)}{\sigma_1} \right\} \exp\left(-\frac{1}{2}s^2\right) ds \\ &= \frac{k_2}{\sqrt{a}} \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_1^3} \left\{ \frac{s^3}{a^{3/2}} + \frac{3s^2b}{a^2} + \frac{3sb^2}{a^{5/2}} + \frac{b^3}{a^3} \right\} \right. \\ &\quad \left. - \frac{1}{\sigma_1} \left\{ \frac{3s}{\sqrt{a}} + \frac{3b}{a} \right\} \right] \exp\left(-\frac{1}{2}s^2\right) ds \\ &= \frac{k_2 b \sqrt{2\pi}}{\sigma_1 a^{3/2}} \left[\frac{3}{a\sigma_1^2} + \frac{b^2}{\sigma_1^2 a^2} - 3 \right] . \end{aligned} \quad (4.119)$$

Therefore since, $k_2 = \lambda_3/6\sigma_1^3$

$$p'' = \frac{\lambda_3 b \sqrt{2\pi}}{6\sigma_1^4 a^{3/2}} \left[\frac{3}{a\sigma_1^2} + \frac{b^2}{\sigma_1^2 a^2} - 3 \right] . \quad (4.120)$$

Now

$$\begin{aligned}\frac{k}{\sigma_1 \sqrt{a}} &= \frac{\sigma_1 \sigma_v (1-\rho^2)^{\frac{1}{2}}}{\sigma_1 2\pi \sigma_u \sigma_v (1-\rho^2)^{\frac{1}{2}} \{\sigma_1^2 + \sigma_v^2 (1-\rho^2)\}^{\frac{1}{2}}} \\ &= \frac{1}{2\pi \sigma_u \{\sigma_1^2 + \sigma_v^2 (1-\rho^2)\}^{\frac{1}{2}}}\end{aligned}$$

since

$$\sigma_1^2 + \sigma_v^2 (1-\rho^2) = (1-\rho^{*2}) (\sigma_1^2 + \sigma_v^2) = \sigma_y^2 (1-\rho^{*2})$$

and

$$\sigma_u^2 = \sigma_x^2$$

$$\frac{k}{\sigma_1 \sqrt{a}} = \frac{1}{2\pi \sigma_x \sigma_y (1-\rho^{*2})^{\frac{1}{2}}} \quad (4.121)$$

Finally

$$\begin{aligned}p' &= \frac{\sqrt{2\pi}}{\sqrt{a}} + p'' \\ &= \frac{\sqrt{2\pi}}{\sqrt{a}} + \frac{\lambda_3 b \sqrt{2\pi}}{6\sigma_1^4 a^{3/2}} \left[\frac{3}{a\sigma_1^2} + \frac{b^2}{\sigma_1^2 a^2} - 3 \right] \quad (4.122)\end{aligned}$$

and

$$\begin{aligned}p &= \frac{1}{2\pi \sigma_x \sigma_y (1-\rho^{*2})^{\frac{1}{2}}} \int_0^\infty \int_0^\infty p' \exp\left\{-\frac{1}{2(1-\rho^{*2})} \left[\frac{y^2}{\sigma_y^2} + \frac{x^2}{\sigma_x^2} - \frac{2\rho^{*} xy}{\sigma_x \sigma_y} \right]\right\} dx dy. \quad (4.123)\end{aligned}$$

Now the last term of (4.91) is given by

$$\begin{aligned}q &= \Pr\{\bar{X}_2 - \bar{X}_1 \leq 0, X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) < 0 | X \in \Pi_1\} \\ &= \int_{-\infty}^\infty \Pr\{\bar{X}_2 - \bar{X}_1 \leq 0, \frac{1}{2}(\bar{X}_1 + \bar{X}_2) > x | x\} g(x) dx, \quad (4.124)\end{aligned}$$

where $g(x)$ is given in (4.92).

By the same procedure (4.124) is evaluated and gives a similar result to that of (4.123). Therefore the final result is

$$\begin{aligned} Q &= E\{e_{12}(\bar{X}_1, \bar{X}_2)\} \\ &= [p + q] \end{aligned} \tag{4.125}$$

Further study is proposed for examining the expected error of misclassification (4.125).

CHAPTER 5

ESTIMATION, COMPARISONS AND RECOMMENDATIONS

5.1 Introduction

Most of the discussion of estimation of the probability of misclassification has been directed to the case of small sample sizes. This is due to the fact that the asymptotic expansions of the error as discussed in Chapter 2 do not indicate the behaviour of the error for small sample sizes. We may mention the work of Hills [18], Geisser [15], Lachenbruch and Mickey [31], Sedransk and Okamoto [53], Sorum [57, 58, 59], Broffitt and Williams [8], McLachlan [38, 39] and many others in this respect.

In this chapter we discuss the estimation of the optimum probability of misclassification in the ESD when the skewness factor varies in the range (0.00625 to 0.4]. We also discuss the apparent error rate for the normal distribution and ESD classification rules using data simulated from the ESD. The derivation of classification rules for the normal and ESD using the likelihood criterion is given in section 5.2. The form of the estimators is presented in section 5.3 and the choice of values for the skewness factor is discussed in section 5.4. The remaining part of the chapter is concerned with comparisons of the likelihood ratio (LR) classification rules for the normal distribution and the ESD using data simulated from the ESD. Some recommendations are given in Section 5.5.

5.2 Derivation of Classification Rules

The general theory of classification rules has been reviewed by Das Gupta [11]. Let f_i be the probability density function of X in population Π_i , then the likelihood ratio rule is defined as follows:

Definition: A likelihood ratio (LR) rule is defined by δ , where $\delta_j(x) > 0$, if for some positive constants $(C_1, C_2, \dots, C_M)$,

$$C_j f_j(x) < \max_{1 \leq i \leq M} \{C_i f_i(x)\}$$

where M is the number of populations.

If C_i 's are all equal, the rule is called a maximum-likelihood (ML) rule.

Das Gupta [11] discusses the classification problem when $M = 2$. In this case the classification problem is essentially that of testing a simple hypothesis against a simple alternative hypothesis. We derive the classification rules when the p.d.f's of X in the populations are known to be univariate normal and when they are known to be of the univariate ESD, for the case of $M = 2$.

a) Univariate Normal Distributions (ND)

Let the p.d.f's of X in Π_i be

$$f_i(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu_i}{\sigma}\right)^2\right], \quad -\infty < x < \infty \quad (5.1)$$

($i = 1, 2$).

Since $M = 2$ let us redefine the problem in the usual hypothesis

testing form.

Let θ be the mean of the observation X , then the null Hypothesis is $H_0: \theta = \mu_1$ and the Alternative hypothesis is $H_A: \theta = \mu_2$. When $\mu_1 < \mu_2$ the LR becomes

$$\begin{aligned} L &= \frac{f_1(x)}{f_2(x)} \\ &= \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_1}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{x - \mu_2}{\sigma}\right)^2\right\} \end{aligned} \quad (5.2)$$

Taking logarithms on both sides of (5.2) we obtain

$$\begin{aligned} L' = \ln L &= -\frac{1}{2}\left(\frac{x - \mu_1}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{x - \mu_2}{\sigma}\right)^2 \\ &= -\frac{1}{2\sigma^2}[2x - (\mu_1 + \mu_2)](\mu_2 - \mu_1) \\ &= [x - \frac{1}{2}(\mu_1 + \mu_2)]\left(\frac{\mu_1 - \mu_2}{\sigma^2}\right). \end{aligned} \quad (5.3)$$

The result in (5.3) is the W discriminant function when the distributions in the two populations are univariate normal with the same variance but having different means (see Sedransk and Okamoto [53]).

By the Neyman and Pearson Lemma (see Rao [51] for details) we reject H_0 if

$$L < K \quad (5.4)$$

where K is some constant. In the context of our problem we reject that $\theta = \mu_1$, which means we reject that X is from Π_1 and are in favour of X coming from Π_2 . For simplicity we take $K = 1$. From (5.3) and (5.4) the classification rule

may be written as follows

$$\left. \begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } W > 0 \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } W \leq 0 \end{array} \right\} \quad (5.5)$$

The rule in (5.5) reduces to

$$\left. \begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } X < \frac{1}{2}(\mu_1 + \mu_2) \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } X \geq \frac{1}{2}(\mu_1 + \mu_2) \end{array} \right\} \quad (5.6)$$

If $\mu_1 > \mu_2$, by a similar procedure the rule is

$$\begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } X < \frac{1}{2}(\mu_1 + \mu_2) \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } X \geq \frac{1}{2}(\mu_1 + \mu_2) \end{array}$$

When the parameters μ_1, μ_2 are unknown and are estimated by $\bar{x}_1, \bar{x}_2$, from sample sizes of n_1, n_2 respectively, the classification rule is seen to be

$$\begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } X < \frac{(\bar{x}_1 + \bar{x}_2)}{2} \text{ and } \bar{x}_1 < \bar{x}_2 \\ \text{or if } X > (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 \geq \bar{x}_2 ; \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } X \geq (\bar{x}_1 + \bar{x}_2)/2 \text{ and} \\ \bar{x}_1 < \bar{x}_2 \\ \text{or if } X \leq (\bar{x}_1 + \bar{x}_2)/2 \text{ and } \bar{x}_1 \geq \bar{x}_2 . \end{array}$$

(5.7)

b) Univariate Edgeworth Series Distributions (ESD)

Let the p.d.f.'s of X in Π_i be

$$f_i(x) = \{1 - \frac{\lambda_3}{6} D^3\} \phi\left(\frac{x - \mu_i}{\sigma}\right), \quad -\infty < x < \infty \quad (5.8)$$

($i = 1, 2$). These p.d.f.'s have been defined earlier in (4.1) and (4.2) of Chapter 4.

Once again define the problem in a hypothesis testing form.

Null hypothesis $H_0: \theta = \mu_1$,

Alternative hypothesis $H_A: \theta = \mu_2$.

When $\mu_1 < \mu_2$ the LR becomes

$$\begin{aligned} L &= \frac{f_1(x)}{f_2(x)} \\ &= \frac{\{1 - [\lambda_3/2\sigma^3] (\frac{x-\mu_1}{\sigma}) + [\lambda_3/6\sigma^3] (\frac{x-\mu_1}{\sigma})^3\} \phi(\frac{x-\mu_1}{\sigma})}{\{1 - [\lambda_3/2\sigma^3] (\frac{x-\mu_2}{\sigma}) + [\lambda_3/6\sigma^3] (\frac{x-\mu_2}{\sigma})^3\} \phi(\frac{x-\mu_2}{\sigma})} \end{aligned} \quad (5.9)$$

Define

$$A = \{1 - [\lambda_3/2\sigma^3] (\frac{x-\mu_1}{\sigma}) + [\lambda_3/6\sigma^3] (\frac{x-\mu_1}{\sigma})^3\} \quad (5.10)$$

and

$$B = \{1 - [\lambda_3/2\sigma^3] (\frac{x-\mu_2}{\sigma}) + [\lambda_3/6\sigma^3] (\frac{x-\mu_2}{\sigma})^3\} \quad (5.11)$$

then

$$L = \frac{A \exp[-\frac{1}{2}(\frac{x-\mu_1}{\sigma})^2]}{B \exp[-\frac{1}{2}(\frac{x-\mu_2}{\sigma})^2]} \quad (5.12)$$

By the Neyman and Pearson Lemma we reject H_0 if

$$L < K \quad (5.13)$$

or equivalently

$$\ln L < \ln K \quad (5.14)$$

taking $K = 1$ gives us $\ln L < 0$.

Hence we reject that X is from Π_1 if

$$\ln A - \frac{1}{2} \left\{ \frac{x - \mu_1}{\sigma} \right\}^2 - \ln B + \frac{1}{2} \left\{ \frac{x - \mu_2}{\sigma} \right\}^2 < 0$$

which simplifies to

$$\ln \left[\frac{A}{B} \right] + [x - (\mu_1 + \mu_2)/2] \left(\frac{\mu_1 - \mu_2}{\sigma} \right) > 0 \quad (5.15)$$

From (5.15) the classification rule becomes:

For $\mu_1 < \mu_2$

Classify X as belonging to Π_1 if $\ln \left[\frac{A}{B} \right] + W > 0$

and (5.16)

Classify X as belonging to Π_2 if $\ln \left[\frac{A}{B} \right] + W \leq 0$

for $\mu_1 > \mu_2$ the classification rule is seen to be

Classify X as belonging to Π_2 if $\ln \left[\frac{B}{A} \right] - W > 0$

and (5.17)

Classify X as belonging to Π_1 if $\ln \left[\frac{B}{A} \right] - W \leq 0$

When the parameters μ_1, μ_2 are unknown they are estimated by $\bar{x}_1, \bar{x}_2$ respectively and plugged in the rules (5.17) before classification begins.

We note here that in the comparisons of errors of misclassification using ESD and Normal classification rules and, using data generated from the ESD the classification rule for

the ESD is not simplified but is left in the form:

$$\begin{aligned}
 &\text{If } \mu_1 < \mu_2 \\
 &\quad \text{Classify } X \in \Pi_2 \quad \text{if } \frac{A \exp[-\frac{1}{2}(\frac{x - \mu_1}{\sigma})^2]}{B \exp[-\frac{1}{2}(\frac{x - \mu_2}{\sigma})^2]} < 1 \\
 &\text{and} \\
 &\quad \text{Classify } X \in \Pi_1 \quad \text{if } \frac{A \exp[-\frac{1}{2}(\frac{x - \mu_1}{\sigma})^2]}{B \exp[-\frac{1}{2}(\frac{x - \mu_2}{\sigma})^2]} \geq 1
 \end{aligned} \quad (5.18)$$

5.3 Defining Estimators

Let $\{X_{ij}, j = 1, 2, \dots, n_i; i = 1, 2\}$ be independent samples of sizes n_1, n_2 from the populations Π_1, Π_2 . The apparent probability of misclassification is defined as

$$E_{12} = \frac{1}{n_1} \sum_{j=1}^{n_1} \gamma_j \quad (5.19)$$

where

$$\gamma_j = \begin{cases} 1 & \text{when } X_{1j} \text{ is classified as belonging to } \Pi_2 \\ 0 & \text{when } X_{1j} \text{ is classified as belonging to } \Pi_1 \end{cases}$$

Similarly, let

$$E_{21} = \frac{1}{n_2} \sum_{j=1}^{n_2} \gamma_j \quad (5.20)$$

where

$$\gamma_j = \begin{cases} 1 & \text{when } X_{2j} \text{ is classified as belonging to } \Pi_1 \\ 0 & \text{when } X_{2j} \text{ is classified as belonging to } \Pi_2 \end{cases}$$

The optimum probability of misclassification is defined in (4.24) and (4.26) of Chapter 4 and is given as

$$\alpha_1[R, f] = 1 - \Phi\left[\frac{\mu_2 - \mu_1}{2\sigma}\right] + \frac{\lambda_3}{6\sigma^2} \left[\left\{ \frac{\mu_2 - \mu_1}{2\sigma} \right\}^2 - 1 \right] \phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) \quad (5.21)$$

$$\alpha_2[R, f] = \Phi\left[\frac{\mu_1 - \mu_2}{2\sigma}\right] - \frac{\lambda_3}{6\sigma^2} \left[\left\{ \frac{\mu_1 - \mu_2}{2\sigma} \right\}^2 - 1 \right] \phi\left(\frac{\mu_1 - \mu_2}{2\sigma}\right) \quad (5.22)$$

where $\alpha_1[R, f]$ is the expression for the optimum error of misclassification when an observation from Π_1 is misclassified and $\alpha_2[R, f]$ is the expression for the optimum error of misclassification when an observation from Π_2 is misclassified.

We shall be using these estimators in sections 5.4 onwards.

5.3 Choice of Values for Skewness Factor

The choice of λ_3 , the skewness factor, is based on the boundary of the unimodal region for Edgeworth distributions, as described by Barton and Dennis [7]. This boundary is defined in terms of β_1 and β_2 where β_1 is the square of skewness and β_2 is kurtosis. The boundary regions for β_1, β_2 are approximately given by

$$0 \leq \beta_1 < 0.25 \quad (5.23)$$

and

$$0 \leq \beta_2 \leq 5.4 \quad (5.24)$$

Rewriting (5.23) we get

$$0 \leq (\beta_1)^{\frac{1}{2}} < 0.5 \quad (5.25)$$

where $(\beta_1)^{\frac{1}{2}} = \lambda_3$

In this study β_2 is taken to be fixed at 3 and λ_3 chosen to lie in the range $(6.25 \times 10^{-3}, 0.4)$.

5.4 Estimation of Probability of Misclassification

a) Optimum Probability of Misclassification

The optimum probabilities of misclassification are computed from equations (5.21) and (5.22). The parameters μ_1 , μ_2 and σ are assumed known with the following values

$$\mu_1 = 0, \mu_2 = 1 \text{ and } \sigma = 1$$

Thus (5.21) and (5.22) become functions of λ_3 , the skewness factor, only. With λ_3 in the range $(6.25 \times 10^{-3}, 0.4)$, the probabilities of misclassification E_{12} and E_{21} are computed. The results are given in Table 5.1. This shows that E_{12} decreases as λ_3 increases while E_{21} increases as λ_3 increases. The total probability of misclassification is fairly constant as λ_3 increases.

b) Apparent Probability of Misclassification

This probability of misclassification is investigated in two cases: (i) when all parameters in the distributions from Π_1 and Π_2 are known and (ii) when the parameters μ_1 , μ_2 are estimated from small samples.

When the parameters are known, the apparent probability of misclassification is obtained as follows. Two independent samples of size 200 each are obtained at each fixed λ_3 value from populations Π_1 , Π_2 whose distributions are of the ESD with the following parameter values:

Table 5.1. Optimum Probability of Misclassification at Different values of Skewness, for the ESD. All parameters known.

Skewness Factor (λ_3)	Optimum Probability of Misclassification		
	E_{12}	E_{21}	Total
6.25×10^{-3}	0.308262	0.308813	0.617075
1.25×10^{-2}	0.307987	0.309088	0.617075
0.25×10^{-1}	0.307437	0.309635	0.617072
0.05	0.306337	0.310735	0.617072
0.10	0.304137	0.312938	0.617075
0.15	0.301936	0.315139	0.617075
0.20	0.299736	0.317339	0.617075
0.25	0.297535	0.319540	0.617075
0.30	0.295335	0.321740	0.617075
0.35	0.293135	0.323940	0.617075
0.40	0.290935	0.326141	0.617076

<u>Population</u>	<u>Parameter Values</u>
Π_1	$\mu_1 = 0, \sigma^2 = 1$
Π_2	$\mu_1 = 1, \sigma^2 = 1$

By the use of the ESD classification rule and the ND classification rule, the proportions misclassified in Π_1 and Π_2 are obtained. This realization is repeated five times, the random numbers which are used in generating the random variables have a different start for each independent sample [see method of generating the distribution in the Appendix]. The probabilities of misclassification obtained from the five samples are averaged. The results are shown in Table 5.2, where the notation E_{12E} means the apparent probability of misclassification when an observation from Π_1 is misclassified by the ESD classification rule and E_{21E} is the apparent probability of misclassification when an observation from Π_2 is misclassified by the ESD classification rule. Similarly E_{12N} and E_{21N} are apparent probabilities of misclassification when the ND classification rule is used. The results show that E_{12E} is always higher than E_{12N} at each λ_3 value and E_{21E} is always less than E_{21N} at each λ_3 value. We also tabulate the total probability of misclassification for the ESD classification rule and ND classification rule. The total probability of misclassification for the ESD and ND rules show no major difference between the two at each value of λ_3 . If the sign of λ_3 is reversed then one would expect the results to be reversed.

Table 5.2. Comparison of the Errors of Misclassification, Using the LR rules for the Normal Distribution (ND) and the ESD, Averaged Over Five Samples. All parameters known. Simulation Size = 200.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{12N}	Total
6.25×10^{-3}	0.329	0.303	0.632	0.327	0.303	0.630
1.25×10^{-2}	0.321	0.321	0.642	0.318	0.322	0.640
0.25×10^{-1}	0.282	0.326	0.608	0.278	0.330	0.608
0.05	0.312	0.289	0.601	0.308	0.296	0.604
0.10	0.328	0.296	0.624	0.315	0.307	0.622
0.15	0.355	0.298	0.653	0.330	0.324	0.654
0.20	0.328	0.305	0.633	0.287	0.331	0.618
0.25	0.339	0.262	0.601	0.298	0.296	0.594
0.30	0.377	0.248	0.625	0.320	0.304	0.624
0.35	0.387	0.255	0.642	0.328	0.310	0.638
0.40	0.396	0.247	0.643	0.297	0.301	0.598

When the parameters μ_1, μ_2 are estimated by $\bar{x}_1, \bar{x}_2$ from small samples the apparent probabilities of misclassification can also be obtained. In this case a sample size 5 is obtained from population Π_1 and the mean $\bar{x}_1$ calculated and another sample of size 5 is obtained from population Π_2 and the mean $\bar{x}_2$ calculated. These samples are generated from the ESD with parameters taking the following values:

<u>Population</u>	<u>Parameter Values</u>
Π_1	$\mu_1 = 0, \sigma^2 = 1$
Π_2	$\mu_2 = 1, \sigma^2 = 1$

λ_3 is taken in the range $(6.25 \times 10^{-3}, 0.4)$.

At each fixed value of λ_3 the means $\bar{x}_1, \bar{x}_2$ are obtained from samples of size 5 and plugged in the classification rules of the ESD and ND. The apparent probability of misclassification at each λ_3 is obtained from a further sample of size 200 from Π_1 and another sample of 200 from Π_2 . This process is repeated for each λ_3 value by estimating means $\bar{x}_1, \bar{x}_2$ from samples of sizes 15 and 25. The results of these calculations are given in Tables 5.3 - 5.4. The total probabilities of misclassification for the ESD and ND are also presented.

Since the estimates of μ_1, μ_2 are obtained from small samples high variation in the estimates is expected and this affects the classification rules. More data are needed to stabilize the effect of random variation in the estimates before any definite statements can be made. However, we can

Table 5.3. Comparison of the Errors of Misclassification Using the LR rules for the Normal Distribution (ND) and the ESD. Means unknown and Estimated by Sample Sizes of 5. Simulation Size = 200.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
6.25×10^{-3}	0.140	0.400	0.540	0.140	0.400	0.540
1.25×10^{-2}	0.220	0.410	0.630	0.220	0.410	0.630
0.25×10^{-1}	0.225	0.465	0.690	0.220	0.475	0.695
0.05	0.210	0.395	0.605	0.205	0.400	0.605
0.10	0.205	0.475	0.680	0.175	0.495	0.670
0.15	0.260	0.285	0.545	0.230	0.320	0.550
0.20	0.305	0.365	0.670	0.295	0.395	0.690
0.25	0.455	0.185	0.640	0.420	0.230	0.650
0.30	0.195	0.465	0.660	0.115	0.545	0.660
0.35	0.225	0.465	0.660	0.125	0.520	0.645
0.40	0.440	0.180	0.620	0.360	0.250	0.610

Table 5.4. Comparison of the Errors of Misclassification Using the LR rules for the Normal Distribution (ND) and the ESD. Means unknown and Estimated by Sample Sizes of 15. Simulation Size = 200.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
6.25×10^{-3}	0.345	0.145	0.490	0.345	0.150	0.495
1.25×10^{-2}	0.310	0.310	0.620	0.310	0.310	0.620
0.25×10^{-1}	0.405	0.280	0.685	0.400	0.285	0.685
0.05	0.230	0.390	0.620	0.225	0.395	0.620
0.10	0.375	0.305	0.680	0.350	0.315	0.665
0.15	0.405	0.180	0.585	0.360	0.225	0.585
0.20	0.355	0.325	0.680	0.320	0.355	0.675
0.25	0.295	0.340	0.635	0.235	0.395	0.630
0.30	0.320	0.350	0.670	0.230	0.385	0.615
0.35	0.260	0.345	0.605	0.200	0.430	0.630
0.40	0.315	0.375	0.690	0.145	0.415	0.560

Table 5.5. Comparison of the Errors of Misclassification Using the LR rules for the Normal Distribution (ND) and ESD. Means unknown and Estimated by Sample Sizes of 25. Simulation Size = 200.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
6.25×10^{-3}	0.270	0.220	0.490	0.270	0.220	0.490
1.25×10^{-2}	0.290	0.330	0.620	0.290	0.235	0.525
0.25×10^{-1}	0.390	0.295	0.685	0.375	0.310	0.685
0.05	0.340	0.270	0.610	0.335	0.280	0.615
0.10	0.375	0.305	0.680	0.360	0.315	0.675
0.15	0.360	0.230	0.590	0.345	0.245	0.590
0.20	0.275	0.430	0.705	0.225	0.480	0.705
0.25	0.375	0.255	0.630	0.320	0.290	0.610
0.30	0.390	0.240	0.630	0.300	0.330	0.630
0.35	0.290	0.300	0.590	0.240	0.345	0.585
0.40	0.405	0.225	0.630	0.305	0.290	0.595

tentatively make observations on these results. It is clear from tables [5.2 - 5.5] that the total probability of misclassification at each value of λ_3 is either under or over-estimated when small samples are used to estimate μ_1 and μ_2 . The differences between taking small samples of 5, 15 and 25 are not apparent. The behaviour of individual probabilities E_{12E} and E_{12N} show that for small sample sizes E_{12E} is equal or greater than E_{12N} at each value of λ_3 , while E_{21E} is equal or less than E_{21N} at each value of λ_3 . The equality of the probabilities occurs when λ_3 is very small, with the disparity increasing as λ_3 increases.

The fact that E_{12E} is equal to or greater than E_{12N} at each value of λ_3 and E_{21E} is equal to or less than E_{21N} at each value of λ_3 in the tabled results needs some algebraic justification.

From (5.5) the normal classification rule, based on the W discriminant function, is

$$\left. \begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } W > 0 \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } W \leq 0 \end{array} \right\} \quad (5.26)$$

From (5.16) the ESD classification rule is

$$\left. \begin{array}{l} \text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } \ln\left[\frac{A}{B}\right] + W > 0 \\ \text{and} \\ \text{Classify } X \text{ as belonging to } \Pi_2 \text{ if } \ln\left[\frac{A}{B}\right] + W \leq 0 \end{array} \right\} \quad (5.27)$$

where

$$A = [1 - (\frac{\lambda_3}{2\sigma^3}) \{\frac{x - \mu_1}{\sigma}\} + (\frac{\lambda_3}{6\sigma^3}) \{\frac{x - \mu_1}{\sigma}\}^3] \quad (5.28)$$

$$B = [1 - (\frac{\lambda_3}{2\sigma^3}) \{\frac{x - \mu_2}{\sigma}\} + (\frac{\lambda_3}{6\sigma^3}) \{\frac{x - \mu_2}{\sigma}\}^3] . \quad (5.29)$$

When $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma = 1$ we have

$$A = 1 - \frac{\lambda_3}{2} X + \frac{\lambda_3}{6} X^3 \quad (5.30)$$

$$B = 1 - \frac{\lambda_3}{2} (X - 1) + \frac{\lambda_3}{6} (X - 1)^3 \quad (5.31)$$

Let us consider the error of misclassifying an observation X from Π_1 . Then, by the normal rule, we classify an observation from Π_2 wrongly if we use the rule

$$\text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } W \leq 0 \quad (5.32)$$

and the corresponding wrong classification of an observation from Π_1 , by the ESD classification rule is to use:

$$\text{Classify } X \text{ as belonging to } \Pi_1 \text{ if } \ln\left[\frac{A}{B}\right] + W \leq 0 \quad (5.33)$$

It is clear from (5.32) and (5.33) that the cut-off points are different depending on the value of $\ln\left[\frac{A}{B}\right]$ in (5.33); $\ln\left[\frac{A}{B}\right]$ is a function of the random variable X and skewness factor λ_3 . If $\ln\left[\frac{A}{B}\right] = 0$ the two procedures are equivalent.

Consider

$$\frac{A}{B} = \frac{1 - \frac{\lambda_3}{2} x + \frac{\lambda_3}{6} x^3}{1 - \frac{\lambda_3}{2}(x-1) + \frac{\lambda_3}{6}(x-1)^3}$$

thus

$$\frac{A}{B} = \frac{1 - \frac{\lambda_3}{2} x + \frac{\lambda_3}{6} x^3}{1 - \frac{\lambda_3}{2} x + \frac{\lambda_3}{2} + \frac{\lambda_3}{6}[x^3 - 3x^2 + 3x - 1]} \quad (5.34)$$

Now

$$\begin{aligned} A-B &= \frac{\lambda_3}{2} x^2 - \frac{\lambda_3}{2} x + \frac{\lambda_3}{6} - \frac{\lambda_3}{2} \\ &= \frac{\lambda_3}{2} x^2 - \frac{\lambda_3}{2} x - \frac{\lambda_3}{3} \\ &= \frac{\lambda_3}{6}[3x^2 - 3x - 2] \end{aligned} \quad (5.35)$$

The quadratic function (5.35) is equal to zero if

$$3x^2 - 3x - 2 = 0 \quad (5.36)$$

The solutions of (5.36) are

$$x = + 1.457 \quad \text{and} \quad x = - 0.457 . \quad (5.37)$$

The graph of (5.35) faces upwards and shows that, when $-0.457 < x < 1.457$, $A < B$. From the simulation results shown in the appendix most of the data lie in the range 1.457 to -0.457 , viz. approximately 60% or over.

Since A is less than B ,

$$\ln\left[\frac{A}{B}\right] = \ln A - \ln B < 0 \quad (5.38)$$

Thus the cut-off point of the ESD classification rule is higher than the ND classification rule cut-off point. This leads to E_{12E} being greater than E_{12N} .

Similarly if we wrongly classify an observation from Π_2 the cut-off point of the ESD classification rule is lower than the ND classification rule cut-off point and as such E_{21E} is less than E_{21N} .

5.5 Recommendations and Conclusions

From the results shown in section 5.4 the following is clear:

- (i) The skewness factor (λ_3) has very little effect, if any, on the total probability of misclassification, although it shows a decrease or increase in individual errors of misclassification. We may therefore recommend that the normal rule can be used without much loss in classifying observations which are from a skewed distribution. One advantage of the normal rule is that it is in a simple form and easy to use.
- (ii) Estimation of the errors when small samples are used to estimate means shows either underestimation or overestimation of the optimum error of misclassification. This conclusion is based on one realization of data and as such only applies to the present situation. Further work is proposed to study the effects of small sample size estimates of the parameters.

APPENDIX A1

COMPUTER SIMULATION OF THE ESD

A1.1 Introduction

To generate random variables from the ESD we use the method of Numerical Inverse Interpolation. Schmidt and Taylor [52] describe this method in detail. This method is usually adopted when the distribution function of the random variable X can be evaluated exactly for X but the inverse cannot be evaluated exactly. That is if

$$X = \psi[F(X)] , \quad (\text{A1.1})$$

where $F(X)$ is the distribution function of X , it is either impossible or difficult to evaluate X , given $F(X)$. This happens to be the case in the ESD.

Let X have the p.d.f. given by

$$f(x) = [1 - \frac{\lambda_3}{6} D^3] \phi\left(\frac{x - \mu}{\sigma}\right), \quad -\infty < x < \infty \quad (\text{A1.2})$$

The distribution function of X is given by

$$\begin{aligned} F(x) &= \int_{-\infty}^x [1 - \frac{\lambda_3}{6} D^3] \phi\left(\frac{y - \mu}{\sigma}\right) dy \\ &= \int_{-\infty}^x [1 + \frac{\lambda_3}{6\sigma^3} H_3\{\frac{y - \mu}{\sigma}\}] \phi\left(\frac{y - \mu}{\sigma}\right) dy \\ &= \phi\left\{\frac{x - \mu}{\sigma}\right\} - \frac{\lambda_3}{6\sigma^3} [(\frac{x - \mu}{\sigma})^2 - 1] \phi\left(\frac{x - \mu}{\sigma}\right) \quad (\text{A1.3}) \end{aligned}$$

From the distribution function (A1.3) we see immediately that (A1.1) is not easy to evaluate.

A1.2 Numerical Inverse Interpolation Method

Let R be a uniformly distributed random number on the interval $(0,1)$, then the method of numerical inverse interpolation is as follows:

- (i) Generate R
- (ii) Find by trial and error, two values A and B of X such that $F(A) < R < F(B)$.
- (iii) By interpolation determine the value of X between A and B such that $F(X) = R$.

To use the above method we tabulate the distribution function in (A1.3) when the parameters take the following values:

$$\mu = 1, \sigma^2 = 1 \text{ and } \lambda_3 \text{ is in the range } (6.25 \times 10^{-3}, 0.4).$$

The cut-off points on the x scale were chosen from the behaviour of the tabulated distribution functions. If x is less than -6 the distribution takes too many negative values in the range of λ_3 and when x is greater than 6 then the distribution function takes values above unity. Thus it was considered reasonable to restrict ourselves to $-6 \leq x \leq 6$. The intervals on the x scale were taken to be 0.1 and this gives us 121 tabulated points of the distribution function which are stored in a vector V . The x values are also stored in a vector F .

A random number R ($0 < R < 1$) is generated and by trial

and error two values of x , A and B such that $F(A) < R < F(B)$ are found. If a particular random number fails to achieve this, we generate another random number and repeat the search through the vector V . To find x between A and B such that $F(X) = R$ we use the Newton-Raphson iteration technique. The Newton-Raphson iteration technique is described below.

A1.3 Newton-Raphson Iteration Technique

Let $F(X)$ be the distribution function of X , then the Taylor series expansion of $F(X)$ about the point $x_1 = A$ to 2 terms is given by

$$F(x_1) = F(A) + hF'(A) \quad \text{where } h = x_1 - A$$

therefore

$$F(x_1) = F(A) + (x_1 - A)F'(A)$$

and solving for x_1 gives

$$x_1 = \frac{F(x_1) - F(A)}{F'(A)} + A, \quad (\text{A1.4})$$

where $F(x_1) = R$ (random number) and $F'(A) = f(A)$ the p.d.f. in (A1.2).

If we expand the distribution function $F(X)$ in a Taylor series to two terms at the point B where $B > A$ we obtain

$$F(x_1) = F(B) + (x_1 - B)F'(B).$$

Solving for x_1 yields

$$x_1 = B - \frac{F(B) - F(x_1)}{F'(B)} \quad (A1.5)$$

We thus have the option to iterate from point B or A. We chose to iterate from point A. The number of iterations depends on the error allowed. We set it equal to 10^{-6} and this gave us 1 or 2 iterations at most.

A1.4 Other Algorithms

To evaluate the upper or lower tail areas of the standardized normal curve corresponding to any given argument we use the algorithm given by Hill [17], whose subprogram bears the name ALNORM. The function RANDOM is used to generate random numbers on the (0,1) scale; this has been developed by Lohmes and Cooley [33]. Subroutine SCAN developed by the author is used to find points A,B on the interval [-6, 6] to be used for finding to random variable X. A new subroutine SINTP is also given here, for iteration from point A using the Newton-Raphson technique.

A1.5 Estimation of Parameters and Goodness-of-Fit Tests

a) Estimation of Parameters

The parameters of interest here are the mean μ , the variance σ^2 and the skewness factor λ_3 . Let y be the random variable of the ESD. Then $x = \frac{y - \mu}{\sigma}$ is the transformed variable. This is necessary since the evaluation of the normal integral requires a standardization of the variable. We take $\sigma^2 = 1$. In all cases the moments about the origin are given by

$$\text{1st Moment} \quad M_1' = \left\{ \sum_{i=1}^n x_i \right\} / n$$

$$\text{2nd Moment} \quad M_2' = \left\{ \sum_{i=1}^n x_i^2 \right\} / n$$

$$\text{3rd Moment} \quad M_3' = \left\{ \sum_{i=1}^n x_i^3 \right\} / n$$

$$\text{4th Moment} \quad M_4' = \left\{ \sum_{i=1}^n x_i^4 \right\} / n$$

where n is the sample size. The moments about the mean of x are as follows

$$\text{1st Moment} \quad M_1 = M_1' + \mu_1$$

$$\text{2nd Moment} \quad M_2 = M_2' - M_1'^2$$

$$\text{3rd Moment} \quad M_3 = M_3' - 3M_1' M_2' + 2M_1'^3$$

$$\text{4th Moment} \quad M_4 = M_4' - 4M_1' M_3' + 6M_1'^2 M_2' - 3M_1'^4$$

These moments are shown in Table A1.1 when a sample of 200 is taken from the ESD. Skewness and kurtosis are also calculated. The first and second moments show some stability irrespective of the value of skewness factor. The third moment is highly variable and so is the fourth moment. As such skewness and kurtosis show no stability. This may be due to the fact that the theoretical values of the third and fourth moments depend on higher moments (as far as the 8th moment).

b) Chi-Square Goodness of Fit Test

The results of the Chi-square-goodness of fit test are presented

Table A1.1 Moments About the Mean of ESD, Skewness and Kurtosis from a Sample of Size 200.

Skewness Factor	Moments				Skewness*[Kurtosis**+3]	
	M ₁	M ₂	M ₃	M ₄		
6.25 x 10 ⁻³	0.8584	1.0653	0.0440	3.3831	0.0400	2.9809
1.25 x 10 ⁻²	1.0028	1.0407	-0.1170	2.9175	-0.1102	2.6939
0.25 x 10 ⁻¹	1.0423	1.0410	0.2194	2.8431	0.2066	2.6238
0.05	0.9992	1.0106	0.0553	2.8962	0.0544	2.8359
0.10	1.0200	0.8602	-0.0667	2.1256	-0.0836	2.8725
0.15	0.8780	0.8743	0.2287	2.5429	0.2798	3.3265
0.20	1.0282	1.1562	0.0944	3.4716	0.0759	2.5968
0.25	1.1184	1.0759	0.3479	3.9584	0.3117	3.4198
0.30	0.9359	0.8145	0.3444	2.2754	0.4685	3.4299
0.35	1.0318	0.9741	0.1298	2.3626	0.1350	2.4901
0.40	1.0859	1.0487	0.6229	3.5805	0.5800	3.2555

* Skewness = $M_3 / (M_2^{3/2})$

** Kurtosis = $M_4 / (M_2^2) - 3$

Table A1.2 Observed and Expected Frequencies, Chi-square statistics from samples of size 200, skewness in the range (0.00625, 0.4)

Skewness Factor (λ_3)	Class Lower Limit	Observed Frequency	Expected Frequency
6.25×10^{-3}	-6.0	20	13.32808
	-1.5	22	18.40367
	-1.0	27	30.03179
	-0.5	37	38.32075
	0.0	42	38.26402
	0.5	31	29.92102
	1.0	12	18.33558
	1.5	9	13.39498
	Chi-Square = 8.429122		d.f. = 7
1.25×10^{-2}	-6.0	19	13.29436
	-1.5	17	18.43740
	-1.0	27	30.08682
	-0.5	30	38.34884
	0.0	42	38.23591
	0.5	36	29.86602
	1.0	14	18.30186
	1.5	15	13.42871
	Chi-Square = 7.520471		d.f. = 7
0.25×10^{-1}	-6.0	13	13.22690
	-1.5	16	18.50486
	-1.0	34	30.19683
	-0.5	42	38.40505
	0.0	34	38.17971
	0.5	23	29.75599
	1.0	18	18.23440
	1.5	20	13.49617
	Chi-Square = 6.287172		d.f. = 7

Skewness Factor (λ_3)	Class Lower Limit	Observed Frequency	Expected Frequency
0.05	-6.0	17	13.09199
	-1.5	19	18.63978
	-1.0	28	13.41687
	-0.5	31	38.51745
	0.0	45	38.06729
	0.5	31	29.53596
	1.0	15	18.09948
	1.5	14	13.63108
	Chi-Square = 4.708631		d.f. = 7
0.10	-6.0	10	12.82216
	-1.5	19	18.90962
	-1.0	30	30.85696
	-0.5	39	38.74226
	0.0	42	37.84246
	0.5	34	29.09588
	1.0	14	17.82966
	1.5	12	13.90091
	Chi-Square = 3.012980		d.f. = 7
0.15	-6.0	14	12.55234
	-1.5	15	19.17945
	-1.0	42	31.29704
	-0.5	42	38.96707
	0.0	36	37.61764
	0.5	31	28.65580
	1.0	9	17.55984
	1.5	11	14.17074
	Chi-Square = 10.117400		d.f. = 7

Skewness Factor (λ_3)	Class Lower Limit	Observed Frequency	Expected Frequency
0.20	-6.0	18	12.28252
	-1.5	20	19.44927
	-1.0	34	31.73713
	-0.5	22	39.19186
	0.0	34	37.39280
	0.5	37	28.21575
	1.0	18	17.29002
	1.5	17	14.44057
	Chi-Square = 13.905150		d.f. = 7
0.25	-6.0	11	12.01270
	-1.5	16	19.71911
	-1.0	23	32.17720
	-0.5	43	39.41668
	0.0	43	37.16798
	0.5	29	27.77563
	1.0	16	17.02020
	1.5	19	14.71040
	Chi-Square = 6.011059		d.f. = 7
0.30	-6.0	6	11.74287
	-1.5	27	19.98895
	-1.0	33	32.61729
	-0.5	42	39.64149
	0.0	44	36.94316
	0.5	24	27.33557
	1.0	15	16.75038
	1.5	9	14.98023
	Chi-Square = 9.737742		d.f. = 7

Skewness Factor (λ_3)	Class Lower Limit	Observed Frequency	Expected Frequency
0.35	-6.0	13	11.47305
	-1.5	18	20.25878
	-1.0	28	33.05737
	-0.5	43	39.86630
	0.0	35	36.71830
	0.5	27	26.89547
	1.0	21	16.48056
	1.5	15	15.25006
	Chi-Square = 2.799382		d.f. = 7
0.40	-6.0	8	11.20322
	-1.5	19	20.52862
	-1.0	29	33.49745
	-0.5	44	40.09111
	0.0	45	36.49346
	0.5	21	26.45541
	1.0	11	16.21073
	1.5	23	15.51989
	Chi-Square = 10.402570		d.f. = 7

$$\chi^2_{[7, 0.05]} = 14.07, \chi^2_{[7, 0.01]} = 18.48, \chi^2_{[7, 0.001]} = 24.32$$

in Table A1.2. These are based on a sample size of 200 observations from the ESD. The observed and expected frequencies are computed and from these the statistic

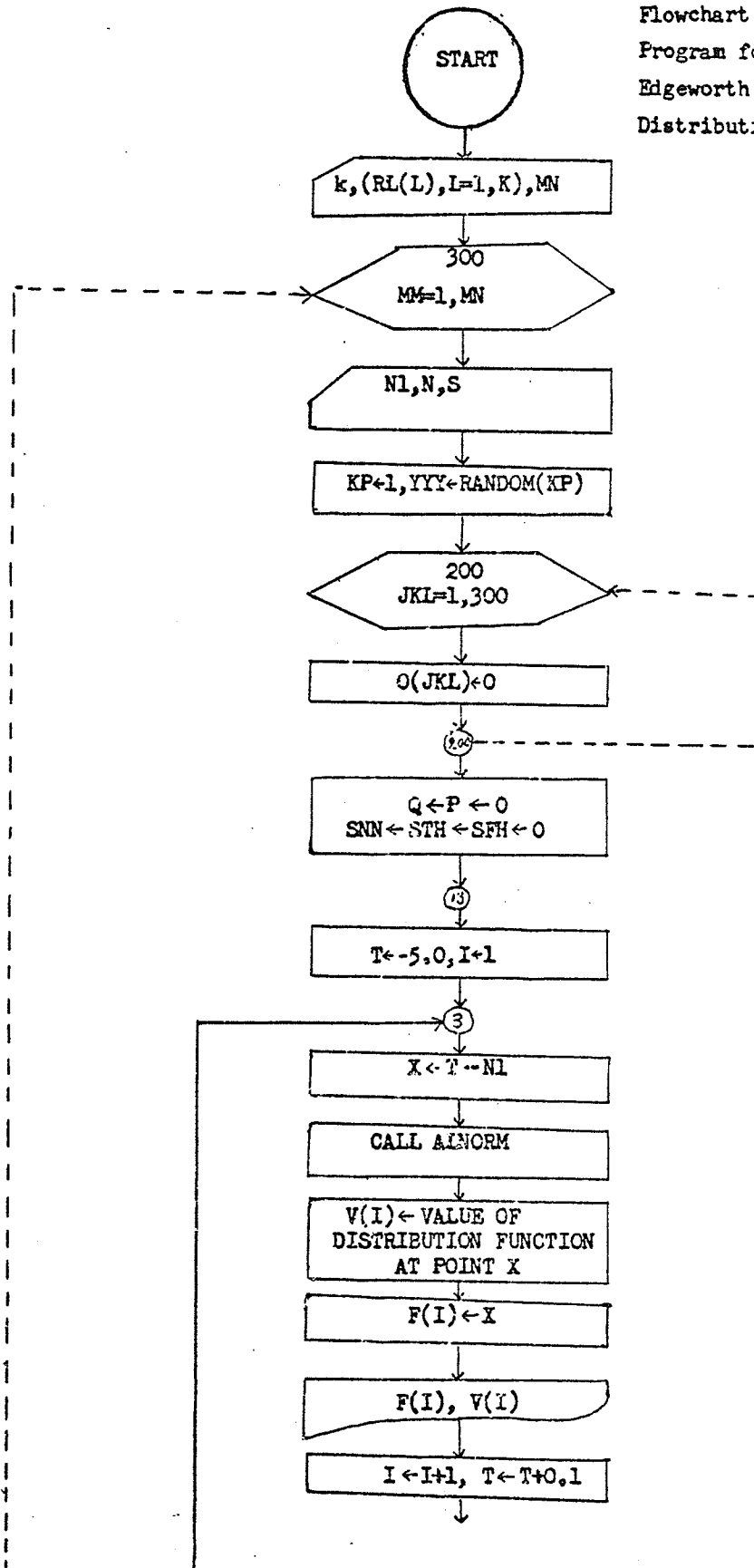
$$\chi^2 = \sum \frac{(o - E)^2}{E}$$

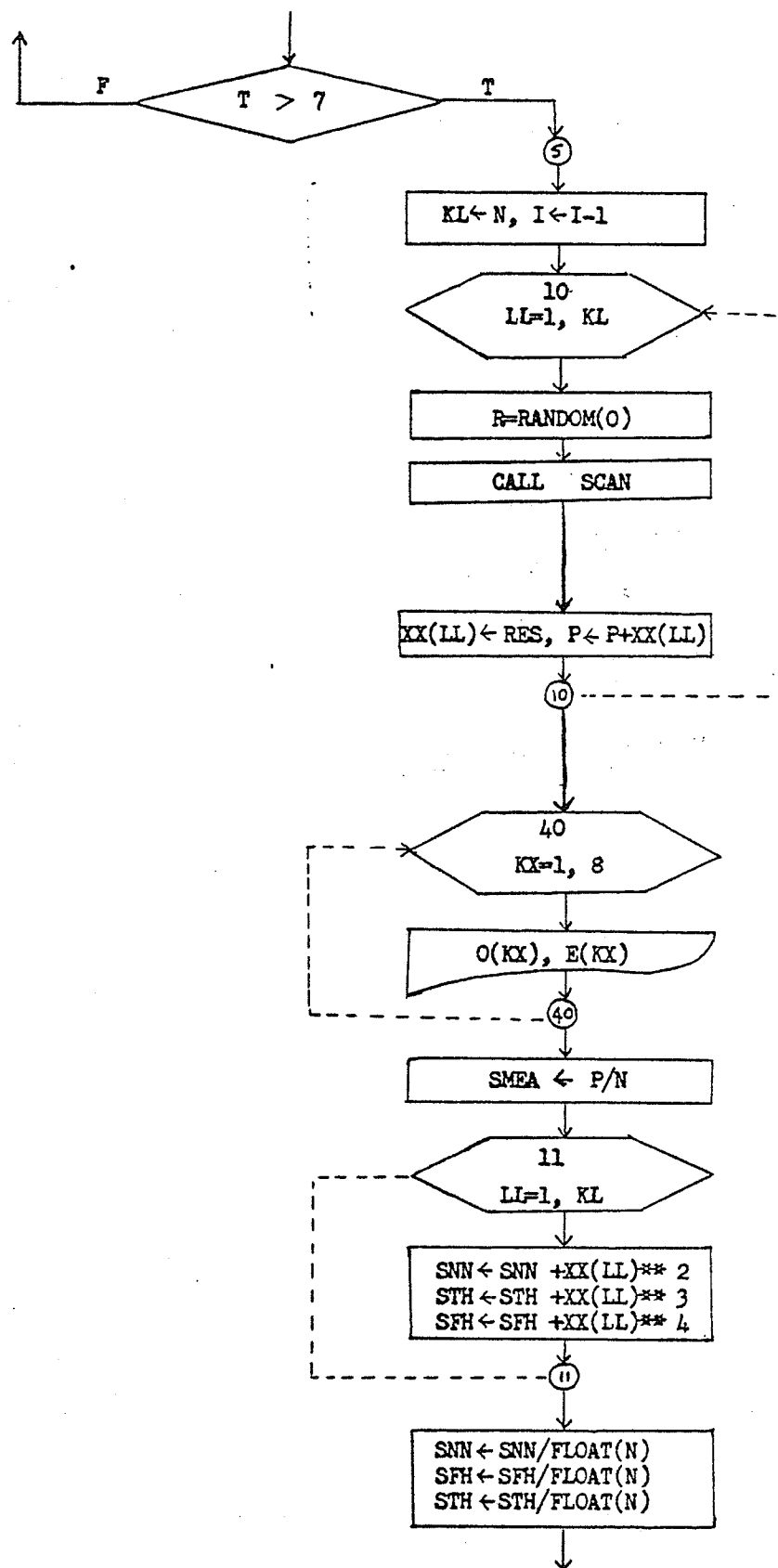
where χ^2 has approximately a Chi-square distribution with $(K - 1)$ degrees of freedom. Here K is the number of classes o the observed frequency and E the expected frequency. From Table A1.2, none of the χ^2 values are significant, i.e. none of the generated data are rejected. We therefore concluded that our observed data are from the ESD.

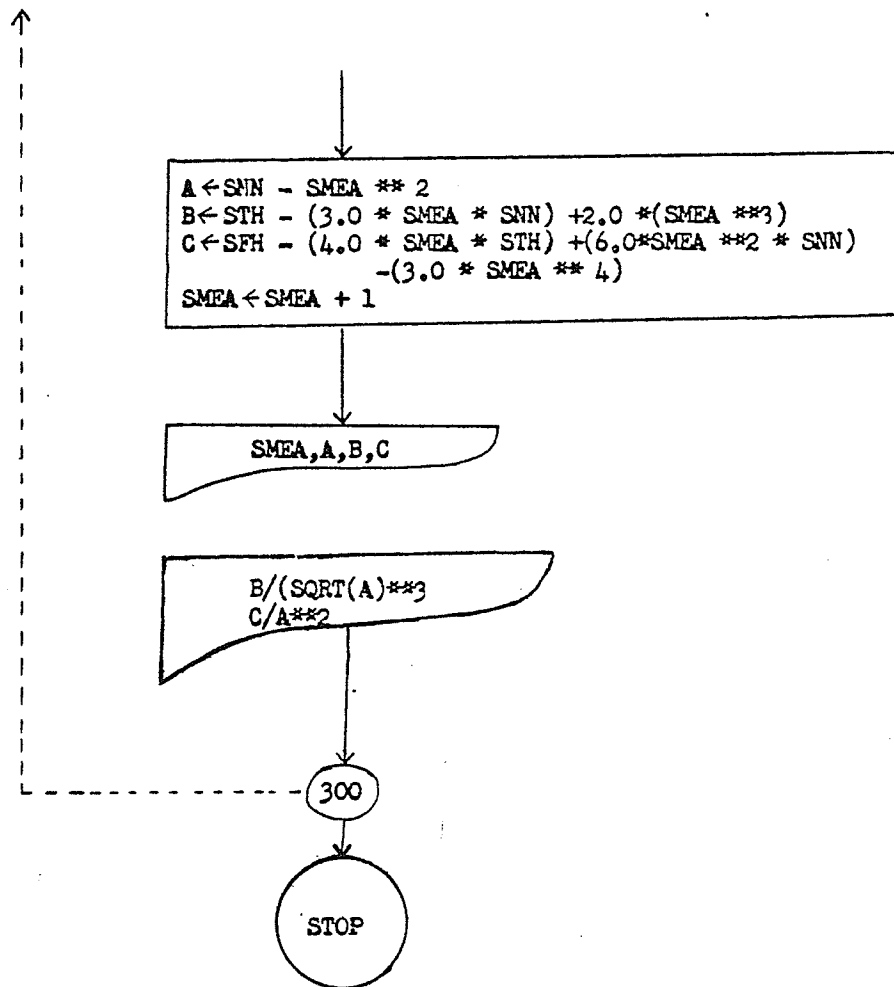
A1.6 Programmes and Flowcharts

The following is a listing of flowcharts and programs which can be used to generate the ESD using techniques described in the Appendix A1.

Flowchart for Mainline
Program for Generating
Edgeworth Series
Distribution







\$JOB WATFIV ENOCK-CHING*ANDA.NOEXT.NOWARN.NOEXT

```

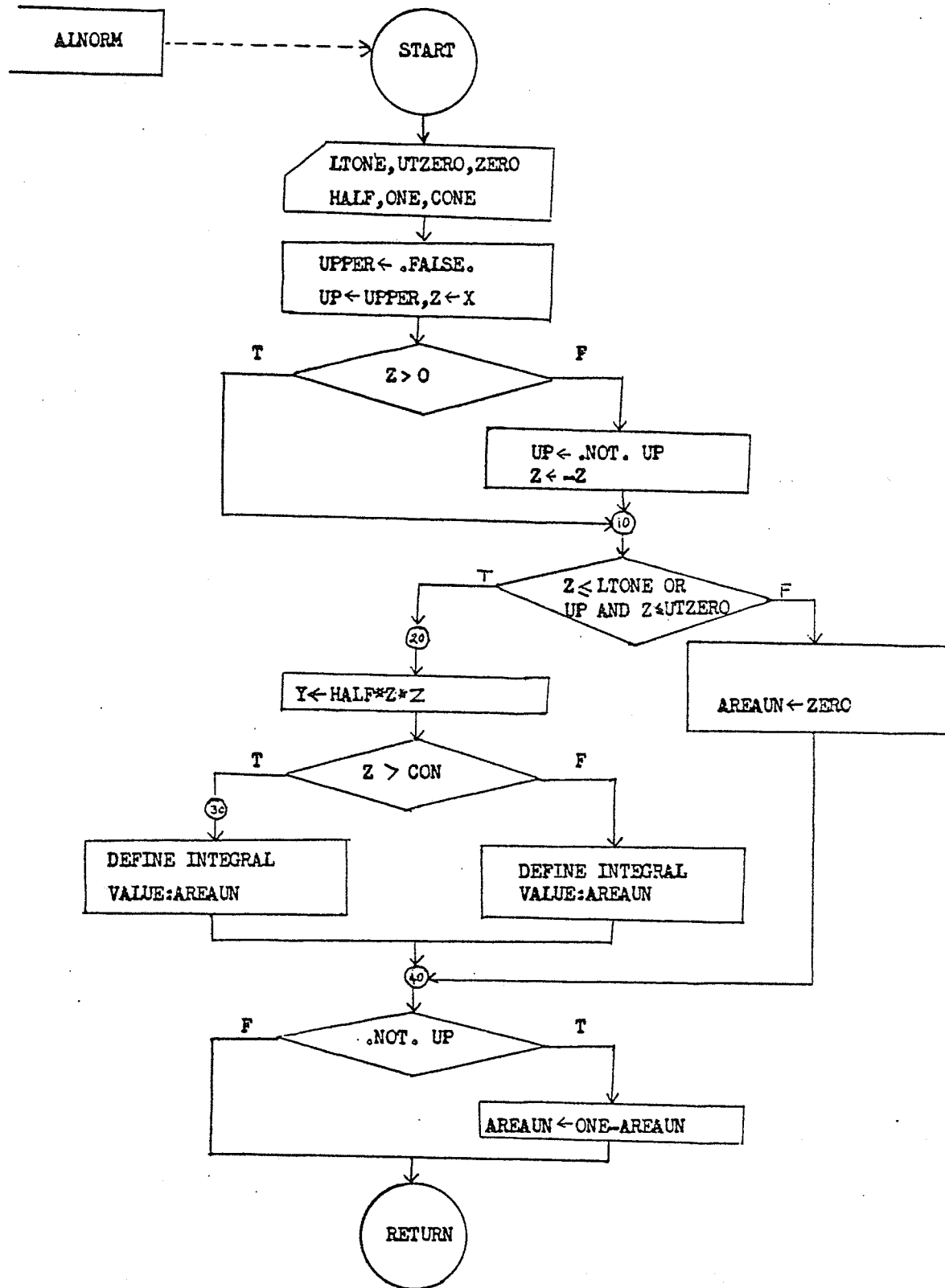
C      THIS IS A MAIN LINE PROGRAM FOR GENERATING THE EDGEWORTH SERIES
C      DISTRIBUTION
C      THE PROGRAM CALCULATES THE DISTRIBUTION FUNCTION AND THE MOMENTS
C      OF THE DISTRIBUTION
1      REAL E(30),O(30),RL(30)
2      REAL V(2000),F(2000),XX(2000)
3      LOGICAL UPPER,J>
4      READ,K,(RL(L),L=1,K)
5      READ,MN
6      DO 300 MM=1,MN
7      READ,N1,N,S
8      KP=1
9      YYY=RANDOM(KP)
10     DO 200 JKL=1,30
11     O(JKL)=0
12     200 CONTINUE
13     PRINT 9
14     9 FORMAT('1')
15     PRINT,'DISTRIBUTION FUNCTION'
16     Q=P=0
17     SNN=STH=SFH=0
18     13 T=-5.0
19     I=1
20     3 X=T-N1
21     CALL ALNORM(X,AREAUN)
22     V(I)=AREAUN-S*((X**2-1)*EXP(-0.5*X*X))/(6.0*(2.0*SQRT(3.141593)))
23     F(I)=X
24     PRINT1,F(I),V(I)
25     1 FORMAT(' ',F5.2,F14.7)
26     I=I+1
27     T=T+0.1
28     IF(T.GT.7.0) GJ TO 5
29     GO TO 3
30     5 KL=N
31     I=I-1
32     DO 10 LL=1,KL
33     R=RANDOM(0)
34     CALL SCAN(F,S,I,V,R,RES,A1,B1,AFA,BFB)
35
36     XX(LL)=RES
37     P=P+XX(LL)
38     10 CONTINUE
39
40     PRINT 61
41     61 FORMAT('1')
42     PRINT,'OBSERVED'
43     DO 40 KX=1,8
44     PRINT 7,O(KX),E(KX)
45     7 FORMAT(' ',F13.7,4X,F13.7)
46     40 CONTINUE

```

```

47      SMEA=P/N
48      DO 11 LL=1,KL
49      SNN=SNN+XX(LL)**2
50      STH=STH+XX(LL)**3
51      11 SFH=SFH+XX(LL)**4
52      SNN=SNN/FLOAT(N)
53      SFH=SFH/FLOAT(N)
54      STH=STH/FLOAT(N)
55      PRINT 22
56      22 FORMAT('1')
57      PRINT,'MEAN(T)=' ,N1,'SAMPLE SIZE=' ,N,'SKEWNESS=' ,S
58      PRINT,'RESULTS OF THE MOMENTS'
59      A=SNN-SMEA**2
60      B=STH-(3.0*SMEA*SNN)+2.0*(SMEA**3)
61      C=SFH-(4.0*SMEA*STH)+(6.0*SMEA**2*SNN)-(3.0*SMEA**4)
62      SMEA=SMEA+N1
63      PRINT 6,SMEA,A,B,C
64      6 FORMAT(' ',MEAN=' ,F13.6,VAR=' ,F13.6,THIRD MOM=' ,F13.6,FOURTH M
        *OM=' ,F13.6)
65      PRINT,'SKEWNESS=' ,+(B/(SQRT(A)**3)),KURT=' ,+(C/A**2)
66      PRINT 60
67      60 FORMAT('1')
69      300 CONTINUE
70      STOP
71      END

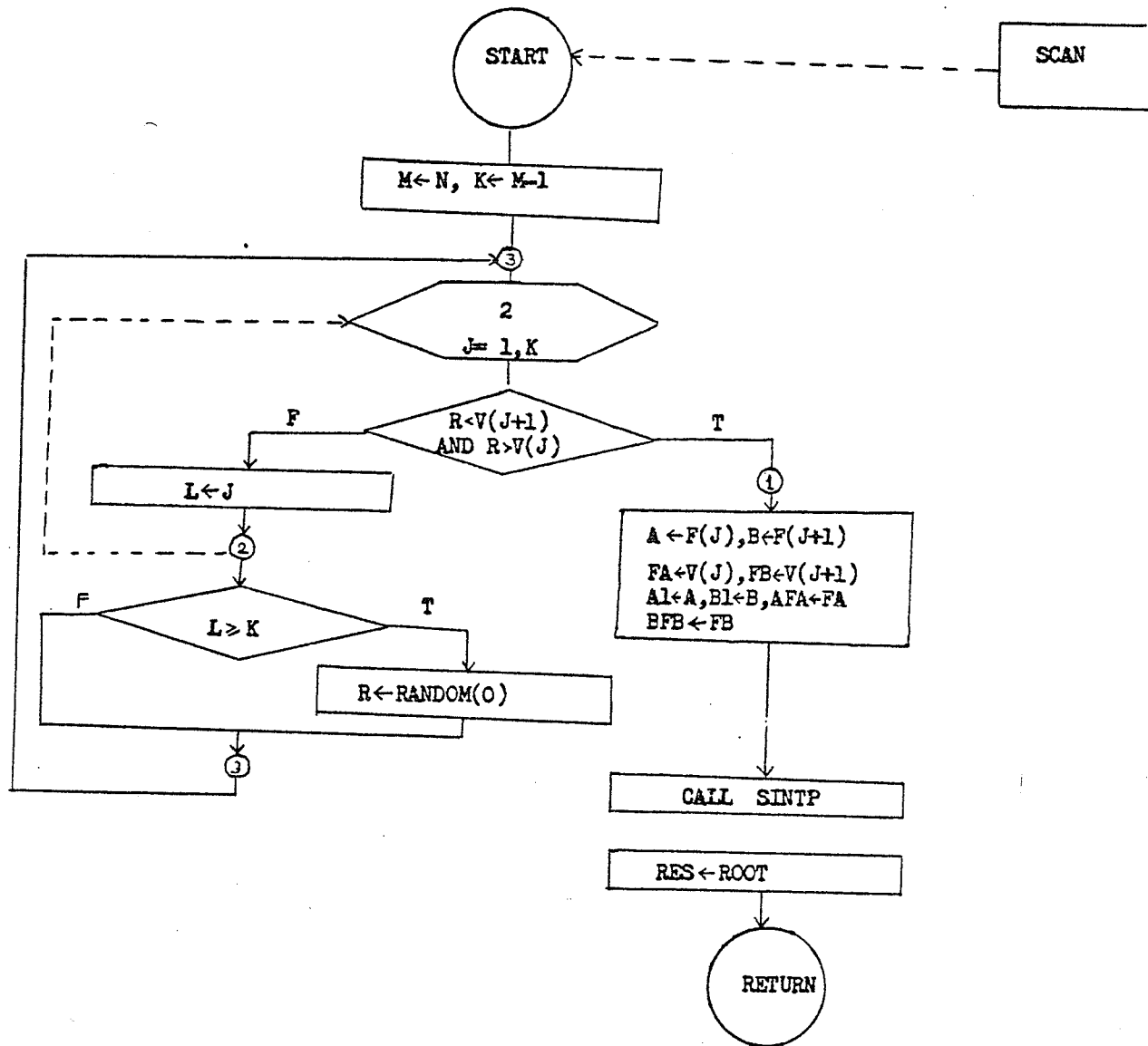
```



```

72      SUBROUTINE ALNDRM(X,AREAUN)
C      THIS SUBROUTINE EVALUATES THE STANDARDIZED NORMAL
C      DISTRIBUTION FUNCTION
73      REAL LTONE,UTZERO,ZERO,HALF,ONE,CONE,Z,Y,X
74      LOGICAL UPPER,J2
75      DATA LTONE,UTZERO/7.0,18.66/
76      DATA ZERO,HALF,ONE,CON/0.0,0.5,1.0,1.28/
77      UPPER=.FALSE.
78      UP=UPPER
79      Z=X
80      IF(Z.GE.ZERO) GO TO 10
81      UP=.NOT.UP
82      Z=-Z
83      10 IF(Z.LE.LTONE.J2.UP.AND.Z.LE.UTZERO) GO TO 20
84      AREAUN=ZERO
85      GO TO 40
86      20 Y=HALF*Z*Z
87      IF(Z.GT.CON) GO TO 30
88      AREAUN=HALF-Z*(0.398942280444-0.399903438504*Y/(Y+5.75885480458-29
      1.8213557808/(Y+2.62433121679+48.6959930692/(Y+5.92885724438))))
89      GO TO 40
90      30 AREAUN=0.398942280385*EXP(-Y)/(Z-3.8052E-8+1.00000615302/(Z+3.9306
      14794E-4+1.98615381364/(Z-0.151679116635+5.29330324926/(Z+4.8385912
      2808-15.1508972451/(Z+0.742380924027+30.789933034/(Z+3.99019417011)
      3))))))
91      40 IF(.NOT.UP) AREAUN=ONE-AREAUN
92      RETURN
93      END

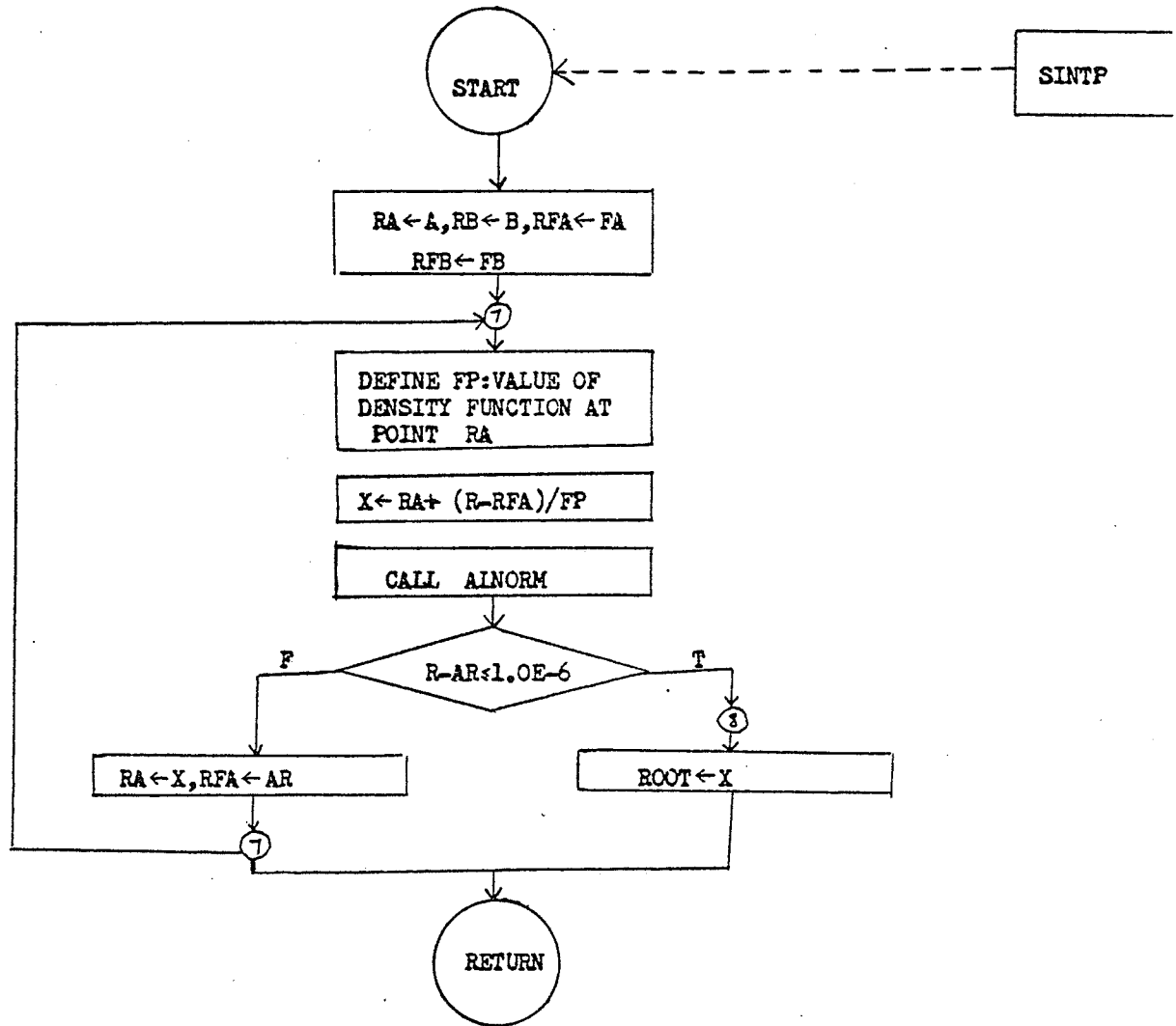
```



```

94      SUBROUTINE SCAN(F,S,N,V,R,RES,A1,B1,AFA,BFB)
C      THIS SUBROUTINE SCANS THROUGH THE DISTRIBUTION FUNCTION TO FIND
C      WHERE THE RANDOM VARIABLE FALLS SO THAT THE NEWTON RAPHSON
C      ROUTINE FOR FINDING THE ROOT(RANDOM VARIABLE) CAN BE USED
95      REAL F(2000),V(2000)
96      M=N
97      K=M-1
98      3 DO 2 J=1,K
99          IF(R.LT.V(J+1).AND.R.GT.V(J)) GO TO 1
100         L=J
101     2 CONTINUE
102         IF(L.GE.K) R=RANDOM(0)
103         GO TO 3
104     1 A=F(J)
105       B=F(J+1)
106       FA=V(J)
107       FB=V(J+1)
108       A1=A
109       B1=B
110       AFA=FA
111       BFB=FB
112       CALL SINTP(S,R,A,B,FA,FB,ROOT)
113       RES=ROOT
114       RETURN
115       END

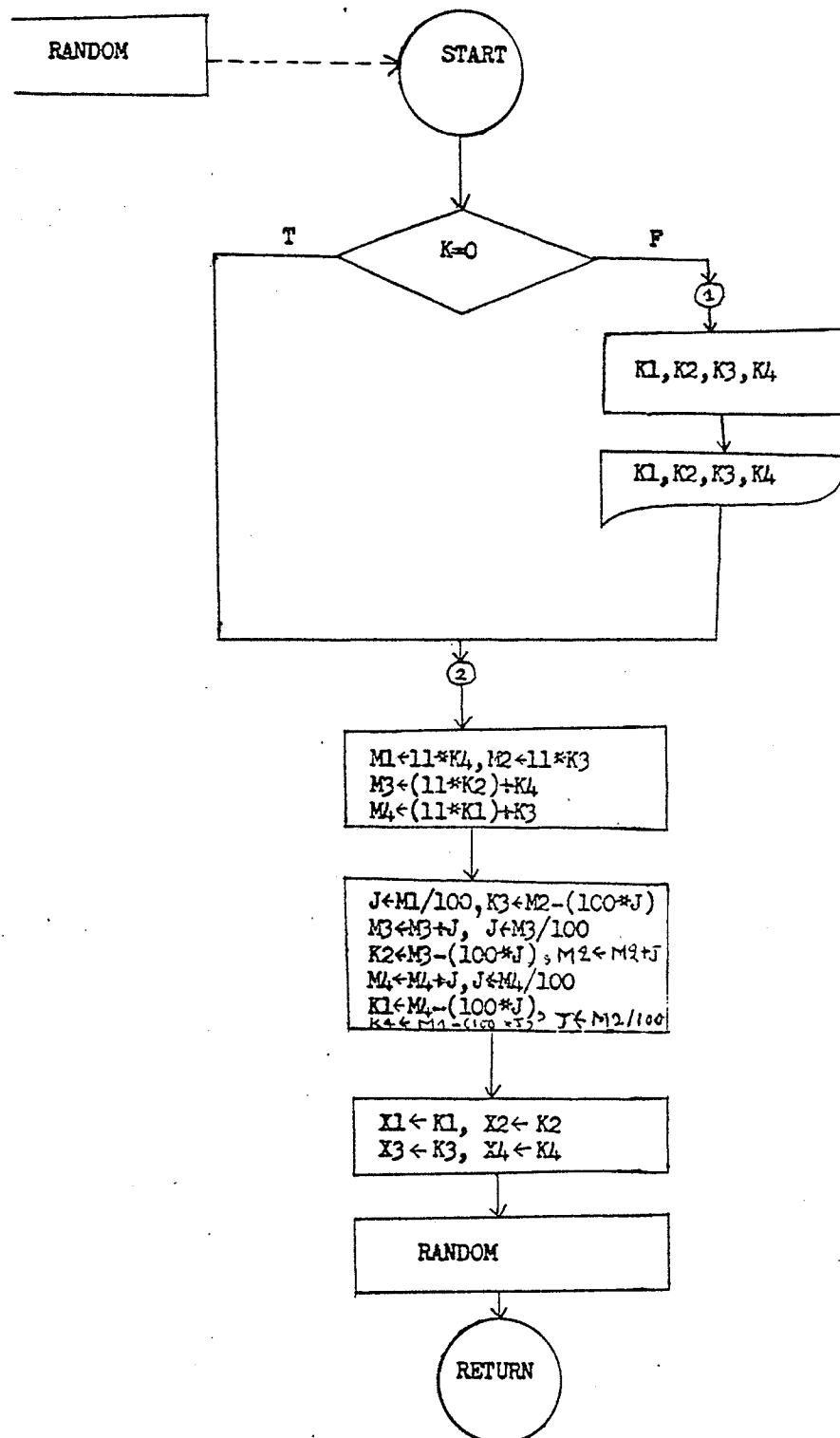
```



```

116      SUBROUTINE SINTP(S,R,A,B,FA,FB,ROOT)
      C      THIS SUBROUTINE ACTUALLY FINDS THE RANDOM VARIABLE
      C      USING THE NEWTON RAPHSON ITERATION ROUTINE
117      LOGICAL UPPER,JP
118      RA=A
119      RB=B
120      RFA=FA
121      RFB=FB
122      7  FP=(1.0-S*(3*RA-RA**3)/6.0)*EXP(-0.5*RA**2)/SQRT(2.0*3.141593)
123      X=RA+(R-RFA)/FP
124      CALL ALNORM(X,AR)
125      IF((R-AR).LE.1.0E-6) GO TO 8
126      RA=X
127      RFA=AR
128      GO TO 7
129      8  ROOT=X
130      RETURN
131      END

```



```

148      FUNCTION RANDOM(K)
C      THIS SUBROUTINE FINDS THE UNIFORM 0,1 RANDOM NUMBER
149      IF(K)1,2,1
150      1 READ(5,3)K1,K2,<3,K4
151      3 FORMAT(4I2)
152      PRINT 27
153      27 FORMAT('1')
154      WRITE(6,4)K1,K2,K3,K4
155      4 FORMAT(31H STARTING NUMBER FOR RANDOM IS 4I2)
156      2 M1=11*K4
157      M2 = 11* K3
158      M3=(11*K2)+K4
159      M4=(11*K1)+K3
160      J=M1/100
161      K4=M1-(100*J)
162      M2=M2+J
163      J=M2/100
164      K3=M2-(100*J)
165      M3=M3+J
166      J=M3/100
167      K2=M3-(100*J)
168      M4=M4+J
169      J=M4/100
170      K1=M4-(100*J)
171      X1=K1
172      X2=K2
173      X3=K3
174      X4=K4
175      RANDOM=X1*1.0E-2+(X2*1.0E-4)+(X3*1.0E-6)+(X4*1.0E-8)
176      RETURN
177      END

```

BIBLIOGRAPHY

- [1] Anderson, T.W. (1951): Classification by multivariate analysis, Psychometrika, 16, 31-50.
- [2] Anderson, T.W. (1958): An introduction to multivariate statistical analysis, John Wiley and Sons, Inc.: New York.
- [3] Anderson, T.W. (1973): Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions, in Discriminant Analysis and Applications (T. Cacoullos, ed.), 17-35. Academic Press: New York.
- [4] Anderson, T.W. (1973): An asymptotic expansion of the distribution of the 'studentized' classification statistic W, Ann. Stat. 1, 964-972.
- [5] Anderson, T.W., Bahadur, R.R. (1962): Classification into two multivariate normal distributions with different covariance matrices, Ann. Math. Stat., 33, 420-431.
- [6] Bartlett, M.S. and Pleare, N.W. (1963): Discrimination in the case of zero mean differences, Biometrika, 50, 17-21.
- [7] Barton, D.E. and Dennis, (1952): The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal, Biometrika, 39, 425-427.
- [8] Broffit, J.D. and Williams, J.S. (1973): Minimum Variance estimators for misclassification probabilities in discriminant analysis, J. Multivariate Anal., 3, 311-327.
- [9] Clumies-Ross, C.W. and Riffenburgh, R.H. (1960): Geometry and linear discrimination, Biometrika, 47, 185-189.
- [10] Das Gupta, S. (1965): Optimum classification rules for classification into two multivariate normal populations, Ann. Math. Stat., 36, 1174-1184.
- [11] Das Gupta, S. (1973): Theories and methods in classification: A review in Discriminant Analysis and Applications (T. Cacoullos, ed.), 77-137, Academic Press, Inc: New York.
- [12] Dunn, O.J. and Varady, P.D. (1966): Probabilities of correct classification in discriminant analysis, Biometrics, 13, 908-924.
- [13] Fisher, R.A. (1936): The use of multiple measurements in tanonomic problems, Ann Eugen., 7, 179-188.

- [14] Gayen, A.K. (1949): The distribution of student's t in random samples of any size drawn from non-normal universes, Biometrika, 87, 236-255.
- [15] Geisser, S. (1967): Estimation associated with linear discriminants, Ann Math. Stat., 38, 807-817.
- [16] Gilbert, E.S. (1969): The effect of unequal variance-covariance matrices on Fisher's linear discriminant function, Biometrics, 25, 505-516.
- [17] Hill, I.D. (1973): Algorithm AS66. The normal integral, Applied Statistics, 22, 424-427.
- [18] Hills, M. (1966): Allocation rules and their error rates, J. Roy. Stat. Soc. Ser. B, 28, 1-32.
- [19] John, S. (1960): On some classification problems I, Sankhya, 22, 301-308.
- [20] John, S. (1960): On some classification statistics, Sankhya, 22, 309-316.
- [21] John, S. (1961): Errors in discrimination, Ann. Math. Stat., 32, 1125-1144.
- [22] John, S. (1963): On classification by the statistics R and Z , Ann. Inst. Stat. Math., 14, 237-246.
- [23] John, S. (1965): Corrections to: On classification by statistics R and Z , Ann. Inst. Stat. Math., 17, 113.
- [24] Johnson, N.L. (1949): Systems of frequency curves generated by methods of translation, Biometrika, 36, 149-176.
- [25] Kendall, M.G. and Stuart, A. (1958): The advanced theory of statistics Vol. 3. Hafner: New York.
- [26] Kendall, M.G. (1966): Discrimination and classification, in Multivariate Anal. Proc. Internatl. Symp. (P.R. Krishna ed.), 165-185, Academic Press, Inc: New York.
- [27] Kudo, A. (1959): The classificatory problem viewed as a two-decision problem. Memoirs of the Faculty of Science, Kyushu Univ., 13, 96-125.
- [28] Lachenbruch, P.A. (1966): Discriminant Analysis when initial samples are misclassified, Technometrics, 8, 657-662.
- [29] Lachenbruch, P.A. (1974): Discriminant analysis when the initial samples are misclassified II: Non-random misclassification models, Technometrics, 16, 419-424.

- [30] Lachenbruch, P.A. (1973): Discriminant Analysis. Hafner Press: New York.
- [31] Lachenbruch, P.A. and Mickey, M.R. (1968): Estimation of error rates in discriminant analysis, Technometrics, 10, 1-11.
- [32] Lachenbruch, P.A., Sneeringer, C. and Revo, L.T. (1973): Robustness of the linear and quadratic discriminant function to certain types of non-normality, Commun. Stat., 1, 39-57.
- [33] Lohnes, P.R. and Cooley, W.W. (1968): Introduction to Statistical Procedures with Computer Exercises. John Wiley and Sons, Inc: New York.
- [34] Mardia, K.V. (1970): Families of bivariate distributions. Charles Griffin: London.
- [35] Marks, S., and Dunn, O.J. (1974): Discriminant functions when covariance matrices are unequal, J. Amer. Stat. Assoc., 69, 555-559.
- [36] McLachlan, G.J. (1972): Asymptotic results for discriminant analysis when initial samples are misclassified, Technometrics, 14, 68-72.
- [37] McLachlan, G.J. (1972): An asymptotic expansion for the variance of the errors of misclassification of the linear discriminant function. Austral. J. Stat., 14, 68-72.
- [38] McLachlan, G.J. (1973): An asymptotic expansion of the expectation of the estimated error rate in discriminant analysis, Austral. J. Stat., 15, 210-214.
- [39] McLachlan, G.J. (1974): An asymptotic unbiased technique for estimating the error rates in discriminant analysis, Biometrics, 30, 239-249.
- [40] McLachlan, G.J. (1974): The asymptotic distribution of the conditional error rate and risk in discriminant analysis, Biometrika, 61, 131-135.
- [41] McLachlan, G.J. (1974): Estimation of the error rates of misclassification on the criterion of asymptotic mean square error, Technometrics, 16, 255-260.
- [42] 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.
- [43] Meyer, P.L. (1966): Introductory Probability and Statistical Applications. Addison-Wesley: Reading, Mass.

- [44] Moran, M.A. (1975): On the expectation of the errors of allocation associated with a linear discriminant function, Biometrika, 62, 141-148.
- [45] Okamoto, M. (1961): Discrimination for variance matrices, Osaka Math. J., 13, 1-39.
- [46] Okamoto, M. (1963): An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Stat., 34, 1286-1301.
- [47] Okamoto, M. (1968): Correction to: An asymptotic expansion for the distribution of the linear discriminant function, Ann. Math. Stat., 39, 1358-1380.
- [48] Penrose, L.S. (1947): Some notes on discrimination, Ann. Eugen., 13, 228-237.
- [49] Rao, C.R. (1950): Statistical inference applied to classificatory problems, Sankhya, 10, 229-256.
- [50] Rao, C.R. (1954): A general theory of discrimination when the information about alternative population distributions is based on samples, Ann. Math. Stat., 25, 651-670.
- [51] Rao, C.R. (1965): Linear Statistical Inference and Its Applications. John Wiley & Sons, Inc.: New York.
- [52] Schmidt, J.W. and Taylor, R.E. (1970): Simulation and Analysis of Industrial Systems. Richard Irwin: Homewood, Illinois.
- [53] Sedransk, N., and Okamoto, M. (1971): Estimation of the probabilities of misclassification for the linear discriminant function in the univariate normal case, Ann. Inst. Stat. Math., 23(3), 419-435.
- [54] Siotani, M. (1975): Comparison of two procedures in discriminant analysis based on Anderson's W-criterion and John-Kudo's Z-criterion, Tech. Report #67, University of Manitoba, Department of Statistics.
- [55] Subrahmaniam, K. (1966): Some contributions to the theory of non-normality - I (univariate case), Sankhya, A, 28, 389-406.
- [56] Smith, C.A.B. (1947): Some examples of discrimination, Ann Eugen., 18, 272-283.

- [57] Sorum, M. (1971): Estimating the conditional probability of misclassification, Technometrics, 13, 333-343.
- [58] Sorum, M. (1972): Three probabilities of misclassification, Technometrics, 14, 309-316.
- [59] Sorum, M. (1972): Estimating the expected and optimal probabilities of misclassification, Technometrics, 14, (4), 935-943.
- [60] Welch, B.L. (1939): Note on discriminant functions, Biometrika, 31, 218-220.