Discriminant Analysis When the Initial Samples are Contaminated

by

Susan W. Ahmed
Department of Biostatistics
University of North Carolina
Chapel Hill, N.C.

Institute of Statistics Mimeo Series # 1012

June, 1975

SUSAN W. AHMED. Discriminant Analysis When the Initial Samples are Contaminated. (Under the direction of PETER A. LACHENBRUCH.)

This paper investigates one aspect of the robustness of Fisher's linear discriminant function—robustness against contamination in the initial samples. Asymptotic results are derived for normal location contamination and for normal scale contamination with the contaminating covariance matrix a multiple of the uncontaminated population covariance matrix. It is shown that when the prior probabilities of the two underlying populations are equal, scale contamination of the initial samples has no effect on our ability to make correct classifications. When the prior probabilities of the two populations are unequal, scale contamination can have a harmful effect. Location contamination can be even more harmful than scale contamination.

Small sample results are presented for the case of scale contamination with equal prior probabilities. Finding that for small samples, Fisher's LDF may perform very poorly when the initial samples are scale contaminated, three types of "robust discriminant functions" are proposed as alternatives to Fisher's LDF. The three types of discriminant functions are evaluated for various combinations of parameters (sample size, distance between the underlying populations, amount of contamination, type of contamination). It is found that for mild contamination, Fisher's LDF performs as well as the "robust procedures." For more severe contamination, a Type I LDF, formed from Fisher's LDF by replacing $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S by robust estimates of location and scale, improves our ability to discriminate.

DISCRIMINANT ANALYSIS WHEN THE INITIAL SAMPLES ARE CONTAMINATED

by

Susan W. Ahmed

A Dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1975

Approved by:

____
Adviser

____
Reader

____
Reader

ACKNOWLEDGMENTS

The author wishes to express her appreciation for the assistance of her advisor, Dr. P. A. Lachenbruch, who suggested the problem and who provided guidance throughout the research. In addition, she would like to thank the other members of her committee, Dr. Lawrence Kupper, Dr. Michael Symons, Dr. Norman Johnson, and Dr. Joan Cornoni for their helpful criticisms and suggestions.

She also wishes to thank her husband, Yahia, for his encouragement throughout her studies and her daughter, Wendy, who was born in the middle of the writing of this dissertation, for her patience during the first six months of her life.

Finally, she wishes to express her thanks to Mrs. Gay Hinnant for her excellent typing of this manuscript.

TABLE OF CONTENTS

Chapter | Page

Chapter Page

LIST OF TABLES

Table		Page

LIST OF FIGURES

Figure | Page

Figure		Page

Figure		Page

CHAPTER I

INTRODUCTION AND REVIEW OF THE LITERATURE

1.1. Introduction

Discriminant analysis deals with the problem of classification. We make a number of measurements on an individual and on the basis of these measurements we wish to classify him into one of several well-defined categories.

More specifically, let $(x_1, \ldots, x_m) = \underline{x}'$ be a vector of observations on m variables. This vector (corresponding to an individual) is known to come from one of k groups or populations $\pi_1, \pi_2, \ldots, \pi_k$. Assume that the probability density function of $\underline{x}$ in π_i is given by $p_X(\underline{x}|\pi_i) = f_i(\underline{x})$ and that the prior probability of π_i is p_i. Assume also that the loss due to assigning an individual to π_i when he is really from π_j is ω_{ij}. Then, having observed $\underline{x}$, a discriminant procedure assigns the individual to π_i if the point $\underline{x}$ is in region R_i, $R_i \cap R_j = \phi, \bigcup_i R_i = \Omega$. In other words, a discriminant procedure partitions m-dimensional space into k mutually exclusive and exhaustive regions such that if $\underline{x}$ is in region R_i, the individual is assigned to π_i. Any particular discriminant procedure tells us how to find R_1, R_2, $\ldots$, R_k.

The optimal discriminant procedure [2,35] consists of choosing $R_1, \ldots, R_k$ to minimize the expected loss:

$$\text{Expected Loss} = \sum_{j=1}^{k} (\text{expected loss}|\pi_j)\ P(\pi_j)$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{k} \left(\begin{array}{l}\text{loss due to}\\ \text{assigning } \underline{x}\\ \text{to } \pi_i\end{array}\middle| \underline{x}\epsilon\pi_j\right)\ P\left(\begin{array}{l}\text{assign}\\ \underline{x} \text{ to}\\ \pi_i\end{array}\middle| \underline{x}\epsilon\pi_j\right)\ P(\pi_j)$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{k} \omega_{ij}\ P(\text{assign to } \pi_i|\pi_j)\ P(\pi_j)$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{k} \omega_{ij} \int_{R_i} \dots \int f_j(\underline{x})d\underline{x}\ p_j$$

$$= \sum_{i=1}^{k} \int_{R_i} \dots \int \sum_{j=1}^{k} p_j\omega_{ij}f_j(\underline{x})d\underline{x}.$$

If we let $\omega_{ij} = 1$ for $i \neq j$ and $\omega_{jj} = 0$ (a loss of 1 for an incorrect classification and no loss for a correct classification) then the formula for the expected loss becomes

$$\text{Expected Loss} = \sum_{i\neq j}\sum P(\text{assign to } \pi_i|\pi_j)\ P(\pi_j)$$

$$= \sum_{i\neq j}\sum P(A_i|\pi_j)\ p_j$$

$$= P(\text{misclassification})$$

$$= 1 - \sum_{j} P(A_j|\pi_j)\ p_j$$

$$= 1 - P(\text{correct classification}).$$

Thus by choosing this particular loss function, the optimal discriminant procedure will minimize the probability of misclassification. To minimize the probability of misclassification, R_i should be taken as the set of points for which $\sum_j p_j\omega_{ij}f_j(\underline{x}) = \sum_{j\neq i}^{k} p_jf_j(\underline{x})$ is minimum, i.e., R_i is the set of points $\underline{x}$ for which $\sum_{j\neq i}^{k} p_jf_j(\underline{x}) = \min_g \sum_{j\neq g}^{k} p_jf_j(\underline{x})$ or equivalently $p_if_i(\underline{x}) = \max_g p_gf_g(\underline{x})$ $g = 1, \dots, k$. Thus the <u>optimal rule</u>

for discrimination is: "Assign $\underline{x}$ to π_i if $p_i f_i(\underline{x}) = \max\{p_1 f_1(\underline{x}), p_2 f_2(\underline{x}), \ldots, p_k f_k(\underline{x})\}$," i.e., assign to the population which has maximum posterior likelihood.

If we consider only two groups, π_1 and π_2, then the optimal rule becomes: "Assign $\underline{x}$ to π_1 if $p_1 f_1(\underline{x}) > p_2 f_2(\underline{x})$; assign $\underline{x}$ to π_2 if $p_1 f_1(\underline{x}) < p_2 f_2(\underline{x})$." If $p_1 f_1(\underline{x}) = p_2 f_2(\underline{x})$ then $\underline{x}$ can be assigned at random to π_1 or π_2.

When the parameters involved in the optimal rule are unknown, the corresponding <u>sample optimal rule</u> can be obtained by taking initial samples of size n_1 and n_2 from π_1 and π_2 and obtaining estimates of the parameters based on these samples. This is the case met most often in practical situations.

If $k = 2$, $f_1(\underline{x}) = N_m(\underline{\mu}_1, \Sigma)$, an m-variate normal distribution with mean vector $\underline{\mu}_1$ and covariance matrix Σ, and $f_2(\underline{x}) = N_m(\underline{\mu}_2, \Sigma)$, then the optimal discriminant rule is: "Assign $\underline{x}$ to π_1 if $\{\underline{x} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)\}' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) > \ln(\frac{1-p}{p})$," where $p = p_1$, $(1-p) = p_2$. The corresponding sample rule assigns $\underline{x}$ to π_1 if $\{\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2)\}' S^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2) > \ln(\frac{1-p}{p})$, where S is a pooled covariance matrix based on samples from π_1 and π_2. This is a common assumption in application: we wish to classify individuals into one of two groups and $\underline{x}$ is assumed to be normally distributed with equal covariance matrices within each group.

In 1936, R. A. Fisher introduced the <u>linear discriminant function</u> [10]. He approached the problem in the following manner: find the linear function $D(\underline{x}) = \underline{\lambda}'\underline{x}$ to be used as a basis for discriminating between π_1 and π_2 which maximizes the quantity

$$\frac{(E(D|\pi_1) - E(D|\pi_2))^2}{V(D)},$$

i.e., find the linear function of the observations which maximizes the ratio of the between group sum of squares to the within group sum of squares. The linear function which satisfies this restriction is $D(\underline{x}) \propto (\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x}$. If we take $D(\underline{x})=(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x} - \frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)$, then the corresponding rule assigns $\underline{x}$ to π_1 if $D(\underline{x}) > \ln(\frac{1-p}{p})$; otherwise $\underline{x}$ is assigned to π_2. The corresponding sample function is $D_s(\underline{x}) = (\bar{\underline{x}}_1-\bar{\underline{x}}_2)'S^{-1}\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1-\bar{\underline{x}}_2)'S^{-1}(\bar{\underline{x}}_1+\bar{\underline{x}}_2)$. The function $D_s(\underline{x})$ is Fisher's linear discriminant function (denoted LDF).

Note that Fisher's LDF is optimal (in the sense that it minimizes the probability of misclassification) when f_1 and f_2 are normal with equal covariance matrices. In general, however, Fisher's rule is different from the optimal rule. Nevertheless, statisticians have applied Fisher's procedure in many different situations: with non-normal distributions, unequal covariance matrices, qualitative data, and when the initial samples are misclassified.

An important question to be considered then is, how robust is the linear discriminant function? How well or how poorly do we do in our classification if we use Fisher's linear discriminant function when the assumptions for optimality do not hold?

1.2. Objectives of the Present Study: Statement of the Problem

This study will investigate one particular aspect of the robustness of the linear discriminant function: robustness against contamination of the initial samples. Very little research has been done on the robustness of the linear discriminant function. Mosteller and Wallace [27] did a study using the LDF and a Bayes procedure; Gilbert [15] has looked at the effects of unequal covariance matrices;

Revo [30] and Gilbert [14] have examined the use of the LDF with qualitative variables; Lachenbruch has investigated the effects of misclassification in the initial samples [20,24]; and Lachenbruch, Sneeringer and Revo have studied the robustness of the LDF to certain continuous non-normal distributions [23]. These results are discussed further in section 1.3.2.

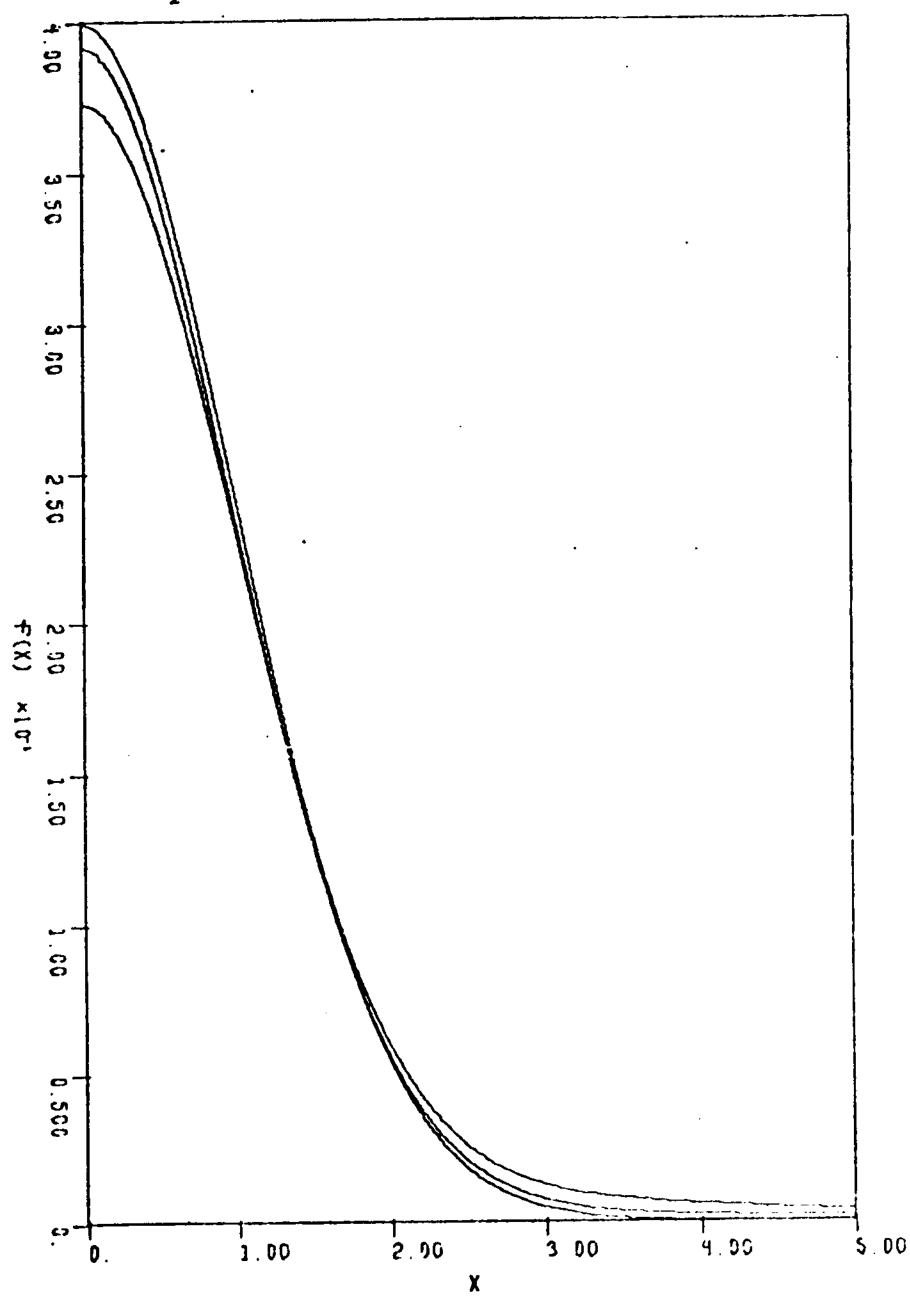
The purpose of the present study is to investigate another aspect of the robustness of the LDF: the effect of contamination of one or both of the initial samples. Assume that we wish to classify individuals into one of two groups:

$$\pi_1: \quad f_1(\underline{x}) = N_m(\underline{\mu}_1, \Sigma)$$
$$\pi_2: \quad f_2(\underline{x}) = N_m(\underline{\mu}_2, \Sigma).$$

We know that Fisher's LDF is the optimal solution to this problem. If the parameters $\underline{\mu}_1, \underline{\mu}_2$, and Σ are unknown, we take samples from π_1 and π_2 and use the sample discriminant function. (Hoel and Peterson [18] have shown that the sample LDF gives an asymptotically optimal solution.)

Suppose that one or both of the initial samples is scale contaminated or location contaminated. We say that sample 1 is <u>scale contaminated</u> if it is really a sample from π_1^c: $(1-\alpha)\ N_m(\underline{\mu}_1, \Sigma) + \alpha G_m(\underline{\mu}_1, \underline{\theta})$ rather than π_1, where α is the proportion of contamination $(0 \leq \alpha \leq 1)$, $\underline{\theta}$ stands for the parameters of G other than $\underline{\mu}_1$, and the <u>contaminating distribution</u> G is assumed to be symmetric, unimodal, centered at $\underline{\mu}_1$, and have greater variability and heavier tails than π_1. Then π_1^c is a symmetric long-tailed distribution. Figure 1.2.1 illustrates a scale contaminated distribution. For this example, we have taken $\alpha = 0.10$, $m = 1$, $\mu_1 = 0$, and $\sigma = 1$. Two contaminating

Fig. 1.2.1. Examples of a scale contaminated density function: α=0.10; $G(\mu_1,\theta)$=N(0,9), Cauchy.*



*The three graphs in the figure (from top to bottom at x=0) correspond to the uncontaminated density N(0,1); $G(\mu_1,\theta)$=N(0,9); $G(\mu_1,\theta)$=Cauchy.

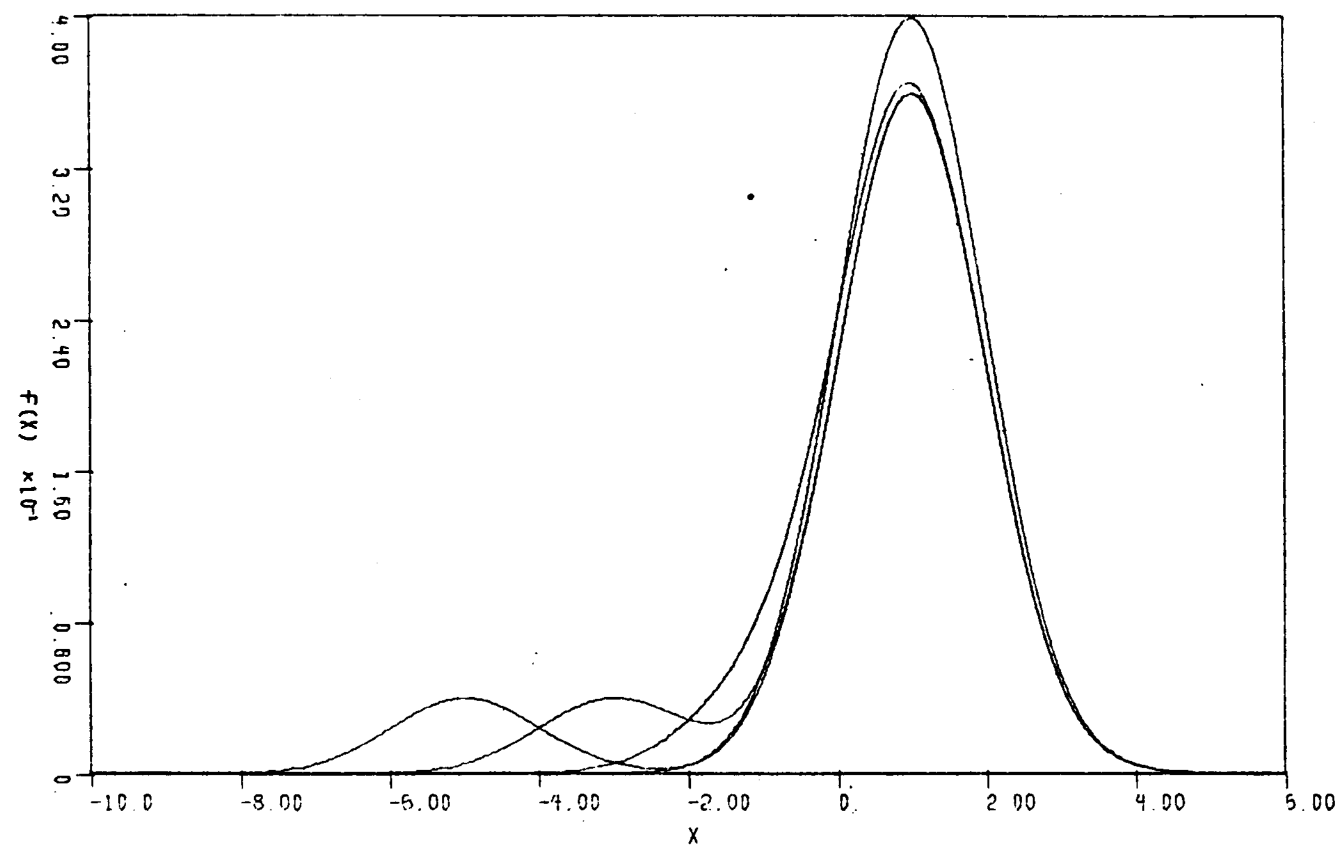
distributions are illustrated: $G(\underline{\mu}_1,\underline{\theta}) = N(0,3)$ and $G(\underline{\mu}_1,\underline{\theta}) = \frac{1}{\pi}\frac{1}{1+x^2}$, along with the N(0,1) density.

Sample 1 is said to be <u>location</u> <u>contaminated</u> if it is really a sample from π_1^c: $(1-\alpha)\ N_m(\underline{\mu}_1,\ddagger) + \alpha G_m(\underline{\nu}_1,\ddagger)$ rather than π_1. In this study, we shall consider the particular location contamination model where $G_m(\underline{\nu}_1,\ddagger) = N_m(\underline{\nu}_1,\ddagger)$, i.e.,the contaminating distribution is normal. The density of π_1 is either an asymmetric unimodal distribution or a distribution with one major mode and one minor mode. Figure 1.2.2 illustrates some location contaminated distributions with $\alpha = 0.10$, $m = 1$, $\mu_1 = 1$, and $\sigma = 1$. Three contaminating distributions are illustrated: N(-1,1), N(-3,1), and N(-5,1), along with the uncontaminated distribution N(1,1).

To see the possibility of such situations arising in practice, consider the following examples.

A scale contaminated sample can arise as an accidental mixture. Suppose that we wish to discriminate between two groups: π_1 is the population of individuals having disease D and π_2 is the population of individuals free of disease D. The classification into π_1 or π_2 will be made on the basis of m measurements $\underline{x}$ thought to be important in determining whether or not an individual has the disease. It is assumed that $\underline{x}$ is normally distributed in both populations. Samples will be taken from the diseased and well populations. The sample from π_1 will be drawn from hospital patients and the m measurements on each of the sample individuals will be made in the hospital with hospital instruments. The sample from π_2 will consist of individuals contacted in a home survey. Some of these measurements will be made in the hospital, while others will be made by local nurses or doctors. Suppose

Fig. 1.2.2. Examples of a location contaminated density function: $\alpha=0.10$; contaminating density $= N(-1,1)$; $N(-3,1)$; $N(-5,1)$.*



*The symmetric unimodal distribution with $f(1)\approx.4$ corresponds to the uncontaminated density function $N(1,1)$. The asymmetric unimodal density function corresponds to contaminating density $N(-1,1)$. The bimodal density function with mode at -3 corresponds to contaminating density $N(-3,1)$. The bimodal density function with mode at -5 corresponds to contaminating density $N(-5,1)$.

that the instruments used by the local nurses and doctors are not as precise as the hospital instruments. Then we have some contamination in the π_2 sample due to the varying precision of the measuring instruments. In the future, when we wish to classify an individual as diseased (π_1) or free of disease (π_2) we will bring him to the hospital and make all measurements with hospital instruments (i.e., we wish to classify him into one of π_1 or π_2, not π_1 or π_2^c). Note that only the sample is contaminated, not the underlying population.

Rather than looking at the way π_2^c is constructed, we can also look at the shape of its density function. It is a long or heavy-tailed distribution. Thus, another example of a scale-contaminated sample is a sample which includes some "gross errors" or "outliers" or in Tukey's words "dirt in the tails" [33]. Hampel [17] suggests that the proportion of "gross errors" or "outliers" in data is normally between 0.1% and 10% depending on circumstances, with several percent being the rule rather than the exception. Any other situation which gives rise to long or heavy-tailed samples will also fit this model.

As an example of location contamination, suppose that we wish to discriminate between individuals with a specified psychiatric disorder (π_1) and normal individuals (π_2). The classification is to be made on the basis of a battery of psychiatric tests. The results of the tests for an individual are represented by the vector $\underline{x} = (x_1, \ldots, x_m)$. The vector of test results $\underline{x}$ is distributed as $N(\underline{\mu}_1, \Sigma)$ in π_1; $\underline{x}$ is distributed as $N(\underline{\mu}_2, \Sigma)$ in π_2. In order to form the linear discriminant function, samples are drawn from π_1 and π_2. The sample from π_1 is drawn from a group of individuals who have recently been diagnosed by a psychiatrist as having the disorder.

The sample from π_2 is drawn from a supposedly normal population. Suppose that among the "normal" sample, there are some undiagnosed cases of the disorder. Then the π_2 sample is location contaminated.

Suppose, then, that although both of the underlying populations are normal, when initial samples are taken, one or both of them are contaminated. What effect does this have on the use of the linear discriminant function? Does the LDF still do a good job of classification? How does the performance of the LDF change as α increases and as G changes? Can the LDF be modified to improve the results with contaminated samples? Suppose we guard against the possibility of contamination of the initial samples by using some "robust" estimation procedures. How do these robust procedures perform when there is no contamination, i.e., when $\alpha = 0$?

Throughout this study, our criterion for evaluating the LDF or any other procedure will be the probability of misclassification, the probability of incorrectly assigning an individual from π_1 to π_2 or assigning an individual from π_2 to π_1. This appears to be a reasonable criterion since our major concern in the classification problem is the correct assignment of an individual to the group of which he is a member.

In Chapter II, we shall investigate several of the questions raised above for large samples. The asymptotic probabilities of misclassification using the LDF with contaminated samples will be derived and compared to the optimal probabilities of misclassification (based on uncontaminated samples).

Chapters III, IV, and V consider the same questions for small sample scale contamination with equal prior probabilities. The small sample investigation is carried out through sampling experiments.

Chapter III describes the approach to the small sample problem and proposes several alternatives to Fisher's LDF when the initial samples are scale contaminated.

Chapter IV presents the results of the sampling experiments for $m = 1$. The performance of the LDF and of the proposed alternatives to the LDF is evaluated.

Chapter V contains the results for $m = 4$.

Chapter VI summarizes our conclusions with suggestions as to when the LDF may be used with contaminated samples and when an alternative should be considered. Also included in Chapter VI are suggestions for further study.

1.3. Review of the Literature

1.3.1. General Background on Discriminant Analysis

The general theory of discriminant analysis is described by Anderson [2]. Fisher introduced his linear discriminant function in 1936 [10] and the optimal discriminant rule described in the introduction was formulated by Welch in 1939 [35]. These are the two basic approaches to the discrimination problem. The first application of discriminant analysis appearing in the literature (1935) is a study concerned with the dating of a series of Egyptian skulls by Barnard [4].

Since that time, the field of discriminant analysis has grown rapidly. The problem of more than two groups has been studied by Lachenbruch [22] and others. New methods for discrimination have been introduced: the logistic discriminant function [32], a general maximum likelihood discriminant [1,9], quadratic discriminant functions [7], and several non-parametric procedures such as statistically equivalent blocks [13] and density estimators [29,31]. In addition, methods have been developed to deal specifically with qualitative data [6,8,14,25, 30].

Other problems encountered in the discrimination procedure have also received attention. Suggestions have been made as to how to select variables to be included in the linear discriminant function [5,26,34]. Lachenbruch and Mickey [21] give a discussion and evaluation of eight approaches to the problem of estimating the probability of misclassification, some of them based on underlying normal distributions, the others empirical estimates.

1.3.2. Robustness of the LDF

The problem of particular interest in this study--robustness of the LDF--has received limited attention.

The first paper in the literature related to this problem is that of Mosteller and Wallace in 1963 [27]. They present a study dealing with the authorship of twelve disputed Federalist papers. The analysis is based on the frequency of occurrence of certain words in the papers. The data is thus discrete or qualitative data. The analysis is carried out by two procedures: (1) Fisher's LDF and (2) a Bayesian approach. The second approach is based on the optimal discriminant rule: assign $\underline{x}$ to π_1 if $p_1 f_1(\underline{x}) > p_2 f_2(\underline{x})$. But since the parameters are unknown, the authors use a Bayesian approach to their estimation, assigning prior distributions for the parameters. The density functions $f_1(\underline{x})$ and $f_2(\underline{x})$ are assumed to be Poisson or negative binomial. The conclusions reached were similar under the two approaches, but weaker for the LDF.

In 1966 and 1974, Lachenbruch studied the effect of incorrect classification of the initial samples [20,24]. In the first of these papers, it is assumed that $\alpha_1 \times 100\%$ of the individuals from π_2: $N_m(\underline{\mu}_2, \Sigma)$ are classified as coming from π_1: $N_m(\underline{\mu}_1, \Sigma)$ and $\alpha_2 \times 100\%$ of the individuals from π_1 are classified as π_2. The individuals are misclassified at random. The discriminant function based on the misclassified samples is

$$D_M(\underline{x}) = K(D_T(\underline{x}) + \frac{\alpha_1 - \alpha_2}{2} \delta^2)$$

where

$$D_T(\underline{x}) = \text{Fisher's LDF (for correctly classified samples)}$$

$$\delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$$

$$K = \frac{1-\alpha_1-\alpha_2}{1 + (\frac{c_1+c_2}{n_1+n_2})\delta^2}$$

$$c_i = \alpha_i(1 - \alpha_i)n_i$$

The asymptotic probabilities of misclassification are given by:

$$P_1 = P(A_2|\pi_1) = \Phi(-(\delta/2)(1 + \alpha_1 - \alpha_2))$$

$$P_2 = P(A_1|\pi_2) = \Phi(-(\delta/2)(1 + \alpha_2 - \alpha_1))$$

$$\alpha_1 + \alpha_2 < 1.$$

If the two proportions of misclassification are equal and the total proportion is not too large, then the probability of misclassification will not be affected. As $|\alpha_1 - \alpha_2|$ increases, the LDF is affected more. Small sample studies reach the same conclusions.

The second paper investigates the same problem but assumes a more practical model of misclassification. The earlier model assumes that the elements are misclassified at random and each observation has the same chance of being initially misclassified. The later model assumes two models. Under the "complete separation model," $\underline{x}$ is initially assigned to π_2 if $(\underline{x} - \underline{\mu}_1)'(\underline{x} - \underline{\mu}_1) > (\underline{x} - \underline{\mu}_2)'(\underline{x} - \underline{\mu}_2)$; otherwise $\underline{x}$ is assigned to π_1. The initial misclassification rate is $\Phi(-\delta/2)$. The second model is a generalization of the first. An individual $\underline{x}$ is initially assigned to π_2 if $(\underline{x} - \underline{\mu}_1)'(\underline{x} - \underline{\mu}_1) > (\underline{x} - \underline{\mu}_2)'(\underline{x} - \underline{\mu}_2)$ and $(\underline{x} - \underline{\mu}_1)'(\underline{x} - \underline{\mu}_1) > \upsilon_1$, where υ_1 is a percentile of the χ^2 distribution. Sampling experiments indicate that the probabilities of misclassification are only slightly affected by this sort of misclassification. In fact, the populations are better separated, i.e., the probability of misclassification based on the misclassified samples is smaller than the corresponding probability based on correctly classified samples. Thus the LDF is robust against limited amounts of misclassification in the initial samples.

Gilbert [14] and Revo [30] have investigated the robustness of the LDF when it is used with qualitative data. Gilbert studied the use of the LDF for discriminating with dichotomous variables. Let $\underline{X} = (X_1, X_2, \ldots, X_m)$ where X_i is a Bernoulli random variable and let $Y = 1$ if $\underline{x} \in \pi_A$, $Y = 0$ if $\underline{x} \in \pi_B$. Let $p_{XY} = P(\underline{X} = \underline{x}, Y = y)$. Then the optimal rule for discriminating between π_A and π_B is: "Assign $\underline{x}$ to π_A if $p_{XB}/p_{XA} < C_A/C_B$" where C_A is the cost due to misclassifying an individual from π_A and C_B is the cost due to misclassifying an individual from π_B. When parameters are unknown, the optimal rule is applied with p_{XB}/p_{XA} replaced by $\hat{p}_{XB}/\hat{p}_{XA}$. Five methods of discrimination are compared:

(1) multinomial model: p_{XB}/p_{XA} is estimated by n_{XB}/n_{XA}

(2) logistic model ($\ell n(p_{XB}/p_{XA}) = 2\underline{\beta}'\underline{x}$) with maximum likelihood estimation of $\underline{\beta}$.

(3) logistic model with minimum logit χ^2 estimates

(4) independent variable model: assuming independence, use the maximum likelihood estimates of p_{XB}/p_{XA}

(5) moment estimates of Fisher's LDF.

Performance of the methods is indicated both by the probability of misclassification and by the correlation between the optimal discriminant function and the function being evaluated. Comparing Fisher's LDF with the optimal rule, assuming parameters known, it was found that in most cases the LDF and the optimal discriminant function were very highly correlated. Computer experiments were used to evaluate the five procedures when parameters were unknown. It was found that procedures (2) through (5) are all better than (1). Based on the probability of misclassification, with six variables the ordering of the procedures from best to worst is (4), (3), (2), (5), (1). The four best procedures are very similar, however. The results indicate that the loss due to using the LDF as opposed to the other procedures is too small to be of much importance.

Revo's study of discrimination with ordered qualitative variables supports Gilbert's conclusion that the LDF is robust against qualitative data. He considered three types of qualitative distributions: Poisson, negative binomial, and discretized normal; m = number of variables = 1, 2; and four discriminant rules:

(1) multinomial model

(2) Fisher's LDF

(3) optimal rule

(4) likelihood ratio procedure based on Hill's "nearest neighbor rule." (If $\underline{X} = (X_1, X_2, \ldots, X_m)$, X_i is a Bernoulli random variable, $\underline{X}$ has a multinomial distribution with 2^m states, the likelihood ratio rule assigns $\underline{x}$ in state j to π_2 if $n_{j2}/n_2 > n_{j1}/n_1$. Hill's N.N. rule assigns $\underline{x}$ to π_2 if $(n_{j2} + n'_{j2})/n_2 > (n_{j1} + n'_{j1})/n_1$, where n'_{j1} is the number of observations from π_1 which fall into states whose $\underline{x}$ values differ from the $\underline{x}$ value in state j with respect to exactly one variable.)

He found that the linear discriminant function performed very well. Using the probability of misclassification as a basis for comparison, it was found that the LDF and the optimal rule performed best, followed by the nearest neighbor rule and last the multinomial model. The LDF did as well as the optimal rule except when the two populations were widely separated.

In 1969, Gilbert [15] investigated the effects of unequal variance-covariance matrices on the performance of the LDF. She considered the case $\pi_1 = N_m(\underline{\mu}_1, \Sigma_1)$, $\pi_2 = N_m(\underline{\mu}_2, \Sigma_2)$ which is equivalent to $\pi_1 = N_m(\underline{0}, I)$, $\pi_2 = N_m(\underline{\gamma}, D)$, where D is diagonal. In this case, the optimal rule is a quadratic discriminant function:

$$Q(\underline{x}) = \tfrac{1}{2}\underline{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\underline{x} + (\underline{\mu}_2'\Sigma_2^{-1} - \underline{\mu}_1'\Sigma_1^{-1})\underline{x} + \tfrac{1}{2}(\underline{\mu}_1'\Sigma_1^{-1}\underline{\mu}_1 - \underline{\mu}_2'\Sigma_2^{-1}\underline{\mu}_2 - \ell n \frac{|\Sigma_2|}{|\Sigma_1|}) .$$

Gilbert considered the special case $\Sigma_2 = d\Sigma_1$ (one covariance matrix is a constant multiple of the other). Then D = dI. The conclusions drawn depend on the values of d, $T^2 = \sum_{i=1}^{m} \frac{\gamma_i^2}{(p + (1-p)d)}$, an estimate of the linear separation between the two populations, and $p = P(\pi_1)$. For $m = 1$, the QDF does little better than the LDF, provided $T^2 \geq 1$. As m increases, the improvement provided by the quadratic discriminant function increases. The squared correlation ρ^2 between the two functions increases as the distance between the populations increases and decreases as d moves away from 1 (the covariance matrices become more different). This effect levels off for $d \leq 0.2$ or $d \geq 5.0$. The squared correlation ρ^2 also decreases as m increases.

If $\rho^2 = 0.8$ is considered good for the LDF, then if $0.2 \leq d \leq 5.0$, Fisher's LDF performs well as long as $T^2 \geq 2$, $m = 1$, or $T^2 \geq 4$, $m = 2$, or $T^2 \geq 8$, $m = 6,10$. Smaller values of T^2 can be tolerated if d is further restricted, while larger values of T^2 permit a wider range of d.

Thus the results indicate that Fisher's LDF will do well if the covariance matrices are not too different and if there is some linear separation between the two populations.

Lachenbruch, Sneeringer, and Revo [23] look at another aspect of the robustness of the linear discriminant function. They consider the robustness of the LDF to certain types of continuous non-normality. They assume m independent variables whose distributions are generated from the normal distribution by applying certain non-linear transformations: (1) log normal: $y = \ell n\, x$, $x = e^y$; (2) logit normal: $y = \ell n \frac{x}{1-x}$, $x = e^y/(1+e^y)$; (3) inverse hyperbolic sine normal: $y = \sinh^{-1}(x)$, $x = \sinh(y)$. The random vector $\underline{y}$ is assumed to be distributed as $N_m(\underline{\mu}_1, I)$ in π_1 and as $N_m(\underline{\mu}_2, I)$ in π_2, $\underline{\mu}_1 = (\delta, 0, \ldots, 0)$, $\underline{\mu}_2 = (0, 0, \ldots, 0)$. Samples of $\underline{x}$ were generated by taking samples of $\underline{y}$ from the m-variate normal distribution and making the appropriate transformation.

The LDF was then used for classification of the observed $\underline{x}$'s. The optimal procedure is to transform the x's to y's and then use the LDF on the y's. Sampling experiments were performed to compare the LDF used on the non-normal distributions to the optimal LDF (using transformed variables). Two other techniques were also considered: (1) modifying the cutoff point for the LDF; (2) using the quadratic discriminant function since the x's will have unequal covariance matrices. For the log-normal and inverse hyperbolic sine normal, there is a significant increase in the mean probability of misclassification due to using the LDF on non-normal data. For the logit normal, there is no such increase. In all cases, the individual probabilities are altered, the logit-normal being the least affected.

Modifying the cutoff point helped to reduce the probability of misclassification when using the non-normal data. The quadratic discriminant function did not seem to help and sometimes even made things worse. Thus the LDF is greatly affected by certain kinds of continuous non-normality.

The literature on the robustness of the LDF indicates that Fisher's LDF is not robust to certain kinds of non-normality; it is somewhat robust to mildly unequal covariance matrices; it is robust to qualitative data and to some misclassification of the initial samples.

1.3.3. Contaminated Distributions

This study will investigate the robustness of the LDF against contamination of one or both of the initial samples. The main surveys of contaminated distributions have been given by Tukey [33] and by Andrews, *et al*. [3]. It is Tukey's discussion which motivates our consideration of the scale contaminated distribution π^c.

Tukey uses the contaminated distribution as a means of investigating the robustness of point estimates or, more precisely, the robustness of the efficiency of point estimates. The sample variance, for example, is the most efficient measure of scale for a normal distribution. Suppose, however, that our distribution is mildly non-normal. The variance may no longer have a high efficiency. The sample mean and variance are very good estimates of location and scale for a normal distribution, but they do not behave well when the underlying distributions are long-tailed.

Tukey investigated several estimates of location and scale. For location he examined the t% trimmed mean with t = 0,1,3,6, and 50. As estimates of scale, he studied logarithmic estimates $\lambda = e^{\tau}$ with τ based on (1) the sample variance; (2) the p^{th} degree absolute moment, $0.2 \leq p \leq 5.0$; (3) the Gaussian mean ($\sum_i e^{-cx_i^2}/n$, $0 \leq c < \infty$); (4) the p% interfractile distance (distance from the observation p% from the bottom of the sample to the observation p% from the top of the sample); (5) the p% trimmed variance. All of Tukey's discussion is based on the contaminated distribution $(1 - \alpha)\, N(0,1) + \alpha N(0,9)$, $0 < \alpha < 0.10$. His major criterion for evaluating the various estimates of scale was the asymptotic effective variance:

$$\begin{array}{l}\text{asymptotic}\\ \text{effective}\\ \text{variance}\end{array} = \frac{\text{asymptotic variance}}{[1 + d/d\tau\ (\text{asymptotic bias})]^2}\,.$$

Tukey's conclusions may be summarized as follows:

(1) The 6% trimmed mean is a good estimator for $0.05 < \alpha < 1.0$; the 3% trimmed mean is good for $0.03 < \alpha < 0.05$ and the 1% trimmed mean for $\alpha < 0.03$.

(2) The sample variance is the worst estimate of scale among those considered.

(3) The use of the Gaussian mean with $c = \frac{1}{4}$ is recommended if α is suspected to be large. For $\alpha < 0.07$, the 2% truncated variance shows good performance.

Tukey's conclusion is, then, that the sample mean and variance are poor estimates of location and scale if scale contamination is suspected.

1.3.4. Robust Estimation

An alternative to the usual LDF which will be investigated in this study is a robust linear discriminant function. The first type of robust LDF to be considered is

$$D_{RS}(\underline{x}) = (\underline{x} - \frac{1}{2}(\tilde{\underline{\mu}}_1 + \tilde{\underline{\mu}}_2))'\tilde{\Sigma}^{-1} (\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)$$

where

$$\tilde{\Sigma} = \frac{(n_1 - 1)\,\tilde{\Sigma}_1 + (n_2 - 1)\,\tilde{\Sigma}_2}{n_1 + n_2 - 2}$$

and $\tilde{\underline{\mu}}_1$, $\tilde{\underline{\mu}}_2$, $\tilde{\Sigma}_1$, $\tilde{\Sigma}_2$ are robust estimates of the corresponding mean vectors and covariance matrices. The second type of function we will consider is based on robust regression coefficients.

A general overview of the concept of robustness has been given by Huber [19]. He has summarized the historical development of the concern over long tailed distributions and the use of the sample mean as an estimate of location. Several different concepts of robustness are reviewed. Three general classes of robust estimators are presented.

Gnanadesikan and Kettenring [16] have given particular attention to multivariate robust estimators of location and scale. They seek estimators which protect against outliers in the tails of the data distributions. Three location estimators are considered:

(1) $\underset{\sim}{y}^*_M$, the vector of medians of the observations on each response variable

(2) $\underset{\sim}{y}^*_{HL}$, the vector of Hodges-Lehman estimators for each variable (the median of all pairwise means of the sample)

(3) $\underset{\sim}{y}^*_{T(\alpha)}$, the vector of α-trimmed means for each response variable, $\alpha = 0.10, 0.25$.

These estimators are vectors of univariate estimators which can be obtained by analyzing the observations on each response separately. To compare the various estimators, Monte Carlo experiments were performed and three measures of the relative efficiency of $\underset{\sim}{y}^*$ with respect to $\bar{\underset{\sim}{y}}$ (the mean vector) were evaluated. These measures are based on the eigenvalues (λ_i) of the Monte Carlo covariance matrix of $\bar{\underset{\sim}{y}}$ and $\underset{\sim}{y}^*$: (1) $\sqrt{\pi\lambda_i}$, (2) $\sum_i \lambda_i$, (3) $\sqrt{\sum_i \lambda_i^2}$. The relative efficiency is the ratio of the value for $\bar{\underset{\sim}{y}}$ to the value for $\underset{\sim}{y}^*$. The vector of medians $\underset{\sim}{y}^*_M$ is the least efficient by any measure, $\underset{\sim}{y}^*_{T(0.25)}$ is somewhat better and the most efficient estimators are $\underset{\sim}{y}^*_{HL}$ and $\underset{\sim}{y}^*_{T(0.10)}$.

The robust estimators of variance s^*_{ii} considered by Gnandesikan and Kettenring are

(1) the α-trimmed variance

(2) the Winsorized variance from an α-Winsorized sample (the α% smallest observations are all replaced by $x_{[\alpha n+1]}$ and the α% largest are replaced by $x_{[n(1-\alpha)]}$).

Both procedures need corresponding location estimators. They suggest an α'-trimmed mean for (1) and an α'-Winsorized mean for (2) with $\alpha' > \alpha$. Both scale estimators must be corrected for bias. A simple method for estimating the covariances as a function of the estimated variances is derived. Monte Carlo studies showed the two estimators to be similar. Finally, methods for obtaining non-singular covariance estimates are discussed.

Andrews, *et al.* [3] have recently completed a very comprehensive study of the robustness of univariate measures of location. The study includes 68 estimators in all, some having appeared earlier in the literature, some being new ones suggested by the authors. The 68 estimators can be divided into four general classes: (1) M-estimates (maximum likelihood type estimators); (2) linear combinations of order statistics; (3) skipped procedures (omitting observations outside a certain range); (4) others. Various properties of these estimators are investigated including: variances for samples of size 5, 10, 20, and 40, percentage points, influence and sensitivity curves (indicating how much the estimator depends on extreme observations), and ease of computation. The distributions considered as alternatives to the normal distribution are contaminated distributions with contaminating densities $G(x) = N(0,k^2)$, Cauchy, $N(0,1)/U(0,1)$, and $N(0,1)/U(0,\frac{1}{3})$. Since these are the same types of contaminating distributions we will be considering in this study, these results are a good basis for our selection of estimators to be used in our first type of robust linear discriminant function.

Gentleman [12] has considered another robust estimate of location, a multivariate location estimate obtained by minimizing q^{th} power deviations. The idea here is the same as that used by Forsythe to obtain robust regression coefficients. Instead of minimizing the squared deviations $\sum_i (y_i - y)^2$, Gentleman suggests minimizing the q^{th}

power deviations: $\sum_i |y_i - y|^q \quad 1 < q < 2$. The estimate is evaluated under several alternative distributions, among them--the contaminated normal.

One of the robust discriminant functions to be considered in this study is based on the relationship between the linear discriminant function and regression analysis. If we perform a regression analysis using $x_1, x_2, \ldots, x_m$ as the independent variables and let the dependent variable y be defined by

$$y = \begin{cases} \dfrac{n_2}{n_1+n_2} & \text{if } \underline{x} \in \pi_1 \\[2ex] -\dfrac{n_1}{n_1+n_2} & \text{if } \underline{x} \in \pi_2 \end{cases},$$

then the resulting regression coefficients will be proportional to Fisher's linear discriminant function coefficients [2,10]. Robust linear discriminant function coefficients might be obtained then, by using robust regression coefficients. Forsythe [11] has given a method for obtaining robust linear regression coefficients by minimizing q^{th} power deviations $(1 < q < 2)$. Let $y_i = \alpha + \beta x_i + \varepsilon_i$; $i = 1, \ldots, n$; $n = n_1 + n_2$. If the ε_i are distributed independently and $N(0,\sigma^2)$ for all i, then the least squares estimators

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$
$$\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

are best linear unbiased estimators. Forsythe considers the case where ε_i is distributed according to a contaminated distribution: $(1-\alpha)\, N(0,1) + \alpha N(S,R^2) \quad 0 < \alpha < 1$. Instead of applying least squares estimates, he minimizes

$$F = \sum_i |y_i - a - bx_i|^q \quad 1 < q < 2.$$

For a symmetric contaminated distribution, he compared the relative mean square errors of the least squares estimates and the q^{th} power estimates and found that as the amount of contamination increased, lower powers of q gave better results. A good compromise is $q = 1.5$. It is always at least 95% as efficient as least squares and can do much better when the underlying distribution is a contaminated normal.

CHAPTER II

THE LDF UNDER CONTAMINATION: ASYMPTOTIC PROPERTIES

2.1. Introduction

In this chapter, we shall develop some theoretical results related to the asymptotic performance of Fisher's linear discriminant function when the initial samples are contaminated. For the asymptotic problem, the contaminating distribution will be assumed to be normal. After deriving an expression for the probability of misclassification under contamination, graphical results will be presented, illustrating the increase in probability of misclassification due to the use of contaminated samples.

2.2. The Model

The general contamination model assuming a normal contaminating distribution may be written as follows:

$$\pi_1: \quad N_m(\underline{\mu}_1, \Sigma)$$

$$\pi_2: \quad N_m(\underline{\mu}_2, \Sigma)$$

$$\pi_1^c: \quad (1-\alpha)\ N_m(\underline{\mu}_1, \Sigma) + \alpha\ N_m(\underline{\nu}_1, \Sigma_1^*) \quad 0 \le \alpha \le 1$$

$$\pi_2^c: \quad (1-\beta)\ N_m(\underline{\mu}_2, \Sigma) + \beta\ N_m(\underline{\nu}_2, \Sigma_2^*) \quad 0 \le \beta \le 1.$$

We wish to classify individuals into one of two groups, π_1 or π_2. Since the parameters $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ are unknown, we will draw an initial sample of size n_1 from π_1 and a sample of size n_2 from π_2 and form the sample linear discriminant function:

$$D_s(\underline{x}) = [\underline{x} - \tfrac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)]' \, S^{-1} \, (\underline{\bar{x}}_1 - \underline{\bar{x}}_2),$$

where $\underline{\bar{x}}_1$ and $\underline{\bar{x}}_2$ are the sample means and S is the pooled covariance matrix. We will then assign an individual with measurements $\underline{x}$ to π_1 if $D_s(\underline{x}) \geq \ln(\frac{1-p}{p})$, where p is the prior probability of population 1; otherwise $\underline{x}$ is assigned to π_2. When the initial samples are drawn, however, they are contaminated, i.e., they are actually samples from π_1^c and π_2^c. The contaminated distributions π_1^c and π_2^c represent a combination of location and scale contamination. The special case of pure location contamination is obtained by setting $\Sigma_1^* = \Sigma = \Sigma_2^*$. The pure scale contamination model is obtained by setting $\underline{\nu}_1 = \underline{\mu}_1$, $\underline{\nu}_2 = \underline{\mu}_2$. The single sample contamination problem is obtained by setting $\alpha = 0$ or $\beta = 0$.

Throughout this study, we shall consider the special case $\Sigma_1^* = k_1\Sigma$, $\Sigma_2^* = k_2\Sigma$, i.e.,

$$\pi_1: \; N_m(\underline{\mu}_1, \Sigma)$$

$$\pi_2: \; N_m(\underline{\mu}_2, \Sigma)$$

$$\pi_1^c: \; (1-\alpha)\, N_m(\underline{\mu}_1, \Sigma) + \alpha\, N_m(\underline{\nu}_1, k_1\Sigma) \quad 0 \leq \alpha \leq 1, \; k_1 \geq 1$$

$$\pi_2^c: \; (1-\beta)\, N_m(\underline{\mu}_2, \Sigma) + \beta\, N_m(\underline{\nu}_2, k_2\Sigma) \quad 0 \leq \beta \leq 1, \; k_2 \geq 1.$$

Since the linear discriminant function is invariant under non-singular linear transformations, the above model can be simplified to the following:

$$\pi_1: \; N_m(\underline{0}, I)$$

$$\pi_2: \; N_m((\delta, 0, \ldots, 0), I)$$

π_1^c: $(1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m((\gamma_{11},\gamma_{12},0,\ldots,0),k_1 I)$

π_2^c: $(1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta\ N_m((\gamma_{21},\gamma_{22},\gamma_{23},0,\ldots,0),k_2 I)$

$0 \le \alpha \le 1 \qquad 0 \le \beta \le 1 \qquad k_1 \ge 1 \qquad k_2 \ge 1 \qquad \delta > 0.$

The simplified model is arrived at by applying a series of nonsingular linear transformations as described below:

(1) We first transform Σ to I. We are given that $\underline{x} \sim N_m(\underline{\mu}_1,\Sigma)$ in π_1; $\underline{x} \sim N_m(\underline{\mu}_2,\Sigma)$ in π_2. The corresponding contaminated populations are π_1^c and π_2^c. There exists a nonsingular matrix C_1 such that $C_1 \Sigma C_1' = I$ (Appendix A, Corollary 4 of Anderson [2]). Applying the linear transformation $\underline{w}_1 = C_1\underline{x}$, our model becomes

π_1: $N_m(C_1\underline{\mu}_1,I)$

π_2: $N_m(C_1\underline{\mu}_2,I)$

π_1^c: $(1-\alpha)\ N_m(C_1\underline{\mu}_1,I) + \alpha\ N_m(C_1\underline{\nu}_1,k_1 I)$

π_2^c: $(1-\beta)\ N_m(C_1\underline{\mu}_2,I) + \beta\ N_m(C_1\underline{\nu}_2,k_2 I)$.

(2) We next translate axes so that $C_1\underline{\mu}_1$ becomes the new origin, i.e., $\underline{w}_2 = \underline{w}_1 - C_1\underline{\mu}_1$. All covariance matrices are left unchanged.

π_1: $N_m(\underline{0},I)$

π_2: $N_m(C_1(\underline{\mu}_2-\underline{\mu}_1),I)$

π_1^c: $(1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m(C_1(\underline{\nu}_1-\underline{\mu}_1),k_1 I)$

π_2^c: $(1-\beta)\ N_m(C_1(\underline{\mu}_2-\underline{\mu}_1),I) + \beta\ N_m(C_1(\underline{\nu}_2-\underline{\mu}_1),k_2 I)$.

(3) We now apply a series of rigid rotations (orthogonal transformations) to transform $C_1(\underline{\mu}_2-\underline{\mu}_1)$, $C_1(\underline{\nu}_1-\underline{\mu}_1)$, and $C_1(\underline{\nu}_2-\underline{\mu}_1)$ to $(\delta,0,\ldots,0)$, $(\gamma_{11},\gamma_{12},0,\ldots,0)$, $(\gamma_{21},\gamma_{22},\gamma_{23},0,\ldots,0)$ respectively. Rotate axes about the origin (leaving the origin fixed) so that $C_1(\underline{\mu}_2-\underline{\mu}_1)$ lies along the first coordinate axis, i.e., apply the orthogonal transformation $\underline{w}_3 = C_2\underline{w}_2$ such that $C_2C_1(\underline{\mu}_2-\underline{\mu}_1) = (\delta,0,\ldots,0)$.

π_1: $N_m(\underline{0},I)$

π_2: $N_m((\delta,0,\ldots,0),I)$

π_1^c: $(1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m(C_2C_1(\underline{\nu}_1-\underline{\mu}_1)=\underline{\nu}_3,k_1I)$

π_2^c: $(1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta\ N_m(C_2C_1(\underline{\nu}_2-\underline{\mu}_1)=\underline{\nu}_4,k_2I)$.

(4) Rotate axes about the first coordinate axis (determined by $\underline{0}$, $(\delta,0,\ldots,0)$) so that $\underline{\nu}_3,\underline{0}$, and $(\delta,0,\ldots,0)$ lie in the same plane. Let $\underline{w}_4 = C_3\underline{w}_3$, C_3 orthogonal, $C_3\underline{\nu}_3 = (\gamma_{11},\gamma_{12},0,\ldots,0)$

π_1: $N_m(\underline{0},I)$

π_2: $N_m((\delta,0,\ldots,0),I)$

π_1^c: $(1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m((\gamma_{11},\gamma_{12},0,\ldots,0),k_1I)$

π_2^c: $(1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta\ N_m(C_3\underline{\nu}_4,k_2I)$.

(5) Rotate axes about the plane determined by $\underline{0}$, $(\delta,0,\ldots,0)$, $(\gamma_{11},\gamma_{12},0,\ldots,0)$ so that $C_3\underline{\nu}_4$, $\underline{0}$, $(\delta,0,\ldots,0)$, $(\gamma_{11},\gamma_{12},0,\ldots,0)$ lie in the same three dimensional space: $\underline{w}_5 = C_4\underline{w}_4$, C_4 orthogonal, $C_4C_3\underline{\nu}_4 = (\gamma_{21},\gamma_{22},\gamma_{23},0,\ldots,0)$.

π_1: $N_m(\underline{0},I)$

π_2: $N_m((\delta,0,\ldots,0),I)$

$$\pi_1^c:\quad (1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m((\gamma_{11},\gamma_{12},0,\ldots,0),k_1 I)$$

$$\pi_2^c:\quad (1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta\ N_m((\gamma_{21},\gamma_{22},\gamma_{23},0,\ldots,0),k_2 I).$$

2.3. The Minimum Probability of Misclassification

Given that we wish to classify an individual into one of the two populations $\pi_1 : N_m(\underline{0},I)$ or $\pi_2: N_m((\delta,0,\ldots,0),I)$, the optimal discriminant function is the same as Fisher's linear discriminant function, i.e., $\underline{x}$ is assigned to π_1 if $(\underline{x} - \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2))'(\underline{\mu}_1-\underline{\mu}_2) > \ell n(\frac{1-p}{p})$, $\underline{\mu}_1 = \underline{0}$, $\underline{\mu}_2 = (\delta,0,\ldots,0)$; otherwise $\underline{x}$ is assigned to π_2. This procedure gives the optimal (minimum) probability of misclassification.

Using the optimal procedure, the probability of incorrectly classifying an individual from π_2 into π_1 is given by:

$$\begin{aligned}
P_2 &\equiv P(\text{assign } \underline{x} \text{ to } \pi_1 | \underline{x}\epsilon\pi_2) \equiv P(A_1|\pi_2) \\
&= P((\underline{x} - \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_2))'(\underline{\mu}_1-\underline{\mu}_2) > \ell n(\tfrac{1-p}{p})|\pi_2) \\
&= P(\underline{x}'(\underline{\mu}_1-\underline{\mu}_2) > \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_2)'(\underline{\mu}_1-\underline{\mu}_2) + \ell n(\tfrac{1-p}{p})|\pi_2) \\
&= \int_h^\infty \phi(t;(\underline{\mu}_1-\underline{\mu}_2)'\underline{\mu}_2,(\underline{\mu}_1-\underline{\mu}_2)'(\underline{\mu}_1-\underline{\mu}_2))dt \\
&\qquad\qquad h = \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_2)'(\underline{\mu}_1-\underline{\mu}_2) + \ell n(\tfrac{1-p}{p}) \\
&= \Phi(-\delta/2 - \frac{\ell n((1-p)/p)}{\delta})
\end{aligned}$$

where $\phi(v;\nu_1,\nu_2)$ represents the normal density function with mean ν_1 and variance ν_2 and Φ denotes the cumulative normal distribution.

Similarly,

$$P_1 \equiv P(A_2|\pi_1) = \Phi(-\delta/2 + \frac{\ell n((1-p)/p)}{\delta}).$$

The total optimal probability of misclassification is $P = pP_1 + (1-p)P_2$. This is the probability of misclassification using the optimal discriminant function, assuming known parameters. Since the sample discriminant function based on uncontaminated samples, $D_s(\underline{x}) = (\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1+\bar{\underline{x}}_2))' \, S^{-1}(\bar{\underline{x}}_1-\bar{\underline{x}}_2)$, approaches $(\underline{x} - \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2))'(\underline{\mu}_1-\underline{\mu}_2)$ as $n \to \infty$, P is also the asymptotic probability of misclassification based on uncontaminated samples.

2.4. The Asymptotic Probability of Misclassification When the Initial Samples are Contaminated

In order to evaluate the asymptotic effect of contamination in the initial samples on the performance of Fisher's linear discriminant function, we derive an expression for the asymptotic probability of misclassification due to using the LDF with contaminated samples and compare this probability to the optimal probability (based on uncontaminated samples).

Theorem 2.4

Given the contamination model

π_1: $N_m(\underline{0},I)$

π_2: $N_m((\delta,0,\ldots,0),I)$

π_1^c: $(1-\alpha)\ N_m(\underline{0},I) + \alpha\ N_m((\gamma_{11},\gamma_{12},0,0,\ldots,0),k_1 I)$

π_2^c: $(1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta\ N_m((\gamma_{21},\gamma_{22},\gamma_{23},0,\ldots,0),k_2 I)$

$0 \le \alpha \le 1 \qquad 0 \le \beta \le 1 \qquad k_1 \ge 1 \qquad k_2 \ge 1 \qquad \delta > 0$

then the asymptotic probabilities of misclassification using Fisher's linear discriminant function to discriminate between π_1 and π_2 with contaminated initial samples from π_1^c and π_2^c are given by:

$$P_1^c = \Phi\left(\frac{\ln(\frac{1-p}{p})}{g_1\delta} - g_2\ \delta/2\right)$$

$$P_2^c = \Phi\left[\frac{-\ell n(\frac{1-p}{p})}{g_1\delta} - g_3 \quad \delta/2\right]$$

$$P^c = pP_1^c + (1-p)P_2^c$$

where

$$g_1 = \{[(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{11}+(\alpha c_{12}-\beta c_{22})a^{12}-\beta c_{23}a^{13}]^2$$
$$+ [(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{21}+(\alpha c_{12}-\beta c_{22})a^{22}-\beta c_{23}a^{23}]^2$$
$$+ [(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{31}+(\alpha c_{12}-\beta c_{22})a^{32}-\beta c_{23}a^{33}]^2\}^{1/2}$$

$$g_2 = -\frac{1}{g_1}\{[\alpha c_{11}-(1-\beta)-\beta c_{21}][a^{11}(\alpha c_{11}+(1-\beta)+\beta c_{21})$$
$$+ a^{21}(\alpha c_{12}+\beta c_{22})+a^{31}\beta c_{23}]$$
$$+ [\alpha c_{12}-\beta c_{22}][a^{12}(\alpha c_{11}+(1-\beta)+\beta c_{21})+a^{22}(\alpha c_{12}+\beta c_{22})+a^{32}\beta c_{32}]$$
$$- \beta c_{23}[a^{13}(\alpha c_{11}+(1-\beta)+\beta c_{21})+a^{23}(\alpha c_{12}+\beta c_{22})+a^{33}\beta c_{23}]\}$$

$$g_3 = -g_2 - \frac{2}{g_1}\{[\alpha c_{11}-(1-\beta)-\beta c_{21}]a^{11}+[\alpha c_{12}-\beta c_{22}]a^{12}-\beta c_{23}a^{13}\}$$

$$c_{11} = \gamma_{11}/\delta \qquad c_{12} = \gamma_{12}/\delta$$
$$c_{21} = \gamma_{21}/\delta \qquad c_{22} = \gamma_{22}/\delta \qquad c_{23} = \gamma_{23}/\delta$$

$$a^{11} = \frac{1}{|A_{11}|}[(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)(a_0+a_2c_{23}^2\delta^2)-(a_2c_{22}c_{23}\delta^2)^2]$$

$$a^{12} = a^{21} = \frac{1}{|A_{11}|}[(a_2c_{22}c_{23}\delta^2)(-a_2(1-c_{21})c_{23}\delta^2)$$
$$- (a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2)(a_0+a_2c_{23}^2\delta^2)]$$

$$a^{13} = a^{31} = \frac{1}{|A_{11}|}[(a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)(a_2c_{22}c_{23}\delta^2)$$
$$+ a_2(1-c_{21})c_{23}\delta^2(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)]$$

$$a^{22} = \frac{1}{|A_{11}|} [(a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_0+a_2c_{23}^2\delta^2)$$
$$- (a_2(1-c_{21})c_{23}\delta^2)^2]$$

$$a^{23} = a^{32} = \frac{1}{|A_{11}|} [(a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)(-a_2(1-c_{21})c_{23}\delta^2)$$
$$- (a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_2c_{22}c_{23}\delta^2)]$$

$$a^{33} = \frac{1}{|A_{11}|} [(a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)$$
$$- (a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)^2]$$

$$a_0 = \frac{(n_1+n_2-2)+(n_1-1)\alpha(k_1-1)+(n_2-1)\beta(k_2-1)}{n_1+n_2-2}$$

$$a_1 = \frac{(n_1-1)\alpha(1-\alpha)}{n_1+n_2-2}$$

$$a_2 = \frac{(n_2-1)\beta(1-\beta)}{n_1+n_2-2}$$

$$|A_{11}| = [a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2][a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2][a_0+a_2c_{23}^2\delta^2]$$
$$- [a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2][a_2c_{22}c_{23}\delta^2][a_2(1-c_{21})c_{23}\delta^2]$$
$$- [a_2(1-c_{21})c_{23}\delta^2][a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2][a_2c_{22}c_{23}\delta^2]$$
$$- [a_2(1-c_{21})c_{23}\delta^2]^2[a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2]$$
$$- [a_2c_{22}c_{23}\delta^2]^2[a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2]$$
$$- [a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2]^2[a_0+a_2c_{23}^2\delta^2].$$

PROOF:

The optimal discriminant function for discriminating between π_1 and π_2 is given by

$$\begin{aligned} D_{OPT}(\underline{x}) &= (\underline{x} - \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_2))'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2) \\ &= (\underline{x} - \tfrac{1}{2}(\underline{\mu}_1+\underline{\mu}_2))'(\underline{\mu}_1-\underline{\mu}_2) \\ &= -x_1\delta + \tfrac{1}{2}\delta^2 \end{aligned}$$

where $\underline{\mu}_1 = \underline{0}$ and $\underline{\mu}_2 = (\delta,0,\ldots,0)$. The corresponding sample discriminant function is given by

$$D_S(\underline{x}) = (\underline{x} - \tfrac{1}{2}(\bar{\underline{x}}_1+\bar{\underline{x}}_2))'S^{-1}(\bar{\underline{x}}_1-\bar{\underline{x}}_2)$$

where $\bar{\underline{x}}_1$ and $\bar{\underline{x}}_2$ are estimates of $\underline{\mu}_1$ and $\underline{\mu}_2$ based on samples from π_1 and π_2 and S is the pooled covariance matrix based on the same samples.

Because of the contamination in the samples, however, we are actually using

$$D_S^c(\underline{x}) = (\underline{x} - \tfrac{1}{2}(\bar{\underline{x}}_1^c+\bar{\underline{x}}_2^c))'(S^c)^{-1}(\bar{\underline{x}}_1^c-\bar{\underline{x}}_2^c)$$

where $\bar{\underline{x}}_1^c$ and $\bar{\underline{x}}_2^c$ are the sample means for the contaminated samples from π_1^c and π_2^c and $S^c = [(n_1-1)S_1^c+(n_2-1)S_2^c]/(n_1+n_2-2)$. The matrices S_1^c and S_2^c are the sample covariance matrices for the contaminated samples.

As n_1 and n_2 become large,

$$\begin{aligned} \bar{\underline{x}}_1^c \to E(\underline{x}|\pi_1^c) &= (1-\alpha)\underline{\mu}_1+\alpha\underline{\gamma}_1 \\ &= (\alpha\gamma_{11},\alpha\gamma_{12},0,0,\ldots,0) \\ \bar{\underline{x}}_2^c \to E(\underline{x}|\pi_2^c) &= (1-\beta)\underline{\mu}_2+\beta\underline{\gamma}_2 \\ &= ((1-\beta)\delta+\beta\gamma_{21},\beta\gamma_{22},\beta\gamma_{23},0,\ldots,0) \end{aligned}$$

$$S_1^c \to V(\underline{x}|\pi_1^c)$$

$$= E[(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)'|\pi_1^c]$$

$$= (1-\alpha)E[(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)'|\underline{x}\sim N(\underline{\mu}_1,I)]$$

$$+ \alpha E[(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)(\underline{x}-(1-\alpha)\underline{\mu}_1-\alpha\underline{\gamma}_1)'|\underline{x}\sim N(\underline{\gamma}_1,k_1I)]$$

$$= (1-\alpha)E[(\underline{x}-\underline{\mu}_1+\alpha(\underline{\mu}_1-\underline{\gamma}_1))(\underline{x}-\underline{\mu}_1+\alpha(\underline{\mu}_1-\underline{\gamma}_1))'|\underline{x}\sim N(\underline{\mu}_1,I)]$$

$$+ \alpha E[(\underline{x}-\underline{\gamma}_1-(1-\alpha)(\underline{\mu}_1-\underline{\gamma}_1))(\underline{x}-\underline{\gamma}_1-(1-\alpha)(\underline{\mu}_1-\underline{\gamma}_1))'|\underline{x}\sim N(\underline{\gamma}_1,k_1I)]$$

$$= (1-\alpha)I+\alpha^2(1-\alpha)(\underline{\mu}_1-\underline{\gamma}_1)(\underline{\mu}_1-\underline{\gamma}_1)'+\alpha k_1I+\alpha(1-\alpha)^2(\underline{\mu}_1-\underline{\gamma}_1)(\underline{\mu}_1-\underline{\gamma}_1)'$$

$$= [1+\alpha(k_1-1)]I+\alpha(1-\alpha)(\underline{\mu}_1-\underline{\gamma}_1)(\underline{\mu}_1-\underline{\gamma}_1)'$$

$$S_2^c \to V(\underline{x}|\pi_2^c)$$

$$= [1+\beta(k_2-1)]I+\beta(1-\beta)(\underline{\mu}_2-\underline{\gamma}_2)(\underline{\mu}_2-\underline{\gamma}_2)'$$

$$S^c \to \{(n_1-1)[(1+\alpha(k_1-1))I+\alpha(1-\alpha)(\underline{\mu}_1-\underline{\gamma}_1)(\underline{\mu}_1-\underline{\gamma}_1)']$$

$$+(n_2-1)[(1+\beta(k_2-1))I+\beta(1-\beta)(\underline{\mu}_2-\underline{\gamma}_2)(\underline{\mu}_2-\underline{\gamma}_2)']\}/(n_1+n_2-2)$$

$$= \left[\frac{(n_1+n_2-2)+(n_1-1)\alpha(k_1-1)+(n_2-1)\beta(k_2-1)}{n_1+n_2-2}\right]I$$

$$+ \frac{(n_1-1)\alpha(1-\alpha)}{(n_1+n_2-2)}(\underline{\mu}_1-\underline{\gamma}_1)(\underline{\mu}_1-\underline{\gamma}_1)'$$

$$+ \frac{(n_2-1)\beta(1-\beta)}{(n_1+n_2-2)}(\underline{\mu}_2-\underline{\gamma}_2)(\underline{\mu}_2-\underline{\gamma}_2)'$$

$$= \begin{bmatrix} A_{11} & 0 \\ 0 & a_0I_{m-3} \end{bmatrix} = A$$

where

$$A_{11}=\begin{bmatrix} a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2 & a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2 & -a_2(1-c_{21})c_{23}\delta^2 \\ a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2 & a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2 & a_2c_{22}c_{23}\delta^2 \\ -a_2(1-c_{21})c_{23}\delta^2 & a_2c_{22}c_{23}\delta^2 & a_0+a_2c_{23}^2\delta^2 \end{bmatrix}$$

$$c_{ij} = \gamma_{ij}/\delta$$

$$a_0 = \frac{(n_1+n_2-2)+(n_1-1)\alpha(k_1-1)+(n_2-1)\beta(k_2-1)}{n_1+n_2-2}$$

$$a_1 = \frac{(n_1-1)\alpha(1-\alpha)}{n_1+n_2-2}$$

$$a_2 = \frac{(n_2-1)\beta(1-\beta)}{n_1+n_2-2} \quad .$$

Therefore, for large samples,

$$D_s^c(\underline{x}) \to \{\underline{x} - \tfrac{1}{2}[(1-\alpha)\underline{\mu}_1+\alpha\underline{\gamma}_1+(1-\beta)\underline{\mu}_2+\beta\underline{\gamma}_2]\}' \; A^{-1}$$

$$\{(1-\alpha)\underline{\mu}_1+\alpha\underline{\gamma}_1-(1-\beta)\underline{\mu}_2-\beta\underline{\gamma}_2\}.$$

The inverse of the matrix A is

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & \frac{1}{a_0} I_{m-3} \end{bmatrix}$$

$$= \begin{bmatrix} a^{11} & a^{12} & a^{13} & \\ a^{21} & a^{22} & a^{23} & 0 \\ a^{31} & a^{32} & a^{33} & \\ & 0 & & \frac{1}{a_0} I_{m-3} \end{bmatrix}$$

where the a^{ij} are the elements of A_{11}^{-1}

$$a^{11} = \frac{1}{|A_{11}|}[(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)(a_0+a_2c_{23}^2\delta^2)-(a_2c_{22}c_{23}\delta^2)^2]$$

$$a^{12} = a^{21} = \frac{1}{|A_{11}|}[(a_2c_{22}c_{23}\delta^2)(-a_2(1-c_{21})c_{23}\delta^2)$$
$$-(a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2)(a_0+a_2c_{23}^2\delta^2)]$$

$$a^{13} = a^{31} = \frac{1}{|A_{11}|}[(a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)(a_2c_{22}c_{23}\delta^2)$$
$$+(a_2(1-c_{21})c_{23}\delta^2)(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)]$$

$$a^{22} = \frac{1}{|A_{11}|}[(a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_0+a_2c_{23}^2\delta^2)-(a_2(1-c_{21})c_{23}\delta^2)^2]$$

$$a^{23} = a^{32} = \frac{1}{|A_{11}|}[(a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)(-a_2(1-c_{21})c_{23}\delta^2)$$
$$-(a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_2c_{22}c_{23}\delta^2)]$$

$$a^{33} = \frac{1}{|A_{11}|}[(a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2)(a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2)$$
$$-(a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2)^2].$$

$$|A_{11}| = \{[a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2][a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2][a_0+a_2c_{23}^2\delta^2]$$
$$-[a_1c_{11}c_{12}\delta^2-a_2(1-c_{21})c_{22}\delta^2][a_2c_{22}c_{23}\delta^2][a_2(1-c_{21})c_{23}\delta^2]$$
$$-[a_2(1-c_{21})c_{23}\delta^2][a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2][a_2c_{22}c_{23}\delta^2]$$
$$-[a_2(1-c_{21})c_{23}\delta^2]^2[a_0+a_1c_{12}^2\delta^2+a_2c_{22}^2\delta^2]$$
$$-[a_2c_{22}c_{23}\delta^2]^2[a_0+a_1c_{11}^2\delta^2+a_2(1-c_{21})^2\delta^2]$$
$$-[a_1c_{11}c_{12}\delta^2-a_2c_{22}(1-c_{21})\delta^2]^2[a_0+a_2c_{23}^2\delta^2]\}.$$

Thus

$$
\begin{aligned}
D_s^c(\underline{x}) &\to \{\underline{x} - \tfrac{1}{2}[\alpha\underline{\gamma}_1+(1-\beta)\underline{\mu}_2+\beta\underline{\gamma}_2]\}' \; A^{-1}\{\alpha\underline{\gamma}_1-(1-\beta)\underline{\mu}_2-\beta\underline{\gamma}_2\} \\
&= [(x_1 - \tfrac{1}{2}\alpha c_{11}\delta - \tfrac{1}{2}(1-\beta)\delta - \tfrac{1}{2}\beta c_{21}\delta)a^{11} + \\
&\quad (x_2 - \tfrac{1}{2}\alpha c_{12}\delta - \tfrac{1}{2}\beta c_{22}\delta)a^{21}+(x_3 - \tfrac{1}{2}\beta c_{23}\delta)a^{31}] \cdot \\
&\qquad [\alpha c_{11}\delta-(1-\beta)\delta-\beta c_{21}\delta] \\
&\quad +[(x_1 - \tfrac{1}{2}\alpha c_{11}\delta - \tfrac{1}{2}(1-\beta)\delta - \tfrac{1}{2}\beta c_{21}\delta)a^{12} + \\
&\quad (x_2 - \tfrac{1}{2}\alpha c_{12}\delta - \tfrac{1}{2}\beta c_{22}\delta)a^{22}+(x_3 - \tfrac{1}{2}\beta c_{23}\delta)a^{32}] \cdot \\
&\qquad [\alpha c_{12}\delta-\beta c_{22}\delta] \\
&\quad +[(x_1 - \tfrac{1}{2}\alpha c_{11}\delta - \tfrac{1}{2}(1-\beta)\delta - \tfrac{1}{2}\beta c_{21}\delta)a^{13} + \\
&\quad (x_2 - \tfrac{1}{2}\alpha c_{12}\delta - \tfrac{1}{2}\beta c_{22}\delta)a^{23}+(x_3 - \tfrac{1}{2}\beta c_{23}\delta)a^{33}][-\beta c_{23}\delta] \\
&= d_0+d_1x_1+d_2x_2+d_3x_3
\end{aligned}
$$

where

$$
\begin{aligned}
d_0 = -\tfrac{1}{2}\delta^2\{&[\alpha c_{11}-(1-\beta)-\beta c_{21}][a^{11}(\alpha c_{11}+(1-\beta)+\beta c_{21}) \\
&\qquad +a^{21}(\alpha c_{12}+\beta c_{22})+a^{31}(\beta c_{23})] \\
&+[\alpha c_{12}-\beta c_{22}][a^{12}(\alpha c_{11}+(1-\beta)+\beta c_{21}) \\
&\qquad +a^{22}(\alpha c_{12}+\beta c_{22})+a^{32}(\beta c_{23})] \\
&-[\beta c_{23}][a^{13}(\alpha c_{11}+(1-\beta)+\beta c_{21}) \\
&\qquad +a^{23}(\alpha c_{12}+\beta c_{22})+a^{33}(\beta c_{23})]\}
\end{aligned}
$$

$$d_1 = \delta\{[\alpha c_{11}-(1-\beta)-\beta c_{21}]a^{11}+[\alpha c_{12}-\beta c_{22}]a^{12}-\beta c_{23}a^{13}\}$$

$$d_2 = \delta\{[\alpha c_{11}-(1-\beta)-\beta c_{21}]a^{21}+[\alpha c_{12}-\beta c_{22}]a^{22}-\beta c_{23}a^{23}\}$$

$$d_3 = \delta\{[\alpha c_{11}-(1-\beta)-\beta c_{21}]a^{31}+[\alpha c_{12}-\beta c_{22}]a^{32}-\beta c_{23}a^{33}\}.$$

We assign $\underline{x}$ to π_1 if $D_s^c(\underline{x}) > \ln(\frac{1-p}{p})$; otherwise $\underline{x}$ is assigned to π_2. The probabilities of misclassification are given by

$$P_1^c = P(D_s^c(\underline{x}) \leq \ln(\frac{1-p}{p})\,|\,\pi_1)$$

$$P_2^c = P(D_s^c(\underline{x}) > \ln(\frac{1-p}{p})\,|\,\pi_2)$$

$$P^c = pP_1^c + (1-p)P_2^c.$$

Since $D_s^c(\underline{x}) \to d_0+d_1x_1+d_2x_2+d_3x_3$ and $(x_1,x_2,x_3) \sim N(\underline{\nu}_i,I)$ in π_i, $\underline{\nu}_1 = (0,0,0)$, $\underline{\nu}_2 = (\delta,0,0)$, therefore for large n,

$$D_s^c(\underline{x}) \sim N(d_0, d_1^2+d_2^2+d_3^2) \text{ in } \pi_1$$

$$D_s^c(\underline{x}) \sim N(d_0+d_1\delta, d_1^2+d_2^2+d_3^2) \text{ in } \pi_2.$$

Then

$$P_1^c = \Phi\left(\frac{\ln(\frac{1-p}{p})-d_0}{\sqrt{d_1^2+d_2^2+d_3^2}}\right) = \Phi\left(\frac{\ln(\frac{1-p}{p})}{g_1\delta} - g_2\delta/2\right)$$

$$P_2^c = \Phi\left(\frac{-\ln(\frac{1-p}{p})+d_0+d_1\delta}{\sqrt{d_1^2+d_2^2+d_3^2}}\right) = \Phi\left(\frac{-\ln(\frac{1-p}{p})}{g_1\delta} - g_3\delta/2\right)$$

$$P^c = pP_1^c + (1-p)P_2^c$$

where

$$\begin{aligned} g_1 = \{&[(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{11}+(\alpha c_{12}-\beta c_{22})a^{12}-\beta c_{23}a^{13}]^2 \\ &+[(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{21}+(\alpha c_{12}-\beta c_{22})a^{22}-\beta c_{23}a^{23}]^2 \\ &+[(\alpha c_{11}-(1-\beta)-\beta c_{21})a^{31}+(\alpha c_{12}-\beta c_{22})a^{32}-\beta c_{23}a^{33}]^2\}^{1/2} \end{aligned}$$

$$\begin{aligned} g_2 = -\frac{1}{g_1}\{&[\alpha c_{11}-(1-\beta)-\beta c_{21}][a^{11}(\alpha c_{11}+(1-\beta)+\beta c_{21}) \\ &+a^{21}(\alpha c_{12}+\beta c_{22})+a^{31}\beta c_{23}] \end{aligned}$$

$$+[\alpha c_{12}-\beta c_{22}][a^{12}(\alpha c_{11}+(1-\beta)+\beta c_{21})+a^{22}(\alpha c_{12}+\beta c_{22})+a^{32}\beta c_{23}]$$

$$-\beta c_{23}[a^{13}(\alpha c_{11}+(1-\beta)+\beta c_{21})+a^{23}(\alpha c_{12}+\beta c_{22})+a^{33}\beta c_{23}]\}$$

$$g_3 = -g_2-2d_1/g_1\delta$$

$$= -g_2 - \frac{2}{g_1}\{[\alpha c_{11}-(1-\beta)-\beta c_{21}]a^{11}+[\alpha c_{12}-\beta c_{22}]a^{12}-\beta c_{23}a^{13}\}.$$

The special cases of pure scale and pure location contamination follow easily from Theorem 2.4 by setting $\gamma_{11} = \gamma_{12} = 0$, $\gamma_{21} = \delta$, $\gamma_{22} = \gamma_{23} = 0$ for pure scale contamination and $k_1 = k_2 = 1$ for pure location contamination. Because of their special interest, these results are stated explicitly in Corollaries 2.4.1 and 2.4.2. Two other special cases are presented in Corollaries 2.4.1.1 and 2.4.2.1. The first, the pure scale contamination problem with equal prior probabilities ($p = \frac{1}{2}$) is of particular interest in this study. Chapters III, IV and V are devoted entirely to small sample results for this particular problem. Corollary 2.4.1.1 indicates that in this case, with large samples, contamination has no effect on our ability to discriminate between π_1 and π_2, i.e., the asymptotic probability of misclassification based on contaminated samples is equal to the optimal probability of misclassification. Corollary 2.4.2.1 is a generalization of a result already known in the literature: the asymptotic probability of misclassification when the initial samples are misclassified at random [20]. Lachenbruch gives an expression for this probability when $p = \frac{1}{2}$. The random initial misclassification problem is a special case of the location contamination problem, obtained by taking

$\underline{\gamma}_1 = (\delta,0,\ldots,0)$ and $\underline{\gamma}_2 = \underline{0}$, i.e., $\underline{\nu}_1 = \underline{\mu}_2$, $\underline{\nu}_2 = \underline{\mu}_1$ in the original model.

Corollary 2.4.1

Under pure scale contamination:

$$\pi_1\colon \quad N_m(\underline{0},I)$$

$$\pi_2\colon \quad N_m((\delta,0,\ldots,0),I)$$

$$\pi_1^c\colon \quad (1-\alpha)N_m(\underline{0},I)+\alpha N_m(\underline{0},k_1 I)$$

$$\pi_2^c\colon \quad (1-\beta)N_m((\delta,0,\ldots,0),I)+\beta N_m((\delta,0,\ldots,0),k_2 I)$$

$$0 \le \alpha \le 1 \qquad 0 \le \beta \le 1 \qquad k_1 \ge 1 \qquad k_2 \ge 1 \qquad \delta > 0$$

the asymptotic probabilities of misclassification using Fisher's linear discriminant function to discriminate between π_1 and π_2 with contaminated initial samples from π_1^c and π_2^c are given by

$$P_1^c = \Phi\left(\frac{\ln(\frac{1-p}{p})}{g\delta} - \delta/2\right)$$

$$P_2^c = \Phi\left(\frac{-\ln(\frac{1-p}{p})}{g\delta} - \delta/2\right)$$

$$P^c = pP_1^c+(1-p)P_2^c$$

where $g = \dfrac{n_1+n_2-2}{n_1+n_2-2+\alpha(k_1-1)(n_1-1)+\beta(k_2-1)(n_2-1)}$.

Corollary 2.4.1.1

Under pure scale contamination with equal prior probabilities for π_1 and π_2, the asymptotic probabilities of misclassification based on contaminated initial samples are equal to the optimal probabilities of misclassification, i.e., $P_1^c = P_1$, $P_2^c = P_2$, $P^c = P$.

PROOF:

If $p = \frac{1}{2}$, then $\ell n(\frac{1-p}{p}) = \ell n\ 1 = 0$.

Thus, $P_1 = P_2 = P_1^c = P_2^c = \Phi(-\delta/2)$.

<u>Corollary 2.4.2</u>

Under pure location contamination:

$$\pi_1:\quad N_m(\underline{0}, I)$$

$$\pi_2:\quad N_m((\delta, 0, \ldots, 0), I)$$

$$\pi_1^c:\quad (1-\alpha)N_m(\underline{0}, I) + \alpha N_m(\underline{\gamma}_1, I)$$

$$\pi_2^c:\quad (1-\beta)N_m((\delta, 0, \ldots, 0), I) + \beta N_m(\underline{\gamma}_2, I)$$

$$\underline{\gamma}_1 = (c_{11}\delta, c_{12}\delta, 0, 0, \ldots, 0)$$

$$\underline{\gamma}_2 = (c_{21}\delta, c_{22}\delta, c_{23}\delta, 0, \ldots, 0)$$

$$0 \leq \alpha \leq 1 \qquad 0 \leq \beta \leq 1 \qquad \delta > 0$$

the asymptotic probabilities of misclassification using Fisher's linear discriminant function to discriminate between π_1 and π_2 with contaminated initial samples from π_1^c and π_2^c are given by

$$P_1^c = \Phi\left(\frac{\ell n(\frac{1-p}{p})}{g_1\delta} - g_2\delta/2\right)$$

$$P_2^c = \Phi\left(\frac{-\ell n(\frac{1-p}{p})}{g_1\delta} - g_3\delta/2\right)$$

$$P^c = pP_1^c + (1-p)P_2^c$$

where g_1, g_2, g_3, a^{ij}, a_1, a_2, and $|A_{11}|$ are as given in Theorem 2.4. The constant a_0 is equal to 1.

Corollary 2.4.2.1

Under the random initial misclassification model, i.e., the pure location contamination model with $\underline{\gamma}_1 = (\delta,0,\ldots,0)$ and $\underline{\gamma}_2 = \underline{0}$, and $\alpha + \beta \leq 1$, the asymptotic probabilities of misclassification are given by

$$P_1^c = \Phi\left[\frac{\ln(\frac{1-p}{p})}{g_1\delta} - g_2\delta/2\right]$$

$$P_2^c = \Phi\left[\frac{-\ln(\frac{1-p}{p})}{g_1\delta} - g_3\delta/2\right]$$

$$P^c = pP_1^c+(1-p)P_2^c$$

where

$$g_1 = (1-\alpha-\beta)/(1+a_1\delta^2+a_2\delta^2)$$

$$g_2 = 1+\alpha-\beta$$

$$g_3 = 1+\beta-\alpha$$

$$a_1 = \frac{(n_1-1)\alpha(1-\alpha)}{n_1+n_2-2}$$

$$a_2 = \frac{(n_2-1)\beta(1-\beta)}{n_1+n_2-2}$$

$$\alpha + \beta \leq 1.$$

Lachenbruch points out the necessity for taking $\alpha + \beta \leq 1$. Otherwise, we could end up classifying an individual with other members of his own population, but calling them members of the other population.

2.5. The Asymptotic Effect of Contamination in the Initial Samples: Scale Contamination

Corollary 2.4.1 gives the asymptotic probabilities of misclassification when Fisher's linear discriminant function is used to discriminate between π_1: $N_m(\underline{0},I)$ and π_2: $N_m((\delta,0,\ldots,0),I)$, based on scale-contaminated initial samples. In order to evaluate the effect of scale contamination in the initial samples, we compare these probabilities to the optimal probabilities of misclassification based on uncontaminated samples, i.e., we compare P_1^c, P_2^c, and P^c to P_1, P_2, and P.

The effect of scale contamination depends upon the parameters p, δ, α, β, k_1 and k_2, and on the relationship between the two sample sizes, n_1 and n_2. The two sets of probabilities were calculated for the following combinations of parameters and sample sizes:

$$0 \le p \le 1$$

$$\delta^2 = 1,2,3,4,5,6,7,8,9$$

$$(\alpha,\beta) = (0.10,0),(0.25,0),(0.10,0.10),(0.10,0.25),(0.25,0.25)$$

$$k_1 = 4,9,25,100$$

$$k_2 = 4,9,25,100$$

$$(n_1 = n_2),\ (n_1 = np,\ n_2 = n(1-p))$$

Figures 2.5.1 through 2.5.24 illustrate the effect of scale contamination for $\delta = 1,2,3$; $(\alpha,\beta) = (0.10,0)$, $(0.25,0)$; $k_1 = 4,9,25,100$; $n_1 = n_2$. The probability of misclassification is plotted as a function of p, the prior probability of population 1. Figures 2.5.1 through 2.5.12 illustrate the effect on P, the overall probability of misclassification. The three graphs in each figure (from top to bottom) correspond to $(\alpha,\beta) = (0.25,0)$; $(\alpha,\beta) = (0.10,0)$; and $(\alpha,\beta) = (0,0)$, the optimal probability of misclassification. Figures 2.5.13 through 2.5.24 illustrate the effect of scale-contaminated initial samples on P_1, the probability of misclassification in population 1. The three graphs in each of these figures (from top to bottom at $p < 0.5$) correspond to $(\alpha,\beta) = (0.25,0)$; $(\alpha,\beta) = (0.10,0)$; $(\alpha,\beta) = (0.0)$. Since $P_2^c(p) = P_1^c(1-p)$, i.e., the probability of misclassification in population 2 when the prior probability of population 1 is p is equal to the probability of misclassification in population 1 when the prior probability of population 1 is (1-p), the graphs of P_2^c as a function of p are reflections of the corresponding graphs for P_1^c, reflected about $p = 0.5$.

Consider first the effect of scale contamination on the individual probabilities, P_1 and P_2. We have $P_1^c > P_1$ for $p < \frac{1}{2}$; $P_1^c < P_1$ for $p > \frac{1}{2}$; $P_1^c = P_1$ for $p = \frac{1}{2}$. The relationship is reversed for P_2: $P_2^c < P_2$ for $p < \frac{1}{2}$; $P_2^c > P_2$ for $p > \frac{1}{2}$; $P_2^c = P_2$ for $p = \frac{1}{2}$. Thus, contamination hurts us in classifying individuals who belong to the population having the smaller prior probability, while it improves our ability to classify individuals from the population with larger prior probability. Note, however, that the detrimental effect in the less likely population is greater than the beneficial effect in the population with larger prior probability. The above relationship holds, no matter which of the two samples is contaminated, i.e., $(\alpha,\beta) = (0.10,0)$ gives the same probabilities of misclassification as $(\alpha,\beta) = (0,0.10)$. In general, scale contamination has a significant effect on the individual probabilities of misclassification, even when the contamination is mild.

The effect of scale-contaminated initial samples on the overall probability of misclassification is illustrated in Figures 2.5.1 through 2.5.12. The graph of P^c as a function of p is symmetric about $p = 0.5$; therefore only the left half of the graph is illustrated.

The overall probability of misclassification is less affected by scale contamination than the individual probabilities. For mild contamination ($k = 4,9$), there is little effect. (When $k = 9$ and $\alpha = 0.25$, we see some mild effects.) When we reach $k = 25$, however, contamination begins to hurt our ability to discriminate between π_1 and π_2 using Fisher's LDF with 10% contamination. For $k = 100$, contamination has a definite harmful effect.

All of the above results are based on single sample scale contamination with $n_1 = n_2$. Similar calculations were made for $n_1 = np$, $n_2 = n(1-p)$, i.e., sample sizes proportional to the prior probabilities of the two populations. In this case, the graph of P^c as a function of p is not symmetric about $p = 0.5$. The contaminated probability of misclassification is closer to the optimal probability when the prior probability of the contaminated population is less than 0.5. Figure 2.5.25 illustrates the effect of scale contamination for $\delta = 2$, $k = 100$, and $n_1 = np$, $n_2 = n(1-p)$. This figure corresponds to Figure 2.5.8 with equal sample sizes. The general conclusions regarding the performance of the LDF with contaminated initial samples are the same, whether we take equal sample sizes or sample sizes proportional to the prior probabilities of the two populations.

Similar conclusions are drawn when both samples are contaminated. Figures 2.5.26 and 2.5.27 illustrate the effect of two-sample scale contamination for $n_1 = n_2$, ($\delta = 3$, $k_1 = k_2 = 9$) and ($\delta = 3$, $k_1 = k_2 = 100$). The four graphs in each figure (from top to bottom) correspond to $(\alpha,\beta) = (0.25,0.25)$; $(\alpha,\beta) = (0.10,0.25)$; $(\alpha,\beta) = (0.10,0.10)$; $(\alpha,\beta) = (0,0)$. As expected, two sample scale contamination hurts our ability to make correct classifications more than single-sample contamination. Even with $k_1 = k_2 = 9$, our probability of misclassification may be significantly increased (depending upon the value of p, the prior probability of population 1).

Fig. 2.5.1. Overall probability of misclassification with single sample scale contamination: $\delta=1$; $k_1=4$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*

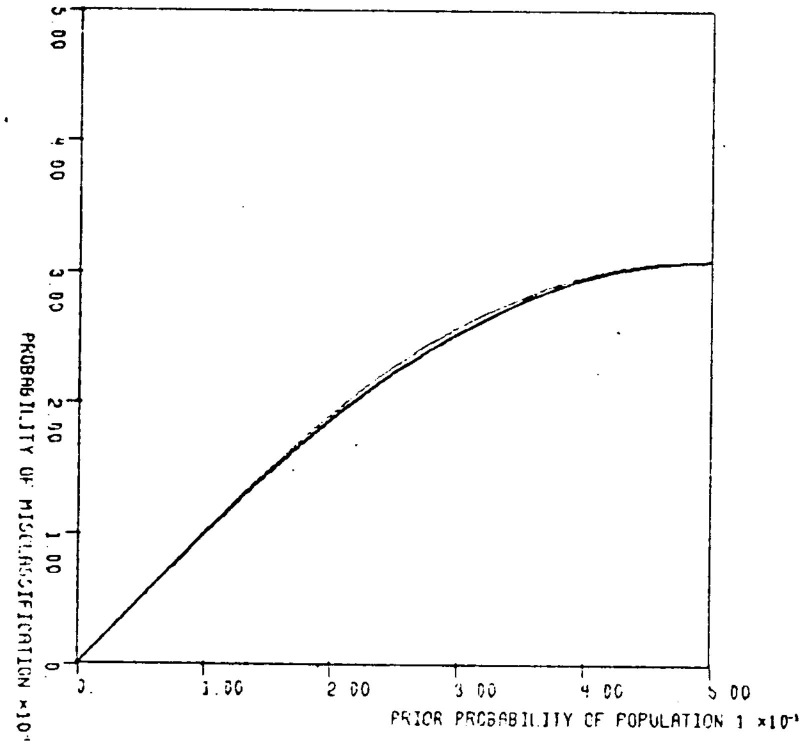
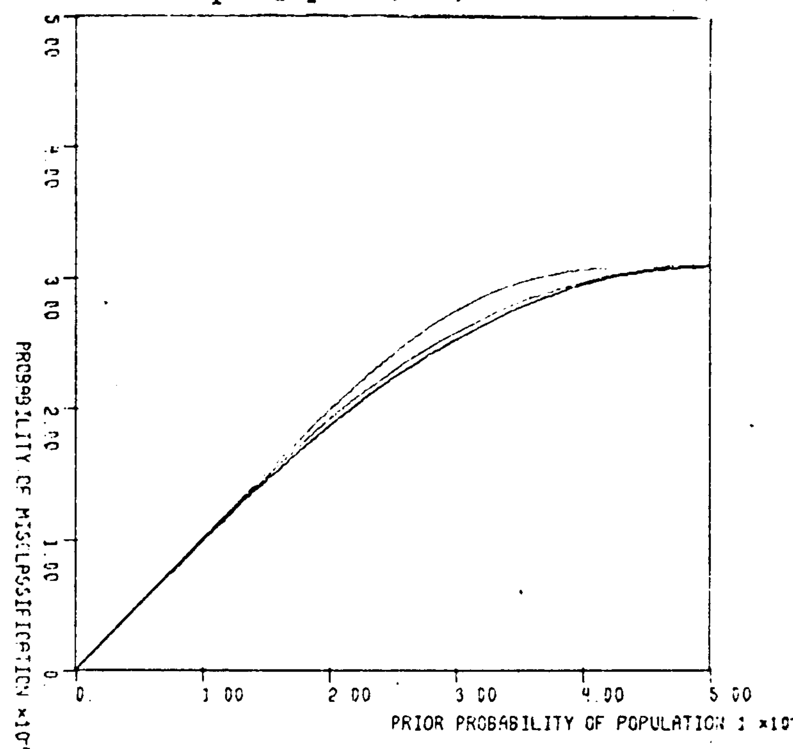


Fig. 2.5.2. Overall probability of misclassification with single sample scale contamination: $\delta=1$; $k_1=9$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.3. Overall probability of misclassification with single sample scale contamination: $\delta=1$; $k_1=25$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*

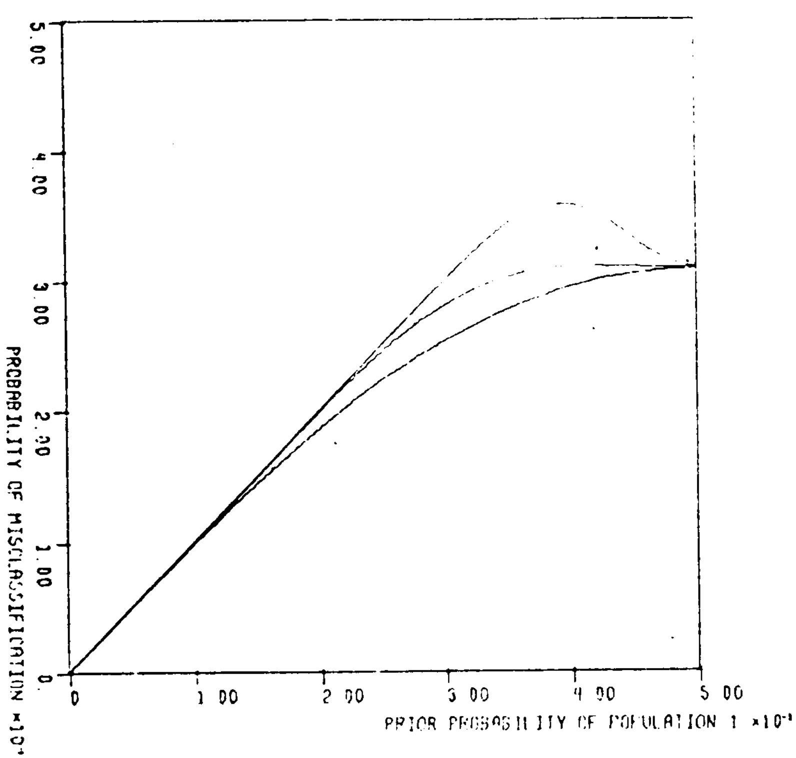
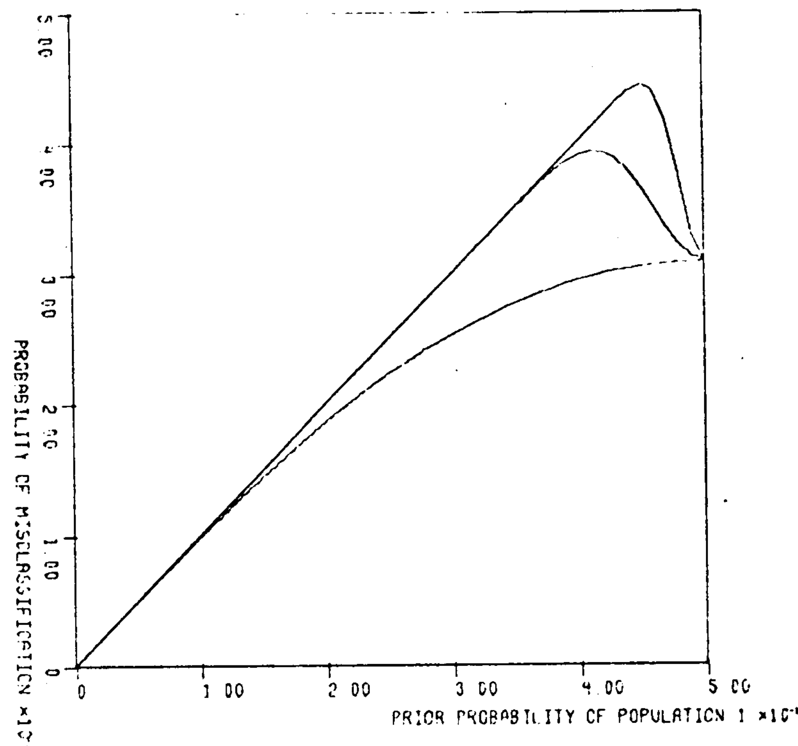


Fig. 2.5.4. Overall probability of misclassification with single sample scale contamination: $\delta=1$; $k_1=100$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.5. Overall probability of misclassification with single sample scale contamination: $\delta=2$, $k_1=4$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*

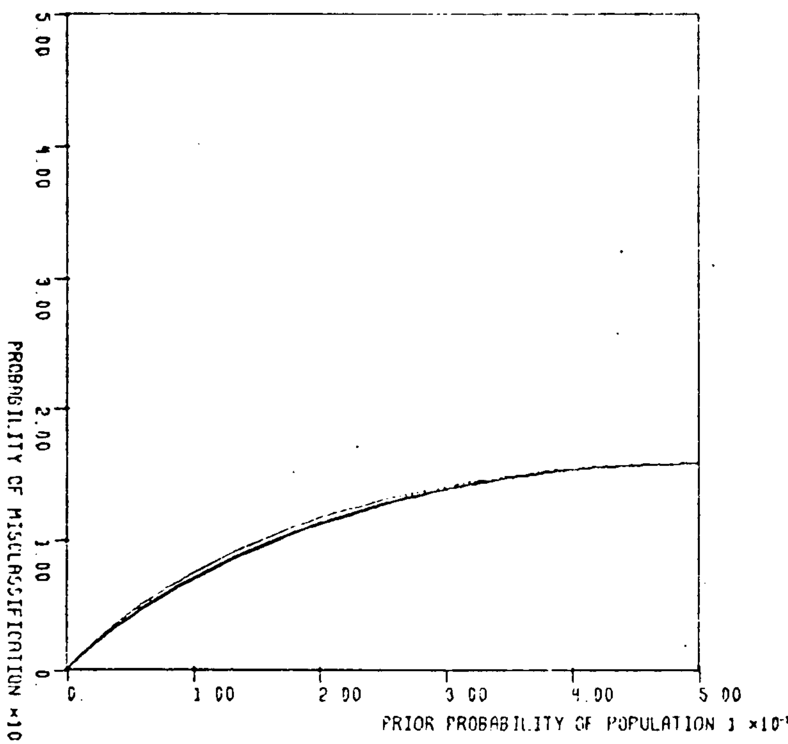
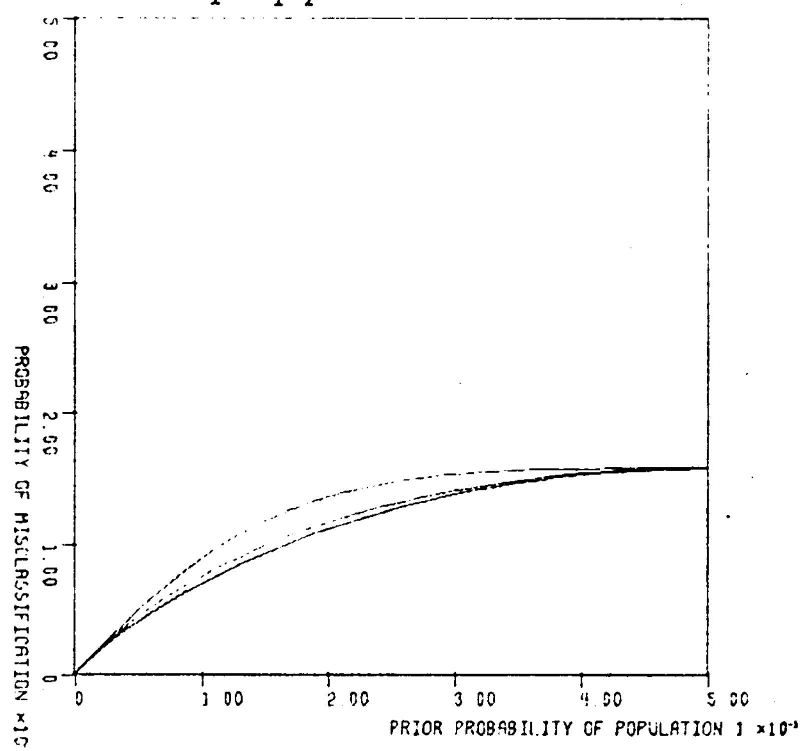


Fig. 2.5.6. Overall probability of misclassification with single sample scale contamination: $\delta=2$; $k_1=9$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.7. Overall probability of misclassification with single sample scale contamination: δ=2; k_1=25; n_1=n_2; α=0.0,0.10,0.25.*

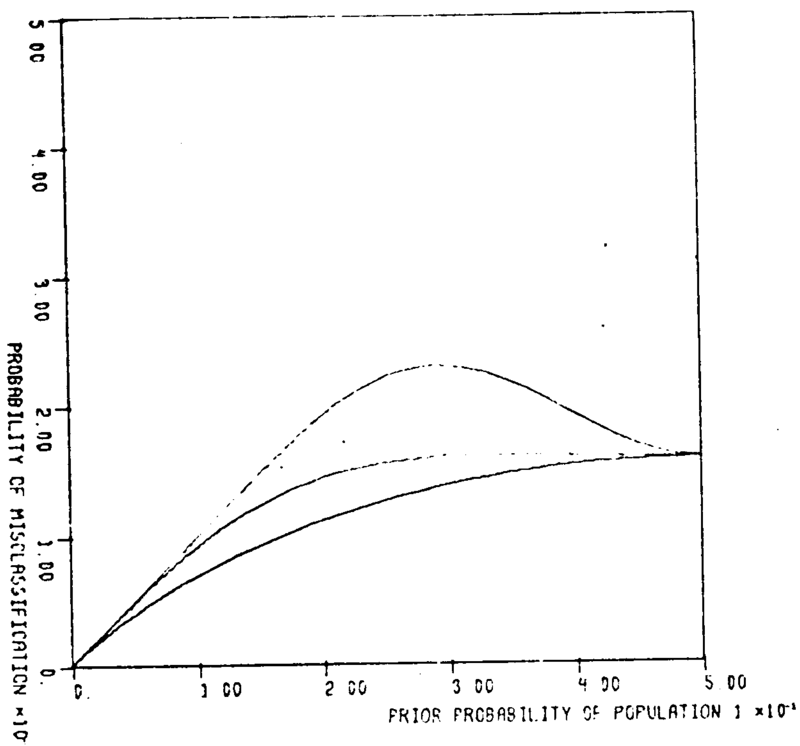
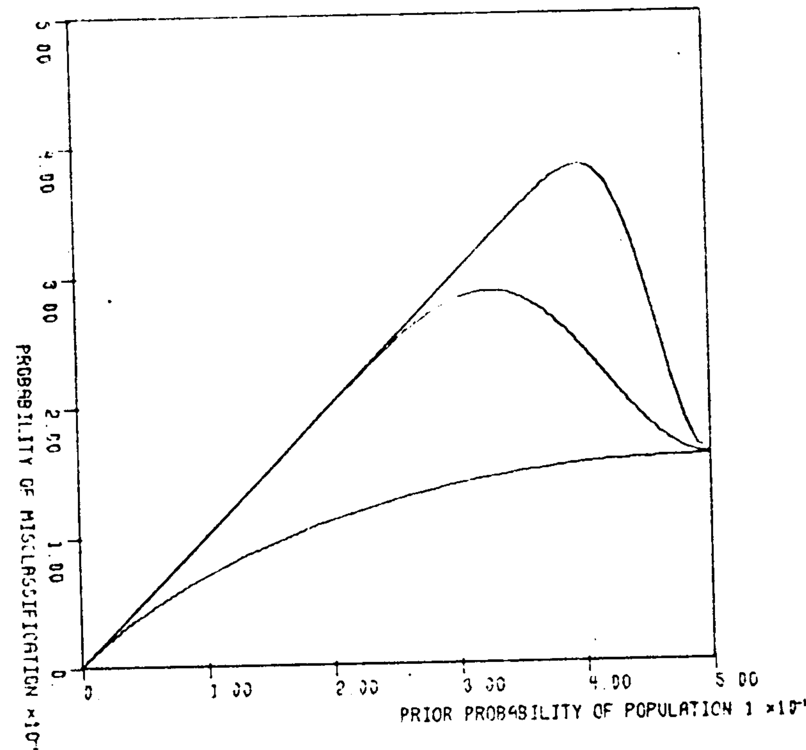


Fig. 2.5.8. Overall probability of misclassification with single sample scale contamination: δ=2, k_1=100; n_1=n_2; α=0.0,0.10,0.25.*



*The three graphs in the figure (from top to bottom) correspond to α=0.25, 0.10, 0.0.

Fig. 2.5.9. Overall probability of misclassification with single sample scale contamination: $\delta=3$; $k_1=4$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*

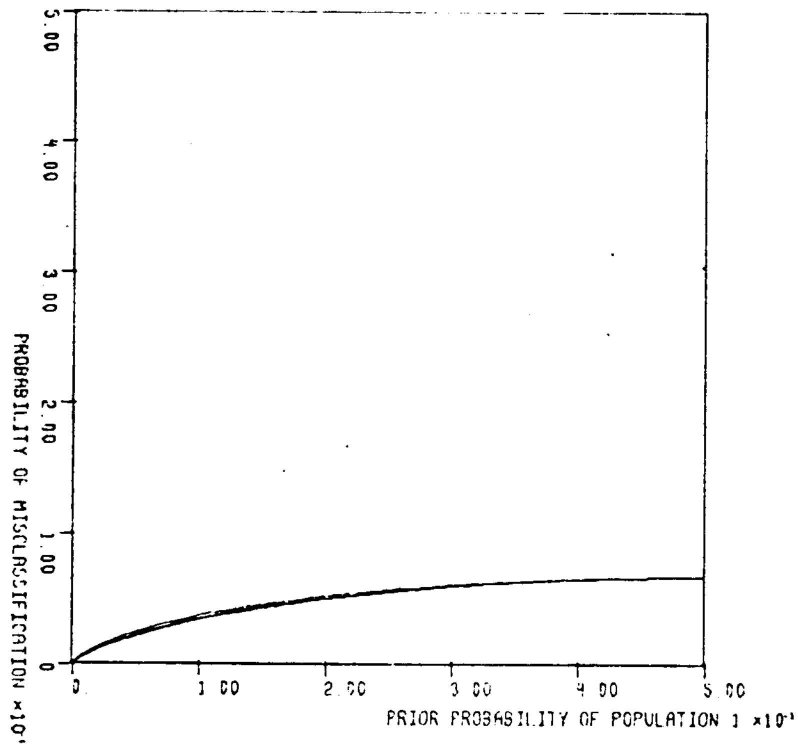
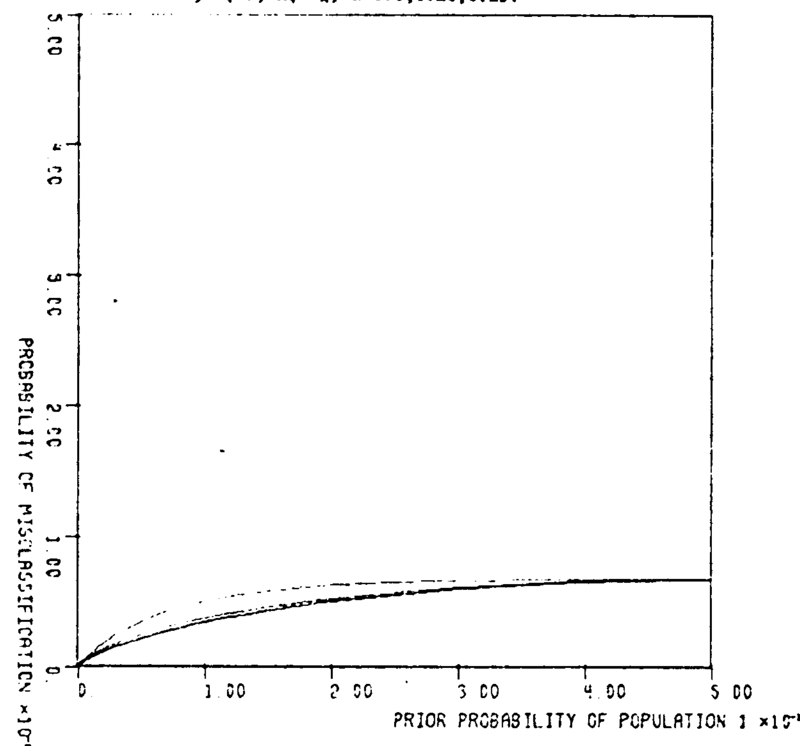


Fig. 2.5.10. Overall probability of misclassification with single sample scale contamination: $\delta=3$; $k_1=9$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25, 0.10, 0.0$.

Fig. 2.5.11. Overall probability of misclassification with single sample scale contamination: $\delta=3$; $k_1=25$; $n_1=n_2$; $\alpha=0.0,0.10,0.25.$*

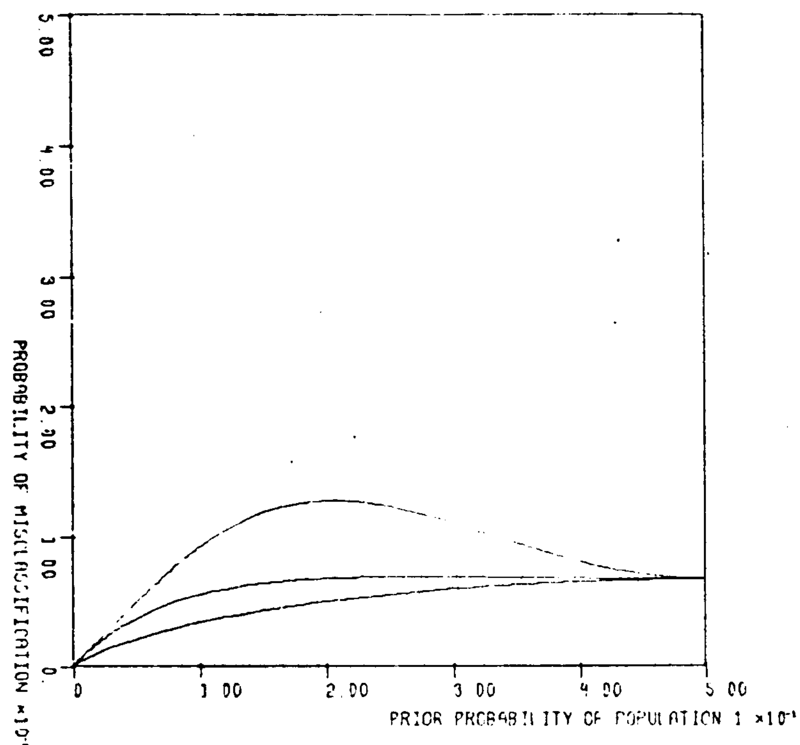
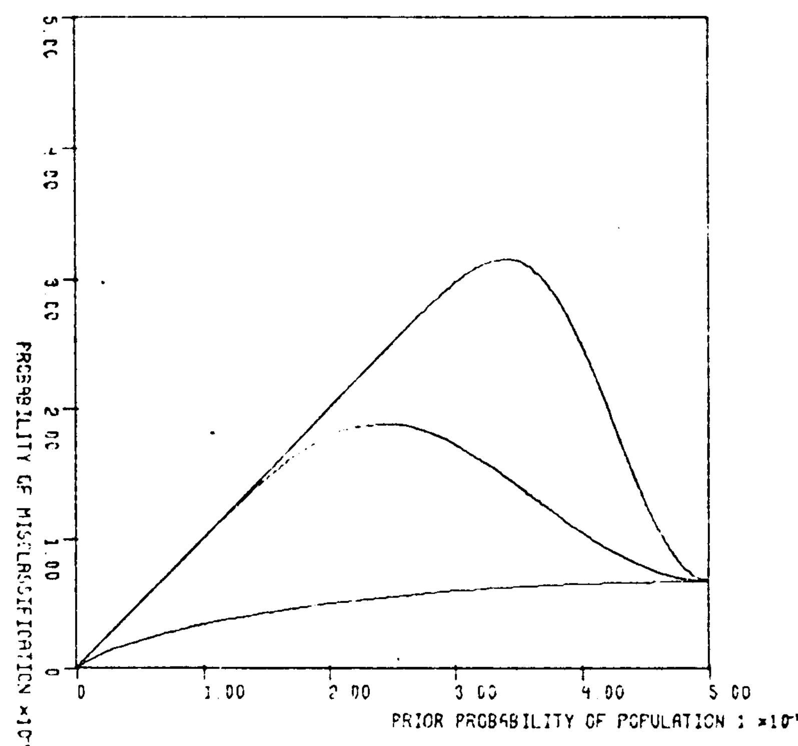


Fig. 2.5.12. Overall probability of misclassification with single sample scale contamination: $\delta=3$; $k_1=100$; $n_1=n_2$; $\alpha=0.0,0.10,0.25.$*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.13. Probability of misclassification in population 1 with single sample scale contamination: $\delta=1$; $k_1=4$; $n_1=n_2$; $\alpha=0.0,0.10,0.25$.*

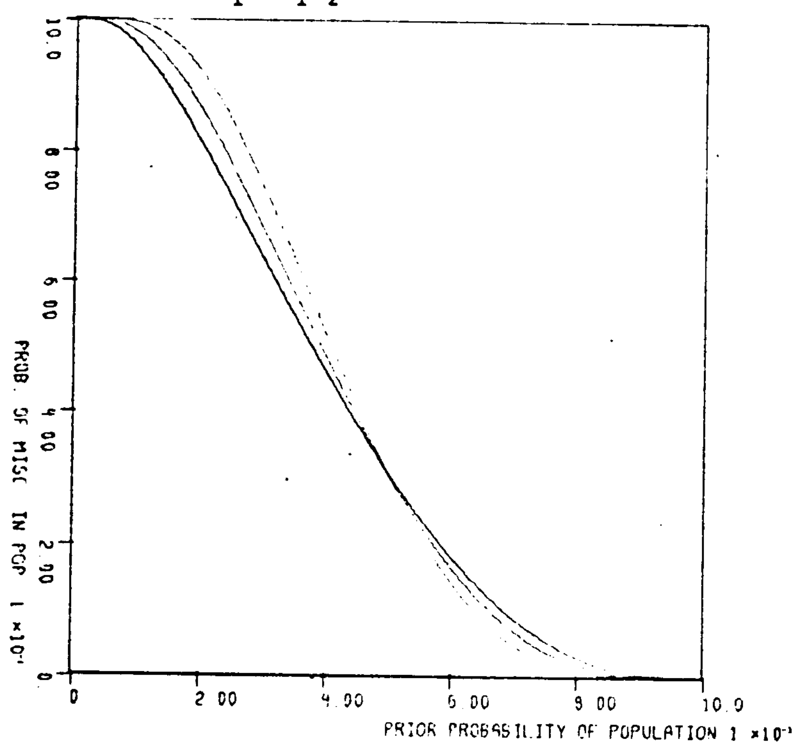
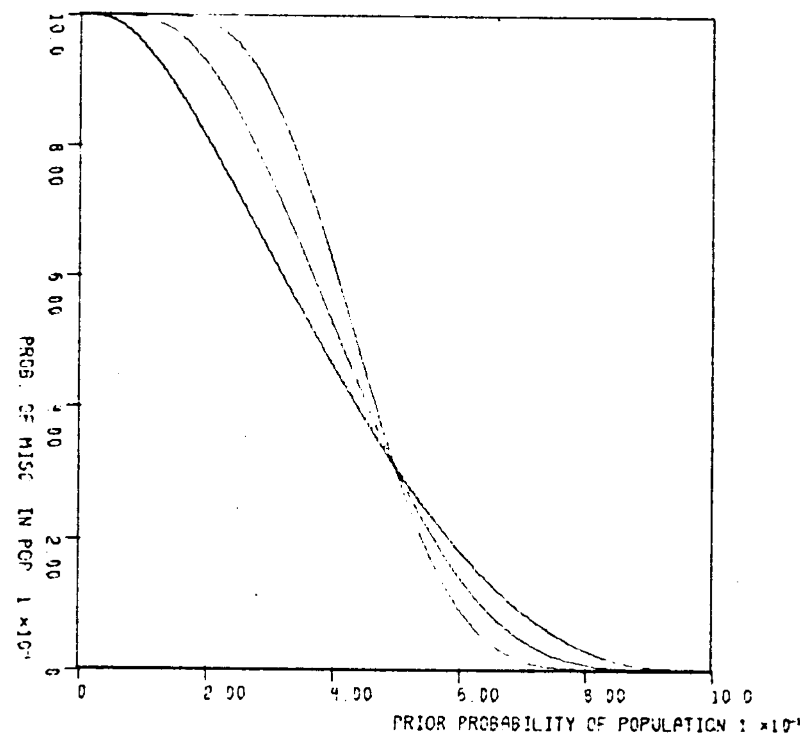


Fig.2.5.14. Probability of miscalssification in population 1 with single sample scale contamination: $\delta=1$, $k_1=9$; $n_1=n_2$; $\alpha=0.0,0.10,0.25$.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.15. Probability of misclassification in population 1 with single sample scale contamination: $\delta=1$; $k_1=25$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*

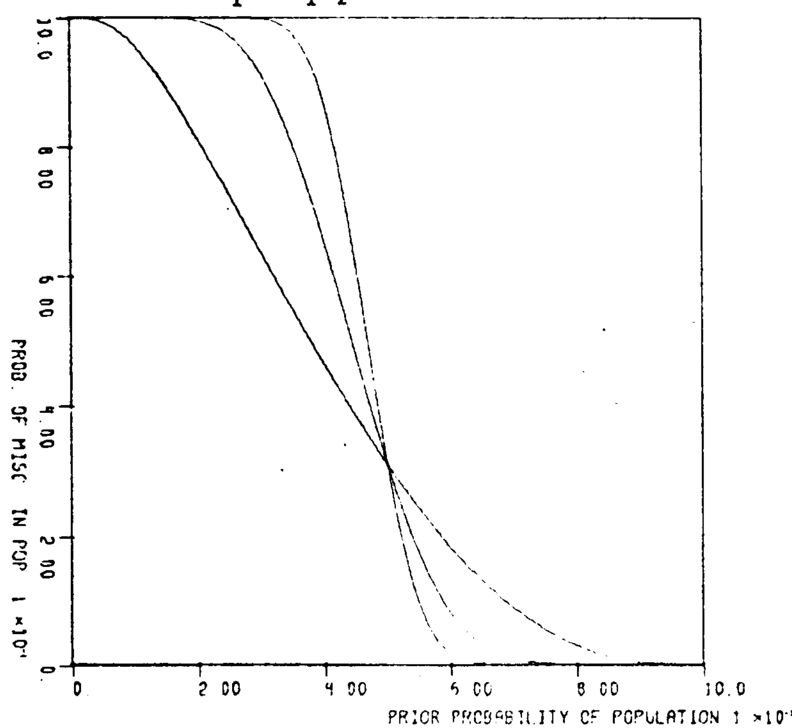
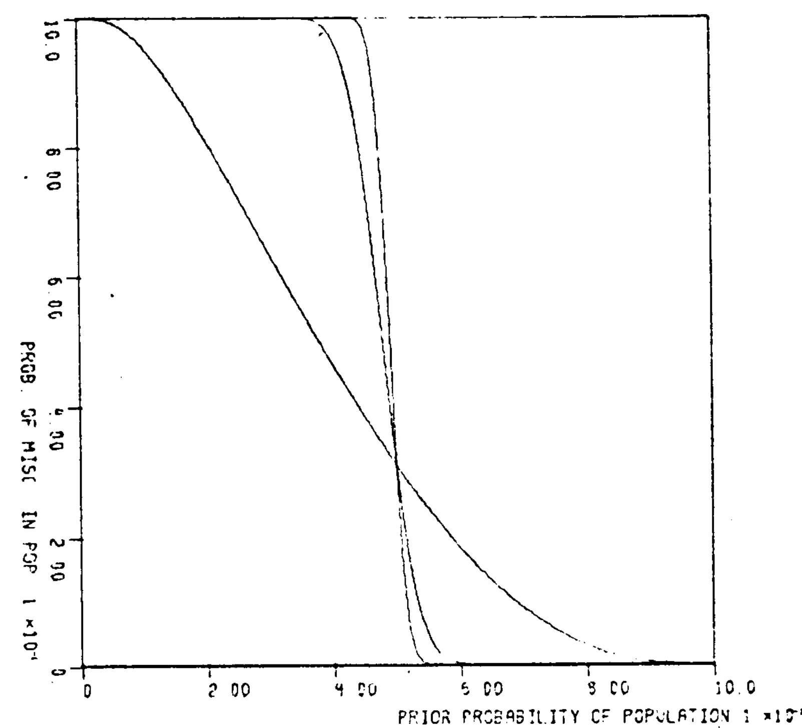


Fig. 2.5.16. Probability of misclassification in population 1 with single sample scale contamination: $\delta=1$; $k_1=100$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25, 0.10, 0.0$.

Fig. 2.5.17. Probability of misclassification in population 1 with single sample scale contamination: $\delta=2$; $k_1=4$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*

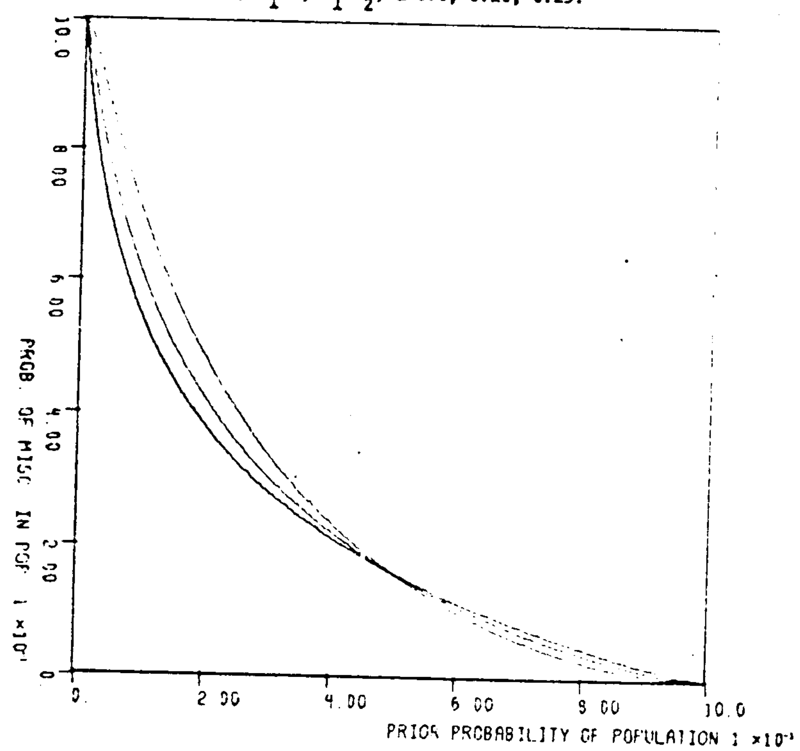
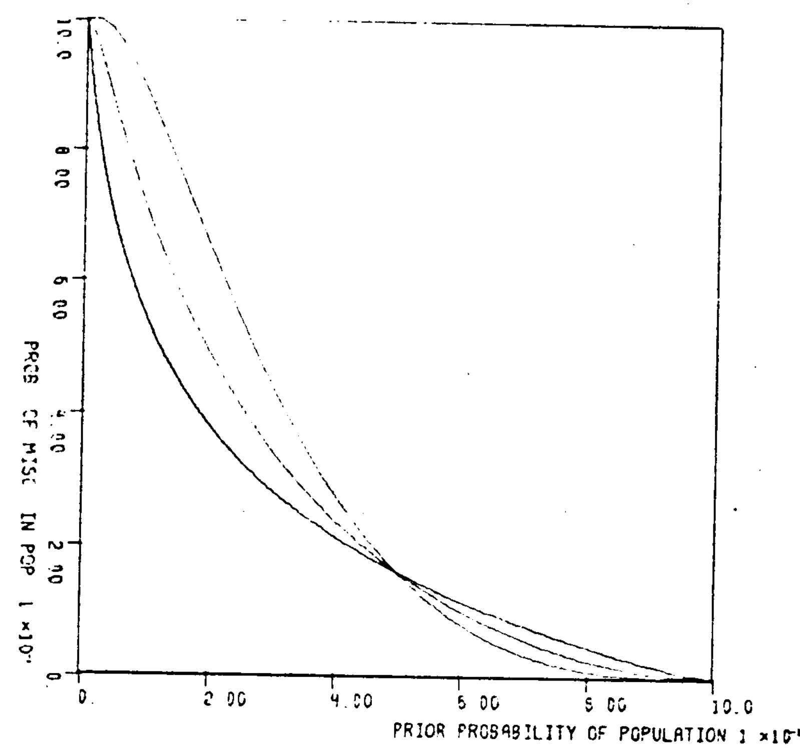


Fig. 2.5.18. Probability of misclassification in population 1 with single sample scale contamination: $\delta=2$; $k_1=9$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25, 0.10, 0.0$.

Fig. 2.5.19. Probability of misclassification in population 1 with single sample scale contamination: $\delta=2$; $k_1=25$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*

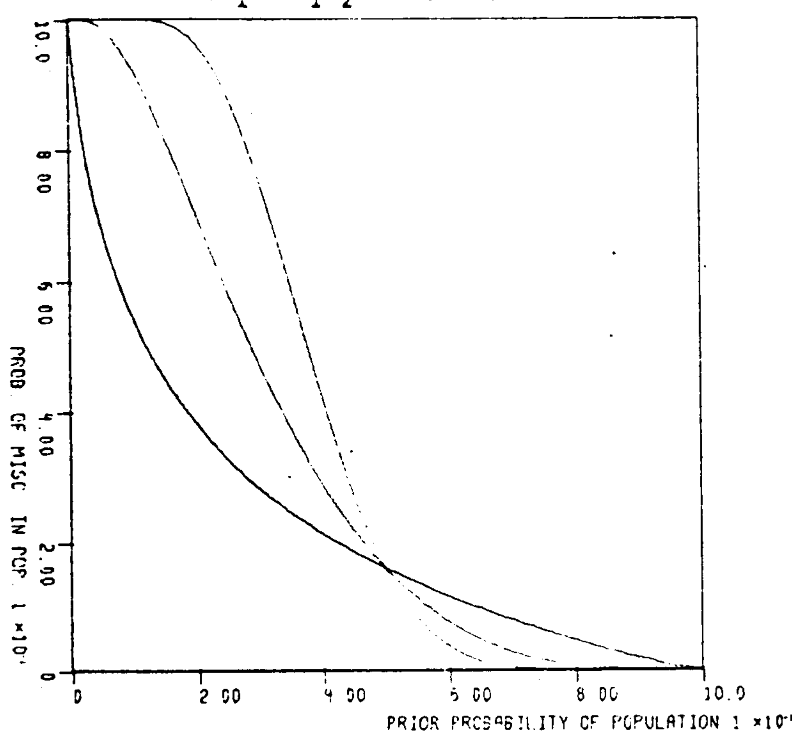
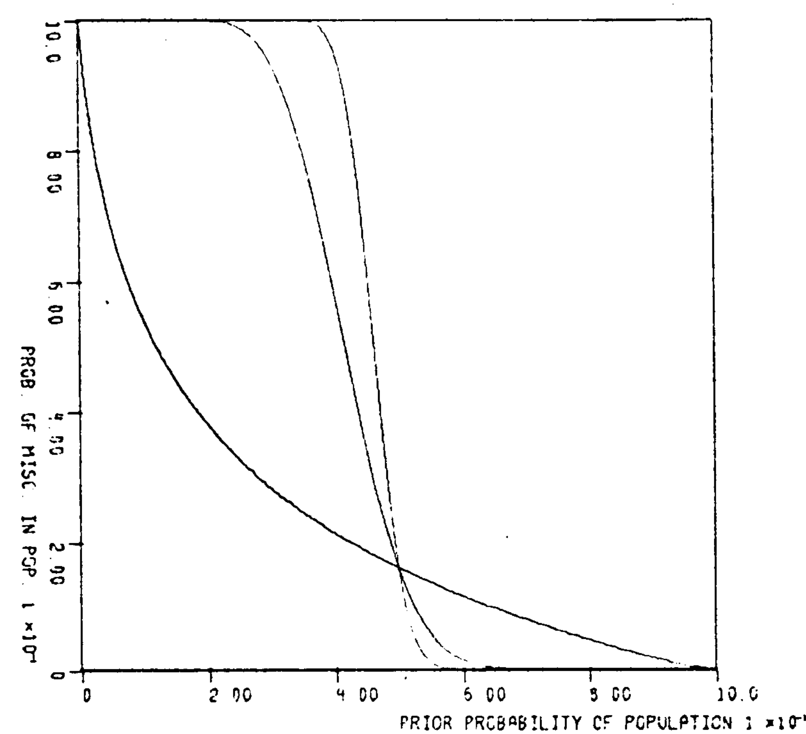


Fig. 2.5.20. Probability of misclassification in population 1 with single sample scale contamination: $\delta=2$; $k_1=100$; $n_1=n_2$; $\alpha=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25, 0.10, 0.0$.

Fig. 2.5.21. Probability of misclassification in population 1 with single sample scale contamination: $\delta=3$; $k_1=4$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*

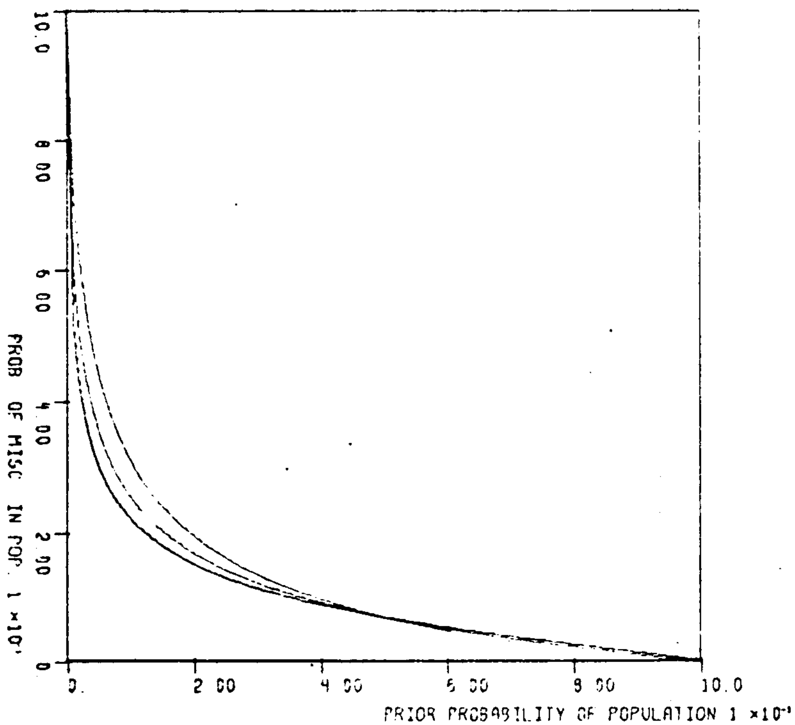
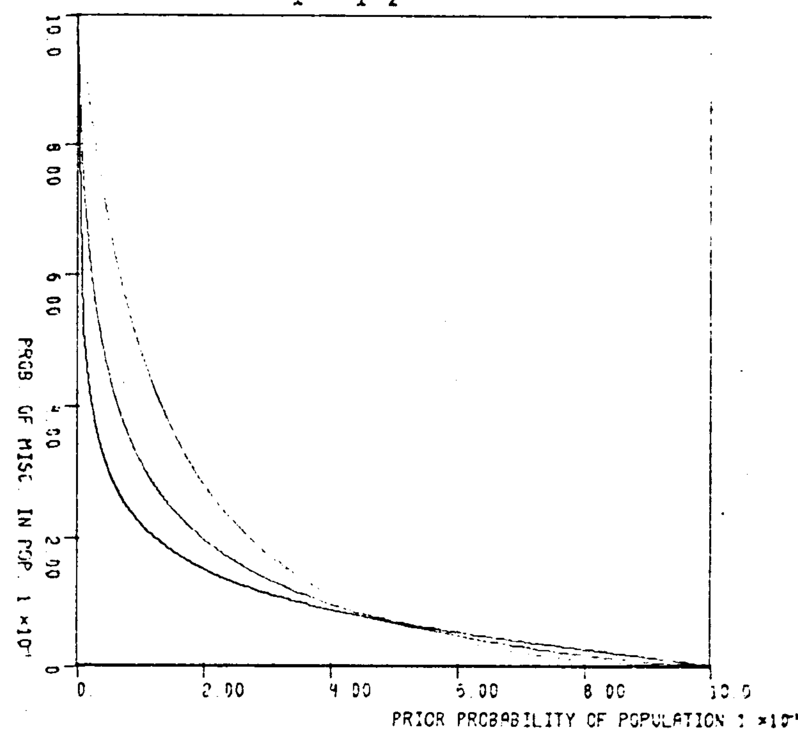


Fig. 2.5.22. Probability of misclassification in population 1 with single sample scale contamination: $\delta=3$; $k_1=9$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.23. Probability of misclassification in population 1 with single sample scale contamination: $\delta=3$; $k_1=25$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*

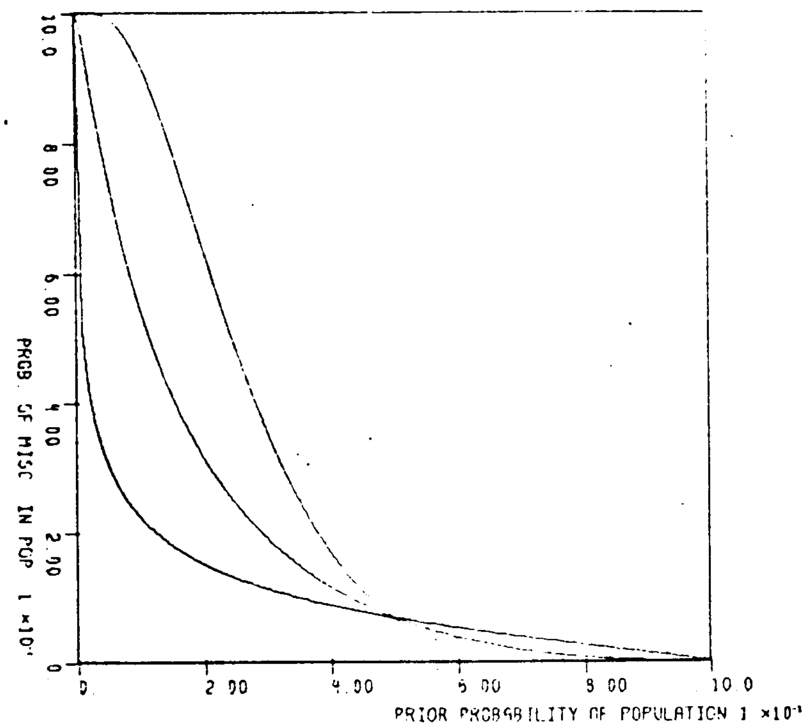
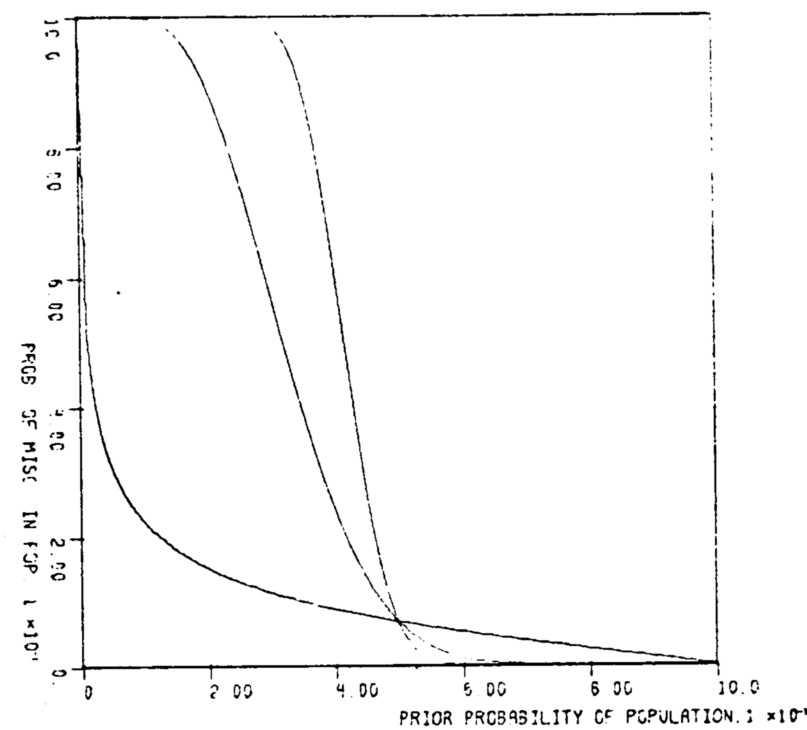
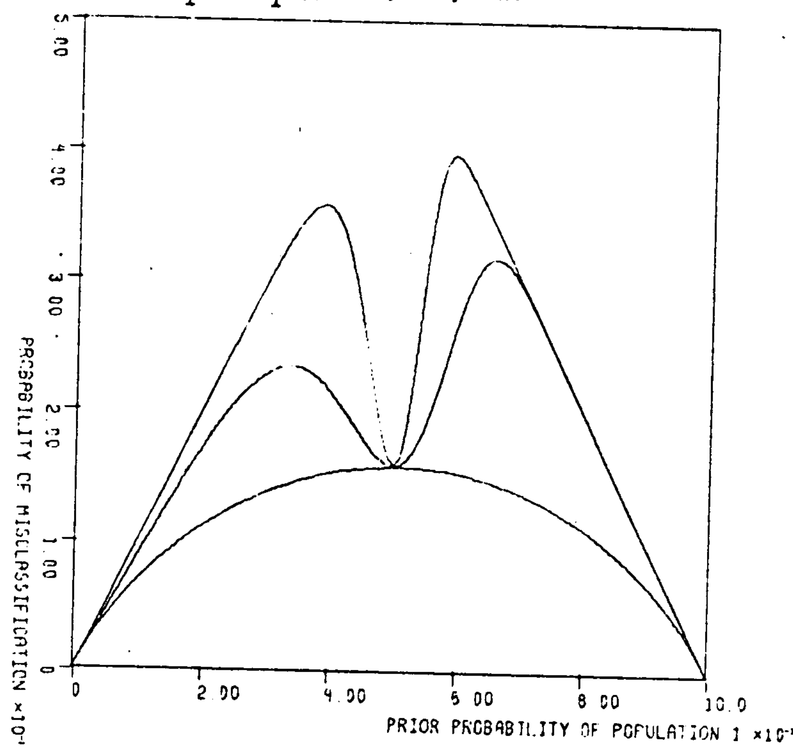


Fig. 2.5.24. Probability of misclassification in population 1 with single sample scale contamination: $\delta=3$; $k_1=100$; $n_1=n_2$; $\alpha=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.25. Overall probability of misclassification with single sample scale contamination: $\delta=2$; $k_1=100$; $n_1=np$; $\alpha=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom) correspond to $\alpha=0.25$, 0.10, 0.0.

Fig. 2.5.26. Overall probability of misclassification with two sample scale contamination: $\delta=3$; $k_1=9$, $k_2=9$; $n_1=n_2$; $(\alpha,\beta)=(.25,.25)$, $(.10,.25)$, $(.10,.10)$, $(0.0,0.0)$.*

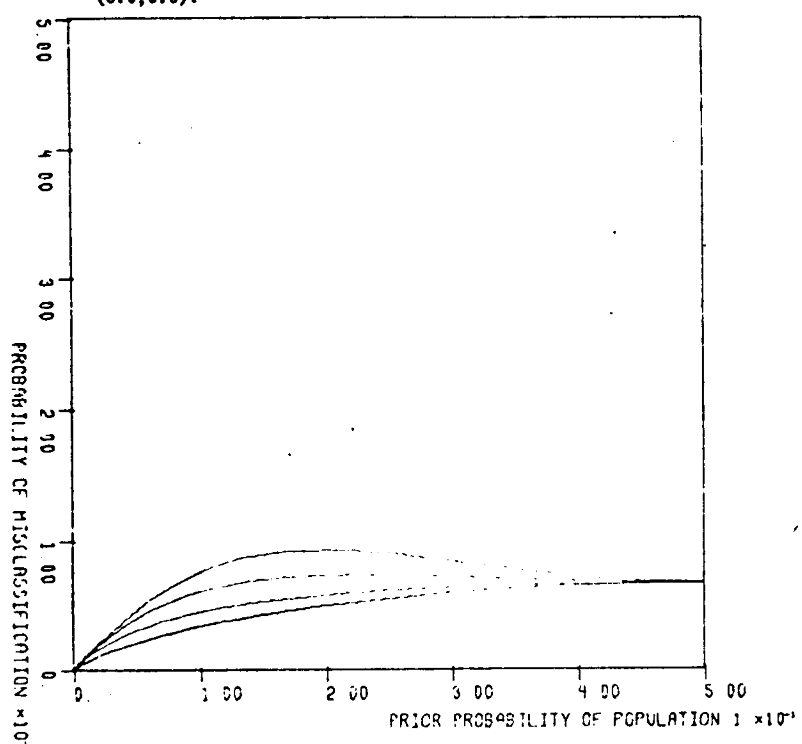
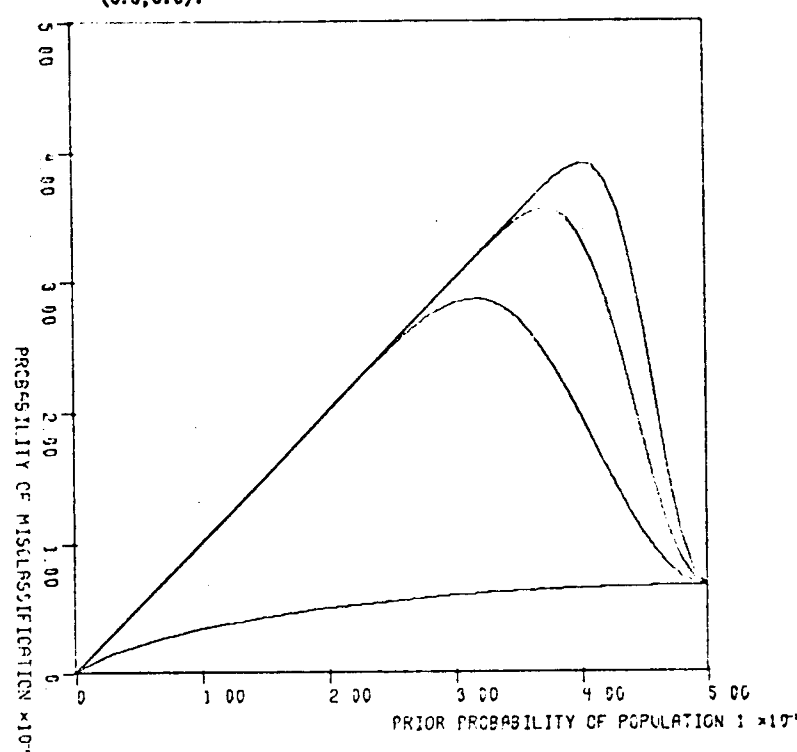


Fig. 2.5.27. Overall probability of misclassification with two sample scale contamination: $\delta=3$; $k_1=100$, $k_2=100$; $n_1=n_2$; $(\alpha,\beta)=(.25,.25)$, $(.10,.25)$, $(.10,.10)$, $(0.0,0.0)$.*



*The four graphs in the figure (from top to bottom) correspond to $(\alpha,\beta)=(.25,.25)$, $(.10,.25)$, $(.10,.10)$, $(0.0,0.0)$.

2.6. The Asymptotic Effect of Contamination in the Initial Samples: Location Contamination

Corollary 2.4.2 gives the asymptotic probabilities of misclassification when Fisher's linear discriminant function is used to discriminate between π_1: $N_m(\underline{0},I)$ and π_2: $N_m((\delta,0,\ldots,0),I)$ based on location-contaminated initial samples. In order to evaluate the effect of location contamination, P_1^c, P_2^c, and P^c can be compared to P_1, P_2, and P respectively.

The effect of location contamination depends upon the parameters p, δ, α, β, c_{11}, c_{12}, c_{21}, c_{22}, and c_{23}, and on the relationship between the two sample sizes n_1 and n_2. Since our major interest in this paper is with scale contamination and because the five location contamination parameters (c_{11}, c_{12}, c_{21}, c_{22}, c_{23}) give rise to so many different possible location contamination situations, we present only a few special cases. Figures 2.6.1 through 2.6.18 illustrate the effect of location contamination for $\delta = 1,3$; $(\alpha,\beta) = (0,0.10)$, $(0,0.25)$; $\underline{c}_2 = (-1,0,0)$, $(0.5,0,0)$,$(3.0,0,0)$; $n_1 = n_2$. Figures 2.6.1 through 2.6.6 illustrate the effect on P; figures 2.6.7 through 2.6.12 illustrate the effect on P_1; figures 2.6.13 through 2.6.18 illustrate the effect on P_2.

The overall probability of misclassification can be greatly affected by location contamination, depending upon the direction of the contamination. When the contaminating mean lies between the two population means, (e.g., $\underline{c}_2 = (0.5,0,0)$), the contamination has little effect. When $\underline{c}_2 = (3,0,0)$ (the contaminating mean is in the direction opposite to the uncontaminated population), the contamination may have some effect, depending upon the size of δ and p. For $\delta = 3$, contamination does increase the probability of misclassification for $p > 0.3$.

As expected, the greatest effect of location contamination occurs when the contaminating mean is on the opposite side of the uncontaminated population mean. When $\underline{c}_2 = (-1,0,0)$ and $\delta = 3$, location contamination greatly increases the probability of misclassification.

Fig. 2.6.1. Overall probability of misclassification with single sample location contamination: $\delta=1$; $\underline{c}_2=(-1.0, 0.0, 0.0)$; $n_1=n_2$; $\beta=0.0, 0.10, 0.25$.*

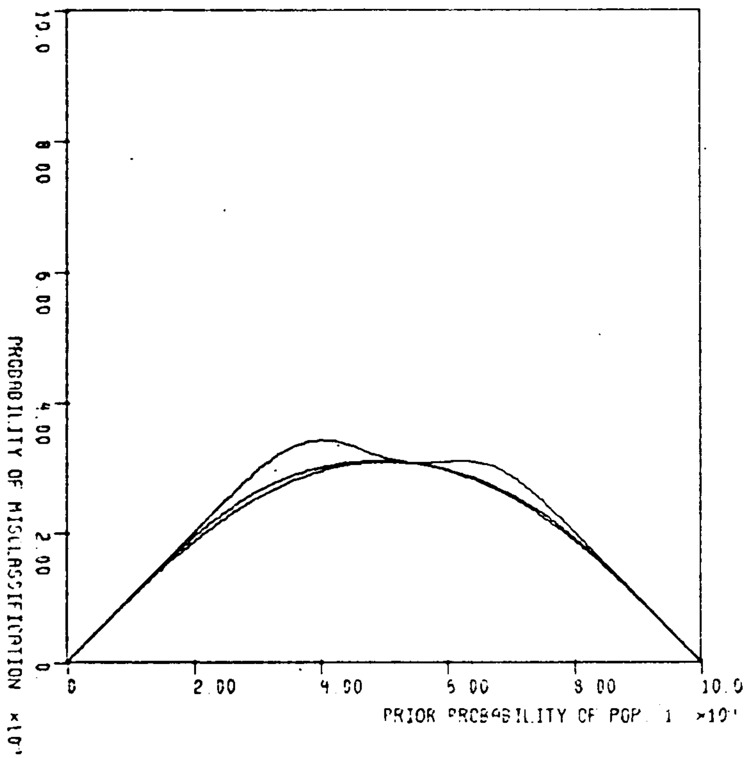
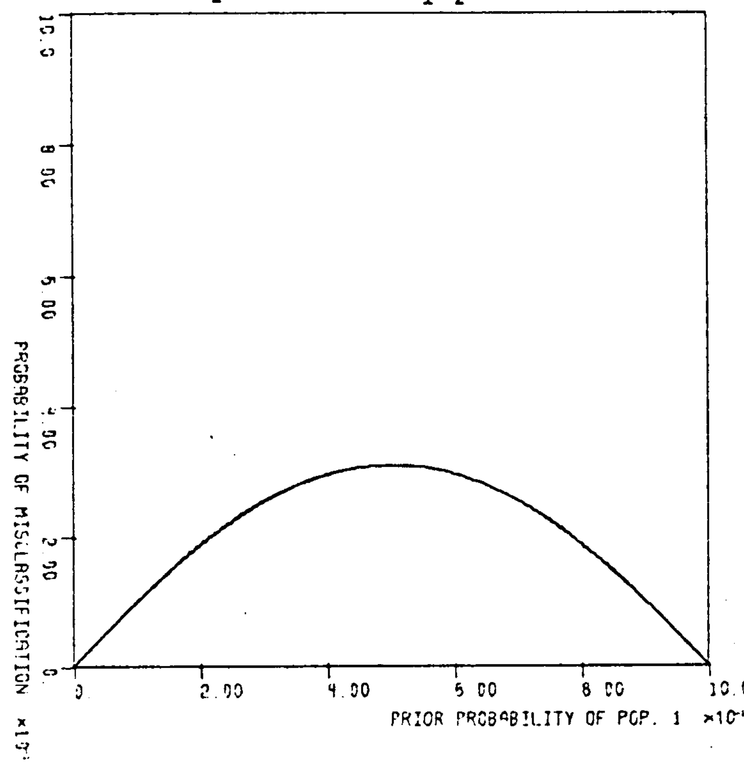


Fig. 2.6.2. Overall probability of misclassification with single sample location contamination: $\delta=1$; $c_2=(0.5, 0.0, 0.0)$; $n_1=n_2$; $\beta=0.0, 0.10, 0.25$.*



*The three graphs in the figure (from top to bottom) correspond to $\beta=0.25, 0.10, 0.0$.

Fig. 2.6.3. Overall probability of misclassification with single sample location contamination: $\delta=1$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25.$*

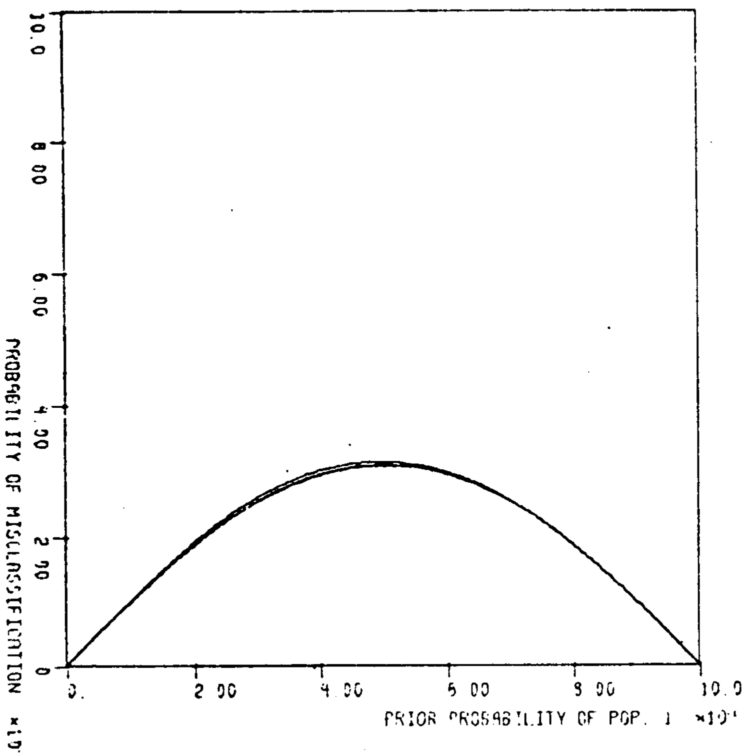
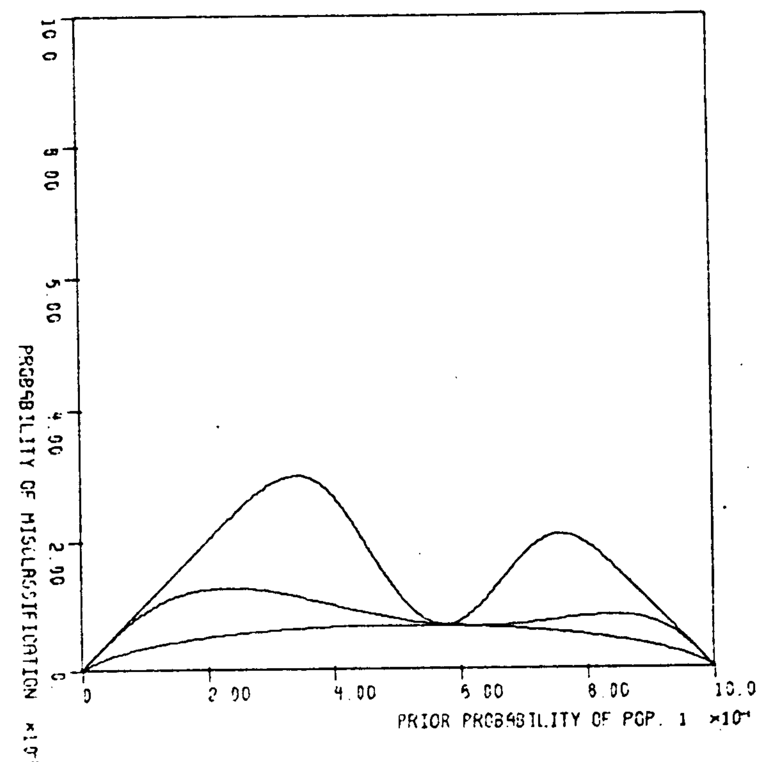


Fig. 2.6.4. Overall probability of misclassification with single sample location contamination:* $\delta=3$; $\underline{c}_2=(-1.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,\ 0.10,\ 0.25.$



*The three graphs in the figure (from top to bottom) correspond to $\beta=0.25,\ 0.10,\ 0.0.$

Fig. 2.6.5. Overall probability of misclassification with single sample location contamination: $\delta=3$; $\underline{c}_2=(0.5,0.0,0.0)$; $n_1=n_2$; $\beta=0.0$, 0.10, 0.25.*

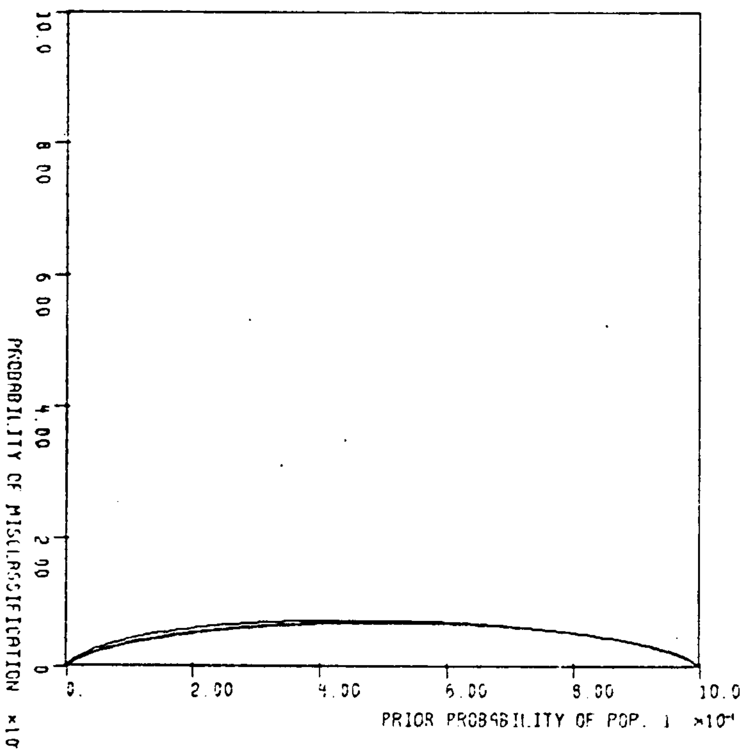
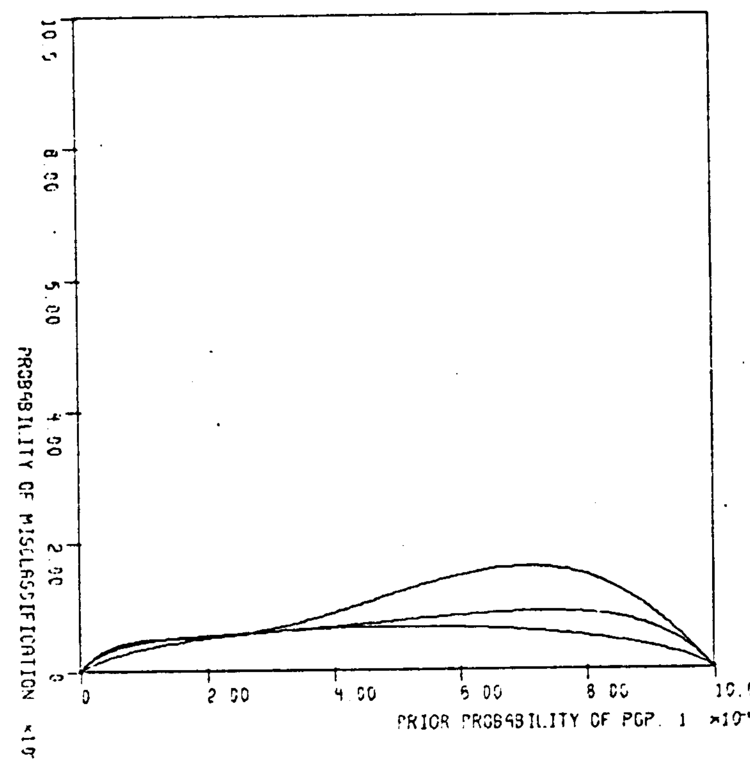


Fig. 2.6.6. Overall probability of misclassification with single sample location contamination: $\delta=3$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom) correspond to $\beta=0.25$, 0.10, 0.0.

Fig. 2.6.7. Probability of misclassification in population 1 with single sample location contamination: $\delta=1$; $\underline{c}_2=(-1.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0$, 0.10, 0.25.*

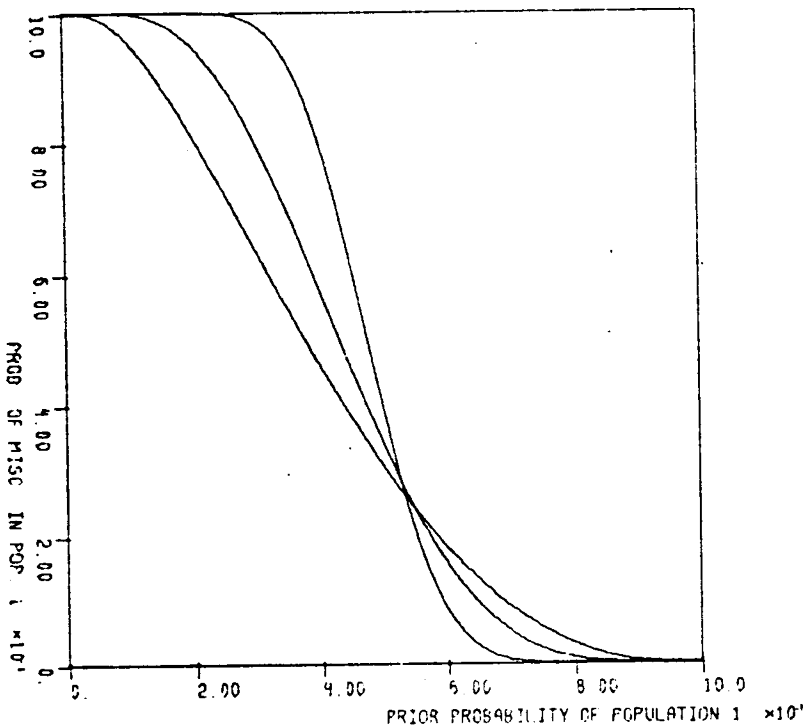
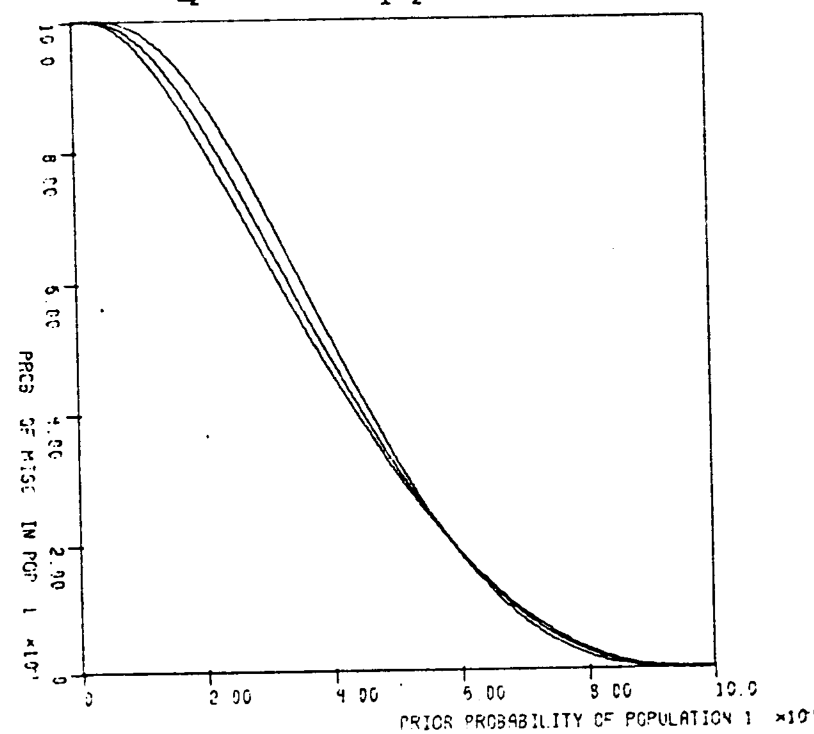


Fig. 2.6.8. Probability of misclassification in population 1 with single sample location contamination: $\delta=1$; $\underline{c}_2=(0.5,0.0,0.0)$; $n_1=n_2$; $\beta=0.0$, 0.10, 0.25.*



*The three graphs in the figure (from top to bottom at $p<.5$) correspond to $\beta=0.25$, 0.10, 0.0.

Fig. 2.6.9. Probability of misclassification in population 1 with single sample location contamination: $\delta=1$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25.$*

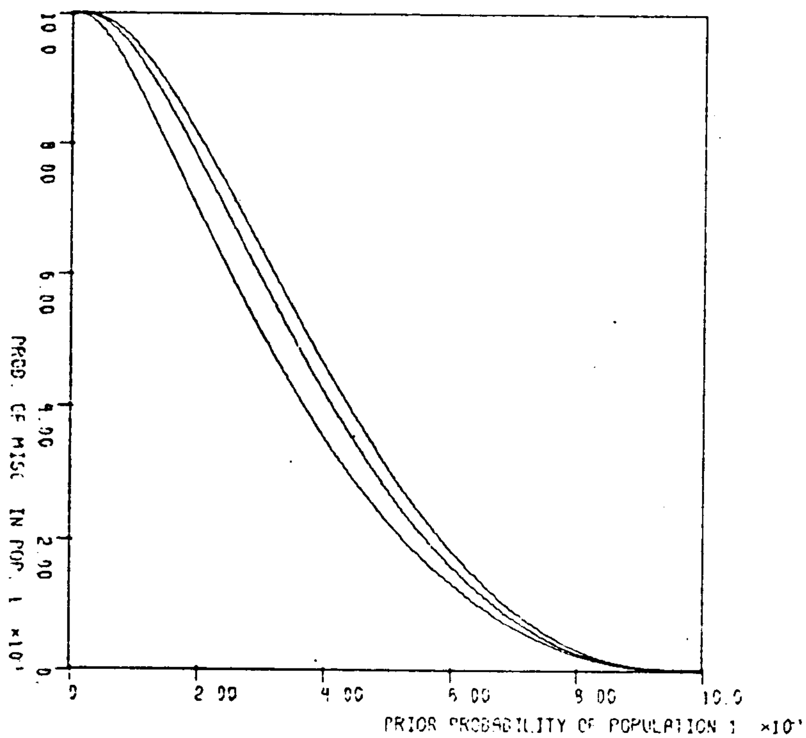
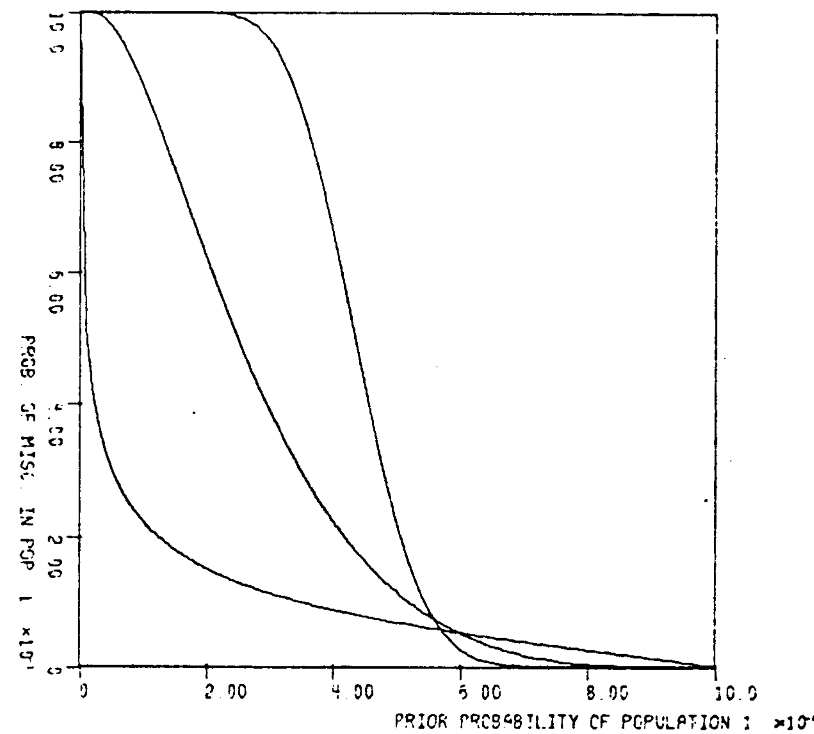


Fig. 2.6.10. Probability of misclassification in population 1 with single sample location contamination: $\delta=3$; $\underline{c}_2=(-1.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25.$**



*The three graphs in the figure (from top to bottom) correspond to $\beta=0.0$, 0.10, 0.25.

**The three graphs in the figure (from top to bottom at $p < .5$) correspond to $\beta=0.25$, 0.10, 0.0.

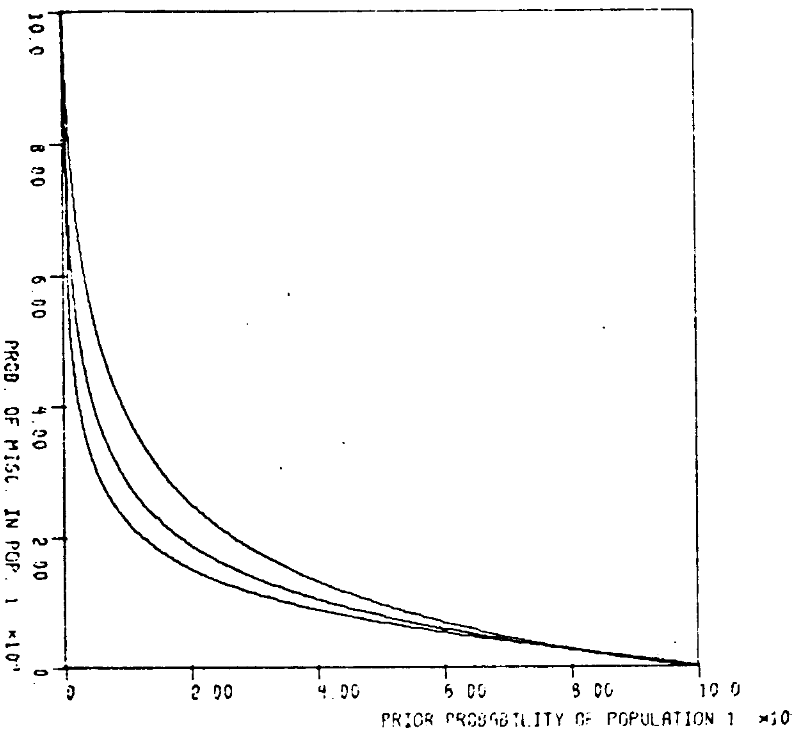


Fig. 2.6.11. Probability of misclassification in population 1 with single sample location contamination: $\delta=3$; $\underline{c}_2=(0.5,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25.$*

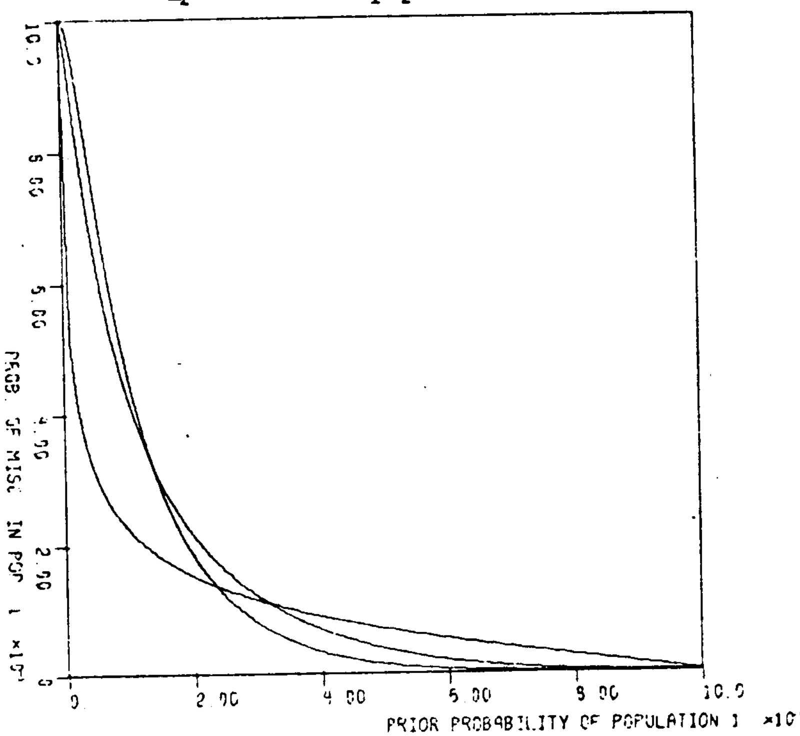


Fig. 2.6.12. Probability of misclassification in population 1 with single sample location contamination: $\delta=3$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25.$*

*The three graphs in the figure (from top to bottom at $p < .1$) correspond to $\beta= 0.25, 0.10, 0.0$.

Fig. 2.6.13. Probability of misclassification in population 2 with single sample location contamination: $\delta=1$; $\underline{c}_2=(-1.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25$.*

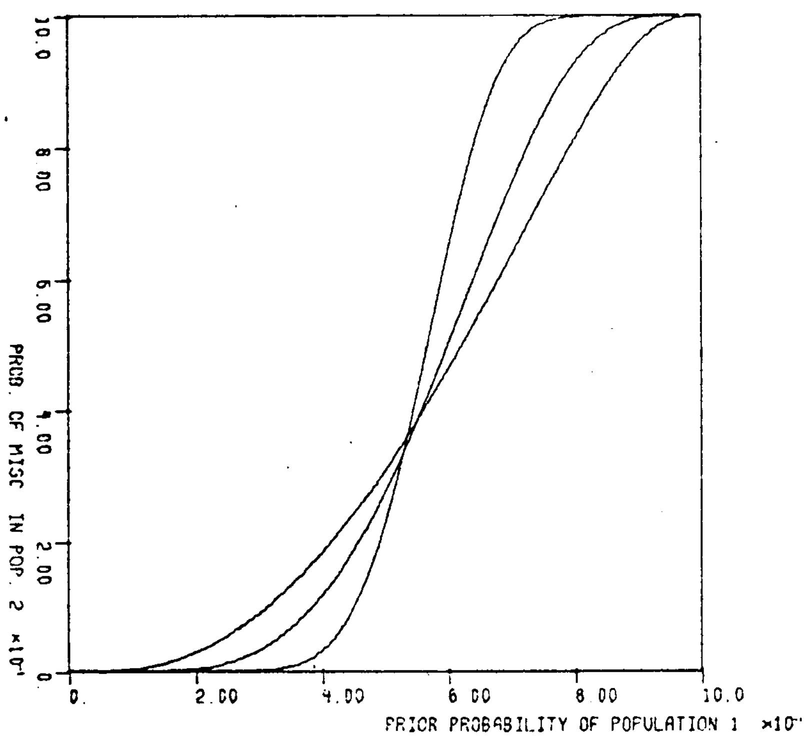
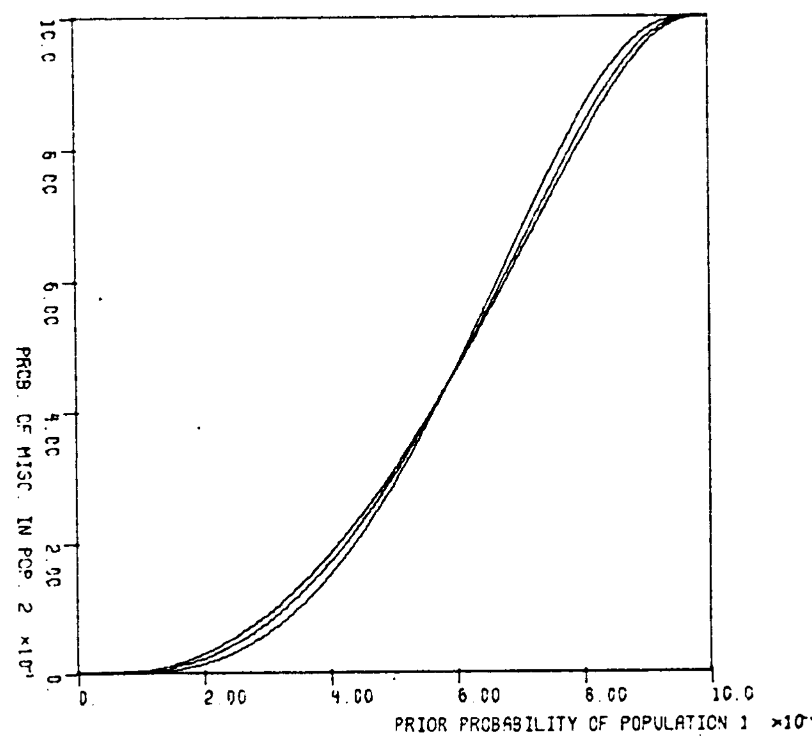


Fig. 2.6.14. Probability of misclassification in population 2 with single sample location contamination: $\delta=1$; $\underline{c}_2=(0.5,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25$.*



*The three graphs in the figure (from top to bottom at $p > .5$) correspond to $\beta= 0.25, 0.10, 0.0$.

Fig. 2.6.15. Probability of misclassification in population 2 with single sample location contamination: $\delta=1$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0$, 0.10, 0.25.*

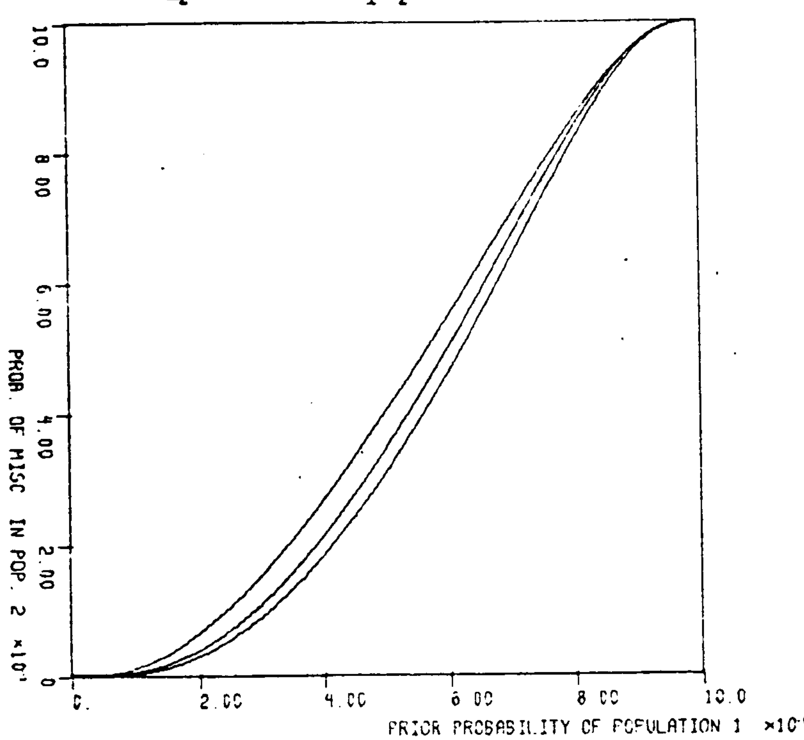
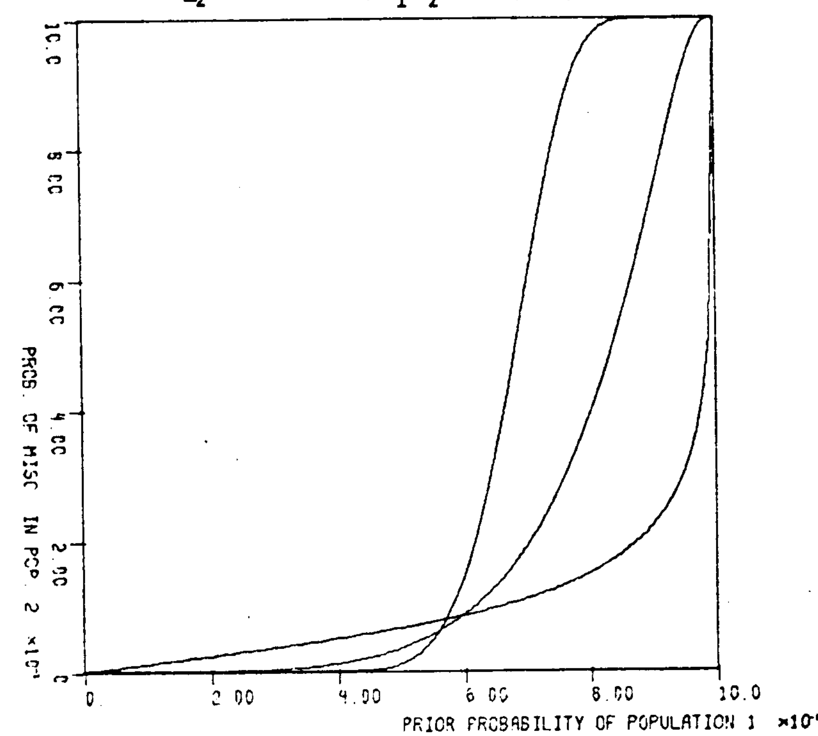


Fig. 2.6.16. Probability of misclassification in population 2 with single sample location contamination: $\delta=3$; $\underline{c}_2=(-1.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25$.**



*The three graphs in the figure (from top to bottom) correspond to $\beta=0.25$, 0.10, 0.0.

**The three graphs in the figure (from top to bottom at $p > .5$) correspond to $\beta=0.25$, 0.10, 0.0.

Fig. 2.6.17. Probability of misclassification in population 2 with single sample location contamination: $\delta=3$; $\underline{c}_2=(0.5,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25$.*

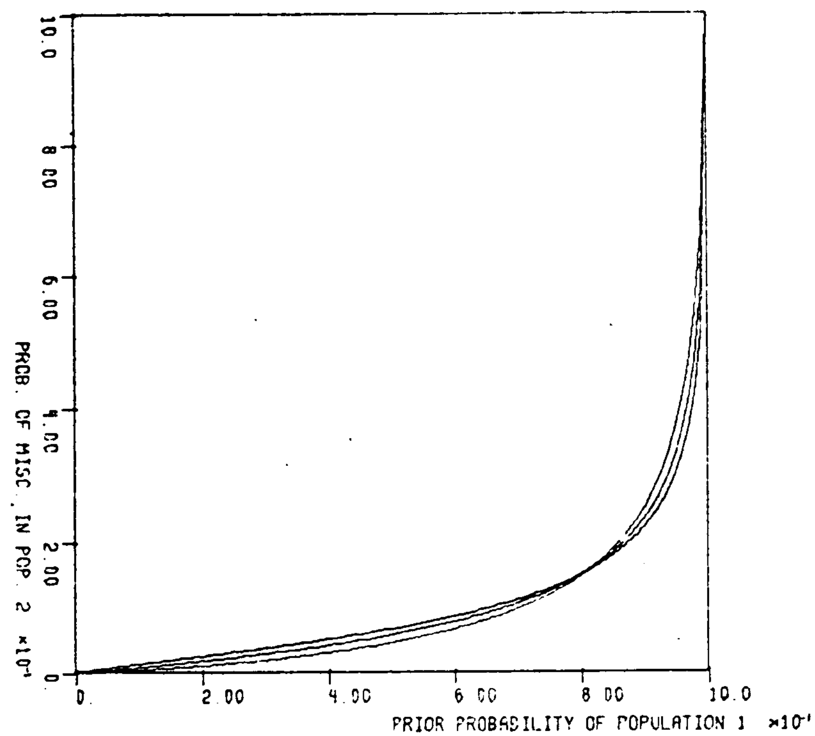
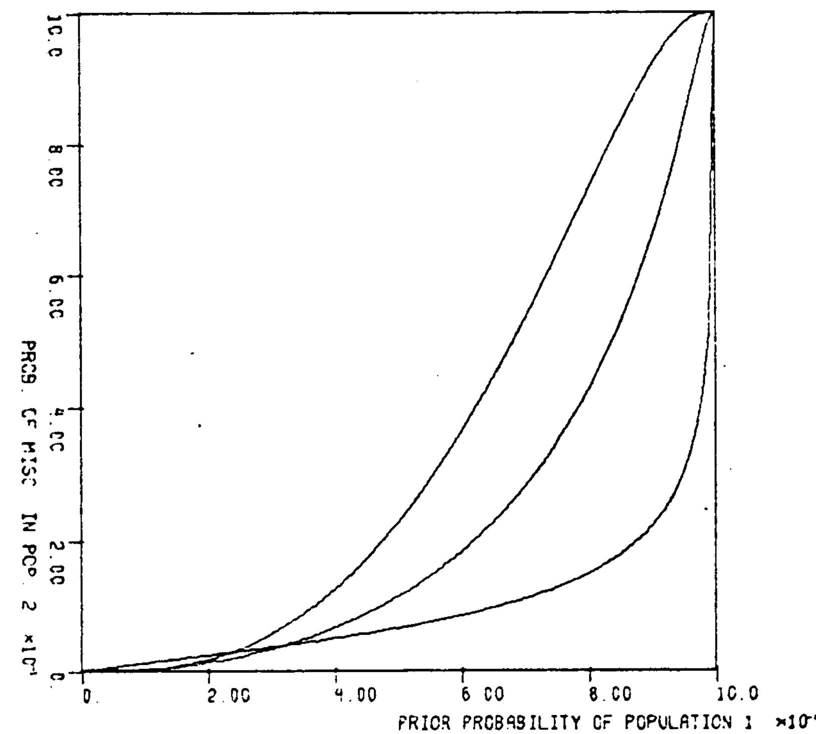


Fig. 2.6.18. Probability of misclassification in population 2 with single sample location contamination: $\delta=3$; $\underline{c}_2=(3.0,0.0,0.0)$; $n_1=n_2$; $\beta=0.0,0.10,0.25$.*



*The three graphs in the figure (from top to bottom at p>.8) correspond to $\beta=0.25$, 0.10, 0.0.

2.7. Summary

In this chapter, we have investigated the asymptotic performance of Fisher's linear discriminant function when the initial samples are contaminated. We have seen that with scale contamination and equal prior probabilities for π_1 and π_2, the LDF is optimal, i.e., the contamination has no effect on our ability to discriminate between the two populations. When the prior probabilities are unequal however, i.e., $p \neq \frac{1}{2}$, scale contamination can hurt our ability to discriminate. With mild contamination (the contaminating covariance matrix four or nine times as large as the uncontaminated matrix), only mild effects are observed. When the contaminating matrix becomes twenty-five to one hundred times the uncontaminated matrix, contamination has a definite harmful effect.

Location contamination can be even more harmful than scale contamination, depending upon the direction of contamination. The worst effects occur when the mean of the contaminating distribution is on the opposite side of the mean of the uncontaminated population.

CHAPTER III

INTRODUCTION TO THE SIMULATION STUDY

3.1. Introduction

In Chapter II, it was shown that for large samples scale contamination of the initial samples has no effect on our ability to discriminate between the two populations π_1: $N(\underline{0},I)$ and π_2: $N((\delta,0,\ldots,0),I)$ when the prior probabilities of the two populations are equal, i.e., $p = \frac{1}{2}$. In the remainder of this study, we shall examine the small sample effects of such scale contamination. We wish to answer the following questions:

(1) How well does the usual sample LDF perform with scale-contaminated initial samples when $p = \frac{1}{2}$?

(2) How is the performance of the LDF affected by changing the amount of contamination, α and β, and the type of contaminating distribution?

(3) In investigating questions (1) and (2) we will find that with small samples, scale contamination can have a harmful effect on our ability to discriminate between π_1 and π_2 using Fisher's LDF. Therefore, several "robust linear discriminant functions" will be proposed. Are any of these functions worthwhile alternatives to the usual LDF when the initial samples are scale contaminated? How do these robust LDFs perform when $\alpha = \beta = 0$, i.e., when there is no contamination?

We seek to answer these questions through a series of computer sampling experiments. This chapter describes the methodology involved in carrying out the experiments and the proposed alternatives to Fisher's LDF.

3.2. Proposed Alternatives to Fisher's LDF When the Initial Samples are Scale Contaminated

3.2.1. Restatement of the Problem

Let us first restate once again our model. We have:

π_1: $N_m(\underline{0}, I)$

π_2: $N_m((\delta, 0, \ldots, 0), I)$

π_1^c: $(1-\alpha)\ N_m(\underline{0}, I) + \alpha G_m(\underline{0}, \underline{\theta})$

π_2^c: $(1-\beta)\ N_m((\delta, 0, \ldots, 0), I) + \beta G_m((\delta, 0, \ldots, 0), \underline{\theta})$

$0 \leq \alpha \leq 1 \qquad 0 \leq \beta \leq 1 \qquad \delta > 0 \qquad p = P(\pi_1) = \frac{1}{2}.$

We wish to classify individuals into π_1 or π_2. When we attempt to draw initial samples of sizes n_1 and n_2 from π_1 and π_2 in order to form the sample linear discriminant function, however, we obtain contaminated samples from π_1^c and π_2^c. If we proceed to form the sample linear discriminant function, we will be using the contaminated LDF:

$$D_s^c(\underline{x}) = \{\underline{x} - \tfrac{1}{2}(\underline{\bar{x}}_1^c + \underline{\bar{x}}_2^c)\}'\ (S^c)^{-1}(\underline{\bar{x}}_1^c - \underline{\bar{x}}_2^c)$$

where $S^c = [(n_1-1)\ S_1^c + (n_2-1)\ S_2^c]/(n_1+n_2-2)$ and $\underline{\bar{x}}_1^c$, $\underline{\bar{x}}_2^c$, S_1^c, and S_2^c are the sample means and covariance matrices based on the contaminated samples, rather than the uncontaminated function

$$D_s(\underline{x}) = \{\underline{x} - \tfrac{1}{2}(\underline{\bar{x}}_1 + \underline{\bar{x}}_2)\}'\ S^{-1}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

which we would like to be using. (Note that $D_s(\underline{x})$ does not actually exist under our model since we are unable to obtain the estimates $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S. The function $D_s(\underline{x})$ is the sample optimal rule, the rule which, if it could be obtained, would best discriminate between π_1 and π_2.)

The sample discriminant function $D_s(\underline{x})$ is obtained by replacing $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ by $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S in the optimal discriminant function

$$D_{OPT}(\underline{x}) = \{\underline{x} - \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2)\}' \, \Sigma^{-1} \, (\underline{\mu}_1-\underline{\mu}_2).$$

The sample means and sample covariance matrix, $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S, are used as estimates of $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ, based on the assumption that the underlying distributions are normal. In order to protect against the long or heavy tails arising in the distribution of the scale-contaminated samples, we suggest three types of alternatives to Fisher's LDF, three types of "robust linear discriminant functions" to be used when there is concern about the possibility of scale contamination in the initial samples: (1) a robust LDF formed by replacing $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ in the optimal discriminant function by $\tilde{\underline{\mu}}_1$, $\tilde{\underline{\mu}}_2$, and $\tilde{\Sigma}$, robust estimates of location and scale, (2) a robust LDF based on robust regression coefficients, and (3) a "trimmed" LDF. In the following three sections, we describe these three types of robust linear discriminant functions and the particular alternatives to Fisher's LDF to be evaluated in this study.

3.2.2. The Type I Robust Linear Discriminant Function: Combinations of Robust Location and Robust Scale Estimates

The first type of alternative to Fisher's linear discriminant function to be considered in this study is a straightforward modification of the LDF. Instead of using $\underline{\bar{x}}_1$, $\underline{\bar{x}}_2$, and S to estimate $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ in the optimal discriminant function, we use $\tilde{\underline{\mu}}_1$, $\tilde{\underline{\mu}}_2$, and $\tilde{\Sigma}$, robust estimates of $\underline{\mu}_1$, $\underline{\mu}_2$, and Σ. Sixteen robust LDFs of this type will be considered in this study. Each of eight robust location estimates will be paired with each of two robust scale estimates. The resulting robust sample linear discriminant function may be written as

$$D_{RS}(\underline{x}) = \{\underline{x} - \frac{1}{2}(\tilde{\underline{\mu}}_1+\tilde{\underline{\mu}}_2)\}' \tilde{\Sigma}^{-1} (\tilde{\underline{\mu}}_1-\tilde{\underline{\mu}}_2).$$

The location vectors $\tilde{\underline{\mu}}_1$ and $\tilde{\underline{\mu}}_2$ will be made up of m univariate estimates. The location estimates studied here were selected on the basis of the extensive study by Andrews, et al. [3], described in section 1.3.4. Two criteria were considered in selecting from among the sixty-eight estimates considered in that study: (1) simplicity: an estimator was desired which could be fairly easily explained to a layman; (2) performance of the estimator in the Andrews study. The following location estimates were selected:

(1) 15% Trimmed Mean (TRIM15)

The 15% trimmed mean of the n ordered observations $x_1, x_2, \ldots, x_n$ is given by

$$\tilde{\mu} = \frac{\{wx_{[0.15n+1]} + x_{[0.15n+2]} + \cdots + x_{n-[0.15n]-1} + wx_{n-[0.15n]}\}}{n(1-2(0.15))}$$

where $w = 1 + [0.15n] - 0.15n$ and [ ] denotes the greatest integer function. This estimate is essentially the mean of the observations which remain after deleting the 15% smallest and 15% largest observations. When 0.15 is not a multiple of 1/n, [0.15n] points are removed from each end and the largest and smallest remaining points are weighted accordingly.

(2) 25% Trimmed Mean (TRIM25)

(3) Huber Proposal 2: $k = 1.2$ (HUB)

Huber estimates are one type of what Andrews, et al., have called M-estimates (maximum likelihood type estimates). An M-estimate of location is a solution $\tilde{\mu}$ of an equation of the form $\sum_i \Psi\left(\frac{x_{(i)} - \tilde{\mu}}{s}\right) = 0$ where Ψ is an odd function and s is estimated independently or simultaneously from an equation of the form $\sum_i \chi\left(\frac{x_{(i)} - \tilde{\mu}}{s}\right) = 0$. (The $x_{(i)}$ are the order statistics of the sample.)

Estimates (3) - (6) are all M-estimates. The corresponding Ψ functions are illustrated in Figure 3.2.2. Note that if $\Psi(x) = x$ for all x, the resulting estimate is the arithmetic mean. The M-estimates used as robust estimates reduce the influence of observations in the tails of the distribution.

Huber's estimates take $\Psi(x;k)$ as

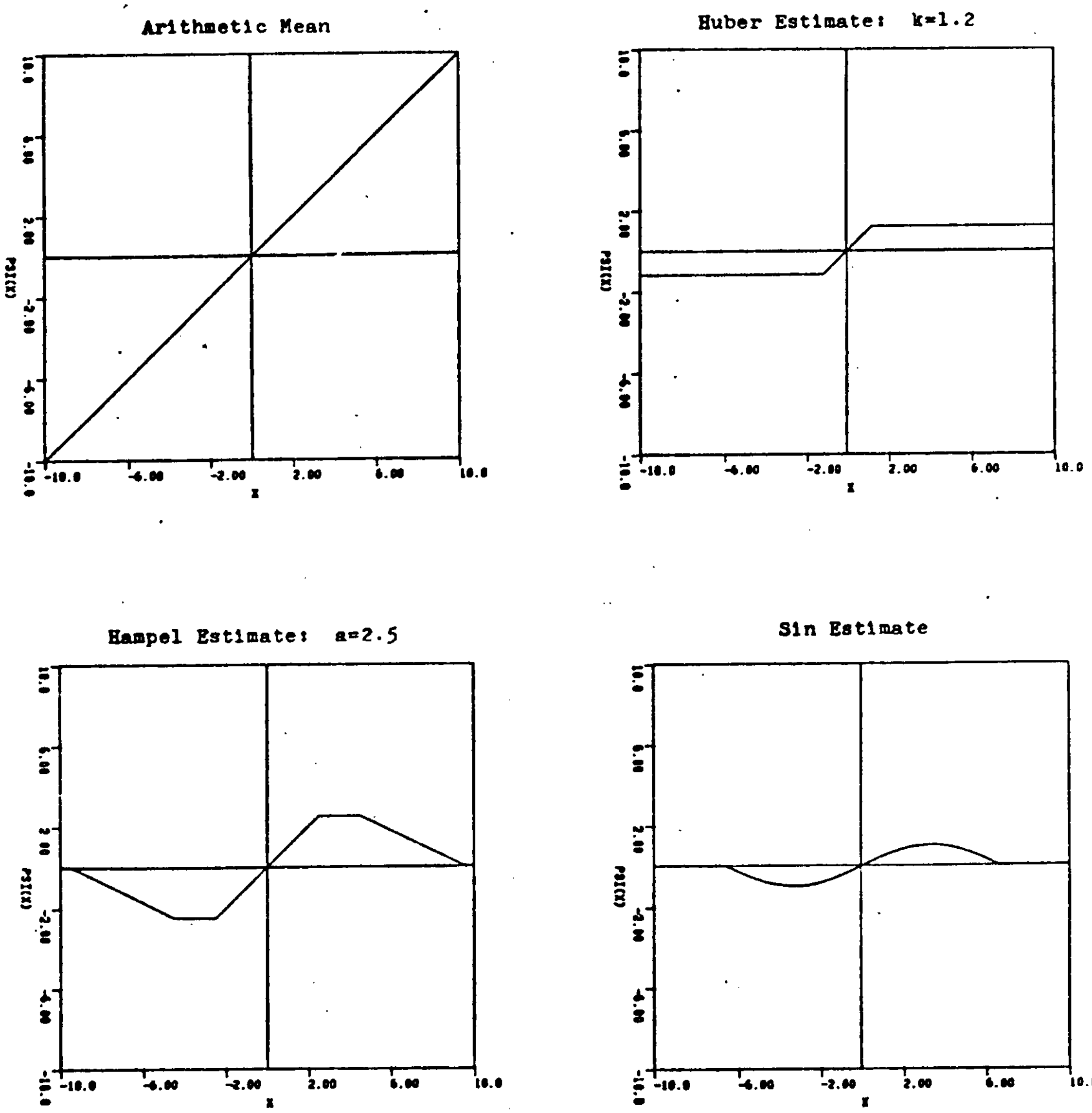
$$\Psi(x;k) = \begin{cases} -k & -\infty < x < -k \\ x & -k \leq x \leq k \\ k & k < x < \infty \end{cases}$$

and $\chi(x) = \Psi^2(x;k) - \int \Psi^2(x;k)\ \Phi\ (dx)$. The basic idea here is that if the observation is within a certain range, or equivalently if the residual $(x-\tilde{\mu})/\tilde{s}$ is within a certain range, say $(-k,k)$, then the observation is left as it is. If the residual is large ($>k$), it is replaced by the value k (x is replaced by $k\tilde{s} + \tilde{\mu}$); if the residual has a large negative value ($< -k$) it is replaced by $-k$. Huber solves the two equations

$$\sum_i \Psi\left(\frac{x_{(i)} - \tilde{\mu}}{\tilde{s}}\right) = 0$$

$$\sum_i \chi\left(\frac{x_{(i)} - \tilde{\mu}}{\tilde{s}}\right) = 0$$

Fig. 3.2.2. Psi Functions Corresponding to M-Estimates of Location.



simultaneously for $\tilde{s}$ and $\tilde{\mu}$, using an iterative procedure beginning with μ_0 = median and s_0 = interquartile range/1.35.

The Huber estimate is similar to a Winsorized estimate. In Winsorization, all observations larger than $x_{[n(1-\alpha)]}$ are replaced by $x_{[n(1-\alpha)]}$ and all observations smaller than $x_{[\alpha n+1]}$ are replaced by $x_{[\alpha n+1]}$.

Our third estimate is Huber's estimate with k = 1.2.

(4) Hampel's Huber: k = 1.5, robust scale (HAMHUB)

Hampel suggested a modification of Huber's Proposal 2. The estimate uses the same Ψ function described in (3), but s is estimated independently by $(\text{med}|x_{(i)} - \text{med}|)/0.6754$.

Our fourth estimate is Hampel's modification of Huber's estimate with k = 1.5.

(5) Hampel's Estimate: a = 2.5, b = 4.5, c = 9.5 (HAMP)

Hampel's estimates are another type of M-estimate with

$$\Psi(x) = \text{sign}(x) \cdot \begin{cases} |x| & 0 \le |x| < a \\ a & a \le |x| < b \\ \dfrac{c-|x|}{c-b}\, a & b \le |x| < c \\ 0 & |x| \ge c. \end{cases}$$

Again $\tilde{\mu}$ is found by solving $\sum_i \Psi\left(\frac{x_{(i)}-\tilde{\mu}}{s}\right) = 0$ iteratively beginning at $\tilde{\mu}_0$ = median $x_{(i)}$. The estimator s is taken to be med $|x_{(i)}$ - med$|$. Estimate 5 is Hampel's estimate with a = 2.5, b = 4.5, c = 9.5.

(6) Sin Estimate (SIN)

The sin estimate is another M-estimate, introduced by Andrews. The Ψ function is taken as

$$\Psi(x) = \begin{cases} \sin(x/2.1) & |x| \leq 2.1\ \pi \\ 0 & |x| \geq 2.1\ \pi. \end{cases}$$

The equation $\sum_i \Psi\ (\frac{x_{(i)}-\tilde{\mu}}{s}) = 0$ is solved iteratively with $s = 2.1 \text{ med } |x_{(i)}-\tilde{\mu}|$. The initial estimate of $\tilde{\mu}$ is taken to be $\tilde{\mu}_0 = \text{med } x_{(i)}$.

Note that this estimate is a smoothed version of Hampel's estimate.

(7) Gastwirth Estimate (GAST)

This simple estimate proposed by Gastwirth is a linear combination of certain of the order statistics: the median and the upper and lower tertiles:

$$\tilde{\mu} = 0.3\ x_{([(n/3)+1])} + 0.4 \text{ med} + 0.3\ x_{(n-[n/3])}.$$

(8) Trimean (TRI)

The trimean is another simple estimate based on a linear combination of order statistics:

$$\tilde{\mu} = \frac{1}{4} h_1 + \frac{1}{2} (\text{med}) + \frac{1}{4} h_2$$

where

$$h_1 = \begin{cases} x_{([n/4])} & n \neq \text{multiple of } 4 \\ \frac{1}{2}(x_{(n/4)} + x_{(n/4+1)}) & n = \text{multiple of } 4 \end{cases}$$

$$h_2 = \begin{cases} x_{(n+1-[(n+3)/4])} & n \neq \text{multiple of } 4 \\ \frac{1}{2}(x_{(n+1-(n/4))} + x_{(n-n/4)}) & n = \text{multiple of } 4. \end{cases}$$

The symbols in parentheses are used to identify the corresponding LDFs.

We have thus selected eight robust estimates of location to be paired with a robust scale estimate in order to produce a robust linear discriminant function. Two of the estimates are trimmed means, four are M-estimates, and two are linear combinations of selected order statistics.

Before describing the covariance estimates to be paired with these robust location estimates, let us try to give some idea of how these eight estimates performed in the Andrews study. Several criteria were considered by Andrews, et al., in the evaluation of the relative performance of the various estimators, including variances and percentage points of the distributions of the estimators, correlations between the estimators, and breakdown points (how many extreme outliers an estimator can tolerate before it breaks down completely). Four general classes of estimators were considered: (1) linear combinations of order statistics (including trimmed means and the Gastwirth and trimean estimates), (2) M-estimates, (3) skipped estimates (estimates which involve simple iterative rejecting based on a simple scale estimate), (4) others. All of the estimates considered in our study fall into one of the first two categories.

There is no one estimator in the Andrews study which can be said to be "best" in all cases. Among the four general classes of estimators, however, the second class--M-estimates--seems to do best. The Hampel estimates and the closely related sin estimate are highly recommended in the Andrews study. They restrict the influence of any fixed fraction of wrong observations on the value of the estimator; they reject outliers which are too far from the bulk of the data; they are little affected by rounding, grouping and other local inaccuracies. For mild contamination and samples of size 20, the Hampel estimate performed better than the sin estimate; for heavy contamination, the sin estimate performed better.

The Huber estimate performed quite well for mild contamination but was not recommended for heavy contamination.

The 25% trimmed mean was recommended by several of the investigators when simplicity is desired or when the estimate is to be calculated by hand.

Based on a study of sixty-five of the estimators and seventeen distributions, Hampel classifies the estimators into five classes: "mildly bad" (an estimator whose variance is two to four times the variance of the best available estimator), "seriously bad" (variance four times the minimal variance), "catastrophic" (variance ten times the minimal variance),"completely disastrous" (variance one hundred times the minimal variance), others. Among the "never bad" estimators (never bad for any distribution or sample size investigated by Hampel) are the 25% trimmed mean, the Gastwirth estimate, the sin estimate, and Hampel's estimate with a = 2.5.

If, therefore, the performance of the Type I LDFs parallels the performance of the corresponding location estimate in the Andrews study, we might expect that LDFs (5) and (6) would come out at the top with (5) performing better than (6) for mild contamination and (6) performing better than (5) for heavy contamination. We would expect LDF (3) to perform well under mild contamination. LDF (2) would be expected to perform better than (1) and (7) better than (8). The Gastwirth and 25% trimmed mean LDFs might be recommended for their simplicity.

Our robust covariance estimates will be based on the procedure described by Gnanadesikan and Kettenring [16]. This procedure guarantees a non-singular estimate. The procedure is outlined below.

(a) Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ denote n observations. Rank the $\underline{x}_i$ in terms of $d_i = (\underline{x}_i - \tilde{\underline{\mu}})' S_0^{-1} (\underline{x}_i - \tilde{\underline{\mu}})$ where $\tilde{\underline{\mu}}$ is a robust location estimate and S_0 is a diagonal weighting matrix whose diagonal elements are robust estimates of the scales of the m variables. The quantity d_i is a weighted distance of the point $\underline{x}_i$ from the location estimate $\tilde{\underline{\mu}}$. We will take S_0 to be diagonal with $s_{jj} = \text{med}\ |x_{ij} - \underset{i}{\text{med}}\ x_{ij}|$.

(b) Delete the 100 a% observations with largest ranks, where $0 < a < 1$ and $[n(1-a)] > m$. (Recall that m denotes the number of variables.)

(c) Compute

$$S_1 = \sum_{g \varepsilon G} (\underline{x}_g - \tilde{\underline{\mu}})(\underline{x}_g - \tilde{\underline{\mu}})'$$

where G denotes those observations not deleted in (b). The restriction $[n(1-a)] > m$ insures that S_1 is nonsingular.

(d) Rank $(\underline{x}_i - \tilde{\underline{\mu}})' S_1^{-1} (\underline{x}_i - \tilde{\underline{\mu}})$ for all n observations.

(e) Delete the 100 b% observations which have largest ranks, $[n(1-b)] > m$.

(f) Compute

$$S = \frac{k}{[n(1-b)]} \sum_{r \varepsilon R} (\underline{x}_r - \tilde{\underline{\mu}})(\underline{x}_r - \tilde{\underline{\mu}})'$$

where R denotes those observations not deleted in (e).

The constant k should be chosen to make S an unbiased estimate of $\ddagger$.

Gnanadesikan and Kettenring have indicated that the problem of determining the constant k is not easy. They suggest that it may depend on factors other than the trimming proportions a and b, such as the sample size and number of variables. In order to investigate the problem of the choice of k for the type of model considered in this study ($\underline{x} \sim N_m(\underline{\mu}, I)$), a small simulation experiment was performed. For each combination of parameters (n,m,a,b) considered, 1000 samples were drawn from a $N_m(\underline{0}, I)$ distribution. For each sample, the Gnanadesikan-Kettenring covariance matrix was calculated as described above. Knowing that $\Sigma = I$ in our model, we determine the value of k which makes $\sum_i s_{ii}/m = 1$, i.e., the average of the diagonal elements equal to 1. The simulation was performed for m = 1, 4; n = 10, 25, 100; a = b = 0.05, 0.10; and m = 8; n = 20, 50; a = b = 0.05, 0.10. Note that the sample sizes for m = 8 are different from those for m = 1, 4. When the small sample studies of the LDF were originally planned, it was intended to consider the eight variable case and to consider unequal prior probabilities ($p \neq \frac{1}{2}$). Samples of size 20 and 50 were considered more reasonable for the eight variable LDF sampling experiments than samples of size 12. Samples of size 100 were not included in the experiments to determine k for m = 8 because of the limitations in computer funds.

The same procedure was then carried out for 1000 samples from the moderately contaminated distribution: 0.90 N(0,1) + 0.10 N(0,1)/U(0,1). (The N(0,1)/U(0,1) distribution has density $f(x) = \frac{1}{\sqrt{2\pi}\, x^2}\,[1-e^{-1/2\, x^2}]$. It is a long tailed distribution. The 75^{th}, 95^{th}, 97.5^{th}, and 99.5^{th} cumulative percentage points of the contaminated density 0.90 N(0,1) + 0.10 N(0,1)/U(0,1) are 0.72, 1.85, 2.36, 7.98 as compared to the corresponding points for the N(0,1) distribution: 0.675, 1.645, 1.96, 2.58.) If we knew the amount and type of contamination in our sample, we would determine k for that particular contaminated distribution. In practice, however, we would not know the amount or type of contamination. Thus, we would base our estimate either on a normal distribution or on some moderately contaminated distribution. The value of k determined on the basis of the normal distribution will be too large if there is any contamination in the sample.

Tables 3.2.2.1 and 3.2.2.2 give the average values of k obtained (over the 1000 replications) and the corresponding standard deviation for each set of parameters. The appropriate value of k does seem to be dependent on sample size, number of variables, the trimming proportions a and b, and the type of distribution sampled. In general, as the trimming proportion increases, the value of k increases; as the sample size increases, the value of k decreases; as the number of variables increases the value of k decreases; as the contamination becomes more severe, the value of k decreases.

We would like to be able to choose a value for k independent of the contaminating distribution in order that our estimate be robust against various long tailed contaminating distributions. Since the best value of k seems to be dependent on the type of contaminating distribution, if we had no idea about the type of contamination in our samples (mild or heavy), we would most likely select a moderate contaminating distribution and base our estimate of k on that distribution.

In the particular case considered in this study ($p = \frac{1}{2}$), the value of k is not important. An individual with measurements $\underline{x}$ is assigned to π_1 if $D_s(\underline{x}) \geq 0$; otherwise $\underline{x}$ is assigned to π_2. Changing the value of k multiplies $D_s(\underline{x})$ by a constant and thus does not change

TABLE 3.2.2.1

BIAS CORRECTION FACTOR IN THE GNANADESIKAN-KETTENRING ROBUST SCALE ESTIMATE
SAMPLED DISTRIBUTION = N(0,1)

		a = b = 0.05			a = b = 0.10		
		m = 1	m = 4	m = 8	m = 1	m = 4	m = 8
n = 10	$\bar{k}$	1.91	1.29	--	2.59	1.44	--
	s_k	1.16	0.30	--	1.59	0.35	--
n = 20	$\bar{k}$	--	--	1.13	--	--	1.19
	s_k	--	--	0.13	--	--	0.15
n = 25	$\bar{k}$	1.64	1.15	--	1.94	1.16	--
	s_k	0.57	0.18	--	0.76	0.19	--
n = 50	$\bar{k}$	--	--	1.11	--	--	1.15
	s_k	--	--	0.08	--	--	0.09
n = 100	$\bar{k}$	1.35	1.07	--	1.64	1.11	--
	s_k	0.21	0.08	--	0.26	0.09	--

TABLE 3.2.2.2

BIAS CORRECTION FACTOR IN THE GNANADESIKAN-KETTENRING ROBUST SCALE ESTIMATE
SAMPLED DISTRIBUTION = 0.90 N(0,1) + 0.10 N(0,1)/U(0,1)

		a = b = 0.05			a = b = 0.10		
		m = 1	m = 4	m = 8	m = 1	m = 4	m = 8
n = 10	$\bar{k}$	1.76	0.95	--	1.80	1.02	--
	s_k	1.49	0.43	--	1.28	0.42	--
n = 20	$\bar{k}$	--	--	0.63	--	--	0.86
	s_k	--	--	0.31	--	--	0.29
n = 25	$\bar{k}$	1.34	0.90	--	1.61	1.03	--
	s_k	0.55	0.30	--	0.68	0.28	--
n = 50	$\bar{k}$	--	--	0.72	--	--	0.89
	s_k	--	--	0.22	--	--	0.20
n = 100	$\bar{k}$	1.10	0.82	--	1.39	1.01	--
	s_k	0.23	0.18	--	0.28	0.15	--

the assignment of any individual. When $p \neq \frac{1}{2}$, $\underline{x}$ is assigned to π_1 if $D_s(\underline{x}) \geq \ln(\frac{1-p}{p})$. In this case, the choice of k does affect the classification of individuals.

The results of the sampling experiments related to the determination of k are presented here to give some idea as to how one might go about getting an estimate of k if such an estimate were required, to indicate the magnitude of the factor k and its dependence on the sample size, number of variables, and amount of contamination. If an estimate of k were required, one might fit an analysis of variance model to the simulated data and estimate k on the basis of the model.

The two covariance estimates to be used in this study are the Gnanadesikan-Kettenring estimates with a = b = 0.05 and a = b = 0.10. Each of these robust scale estimates will be paired with each of the eight robust location estimates described above to give sixteen Type I robust LDFs. The vector $\tilde{\underline{\mu}}$ in the Gnanadesikan-Kettenring estimate will be the corresponding location estimate (1) - (8) being used for that particular LDF. Separate robust scale estimates will be computed for π_1^c and π_2^c and the two estimates pooled to give S^c.

3.2.3. The Type II Robust Linear Discriminant Function: LDF Based on Robust Regression Coefficients (ROBREG)

The second type of robust discriminant function to be considered is based on the relation between Fisher's LDF and regression analysis described in section 1.3.4. If we perform a regression analysis using $x_1,\ldots,x_m$ as the independent variables and let the dependent variable y be defined by

$$y = \begin{cases} \dfrac{n_2}{n_1+n_2} & \text{if} \quad \underline{x} \in \pi_1 \\ -\dfrac{n_1}{n_1+n_2} & \text{if} \quad \underline{x} \in \pi_2 \end{cases}$$

then the resulting regression coefficients will be proportional to Fisher's linear discriminant function coefficients. So a robust discriminant function may be constructed by performing a regression analysis and using robust regression coefficients. A procedure for obtaining such robust coefficients is described by Forsythe [11]. The idea behind this procedure is that instead of minimizing the sum of squared deviations $\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_{1i}-\ldots-\beta_mx_{mi})^2$ as we do in finding the usual least squares regression coefficients, we minimize $\sum_{i=1}^{n}|y_i-\beta_0-\beta_1x_{1i}-\ldots-\beta_mx_{mi}|^q$, $1 < q < 2$.

The resulting discriminant function may be written as $D_{RS}(\underline{x}) = \underline{b}_q'\underline{x}$ where $\underline{b}_q$ is chosen to minimize the sum of q^{th} power absolute deviations. An individual with measurements $\underline{x}$ is assigned to π_1 if $\underline{b}_q'\underline{x} \geq c$; otherwise $\underline{x}$ is assigned to π_2. In general (for any discriminant function $\underline{b}'\underline{x}$), the cutoff point c is chosen to minimize the total probability of misclassification: $P = pP_1 + (1-p)P_2$. The general formula for c is derived below.

We wish to find the value of c which minimizes $P = pP_1 + (1-p)P_2$. Since in π_i, $\underline{x} \sim N(\underline{\mu}_i, I)$ then $\underline{b}'\underline{x} \sim N(\underline{b}'\underline{\mu}_i, \underline{b}'\underline{b})$ in π_i. Therefore,

$$\begin{aligned} P_1 &= P(A_2|\pi_1) = P(\underline{b}'\underline{x} < c|\pi_1) \\ &= P\left(\frac{\underline{b}'\underline{x} - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}} < \frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right) \end{aligned}$$

$$= \Phi\left(\frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right)$$

$$P_2 = P(A_1|\pi_2) = P(\underline{b}'\underline{x} \geq c|\pi_2)$$

$$= \Phi\left(\frac{\underline{b}'\underline{\mu}_2 - c}{\sqrt{\underline{b}'\underline{b}}}\right)$$

$$P = p\Phi\left(\frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right) + (1-p)\ \Phi\left(\frac{\underline{b}'\underline{\mu}_2 - c}{\sqrt{\underline{b}'\underline{b}}}\right).$$

Minimizing P with respect to c, we obtain

$$\frac{dP}{dc} = \left(\frac{p}{\sqrt{\underline{b}'\underline{b}}}\right)\ \phi\left(\frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right) - \left(\frac{1-p}{\sqrt{\underline{b}'\underline{b}}}\right)\ \phi\left(\frac{\underline{b}'\underline{\mu}_2 - c}{\sqrt{\underline{b}'\underline{b}}}\right) = 0$$

$$\frac{\phi\left(\frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right)}{\phi\left(\frac{\underline{b}'\underline{\mu}_2 - c}{\sqrt{\underline{b}'\underline{b}}}\right)} = \frac{1-p}{p}$$

$$\exp\left(-\frac{1}{2}\left(\frac{c - \underline{b}'\underline{\mu}_1}{\sqrt{\underline{b}'\underline{b}}}\right)^2 + \frac{1}{2}\left(\frac{\underline{b}'\underline{\mu}_2 - c}{\sqrt{\underline{b}'\underline{b}}}\right)^2\right) = \frac{1-p}{p}$$

$$c = \frac{(\underline{b}'\underline{\mu}_1)^2 - (\underline{b}'\underline{\mu}_2)^2 + 2\ \underline{b}'\underline{b}\ \ln(\frac{1-p}{p})}{2\ \underline{b}'(\underline{\mu}_1 - \underline{\mu}_2)}\ ,\ \underline{\mu}_1 \neq \underline{\mu}_2.$$

Note that when $p = \frac{1}{2}$, then

$$c = \frac{(\underline{b}'\underline{\mu}_1)^2 - (\underline{b}'\underline{\mu}_2)^2}{2\ \underline{b}'(\underline{\mu}_1 - \underline{\mu}_2)} = \frac{1}{2}(\underline{b}'\underline{\mu}_1 + \underline{b}'\underline{\mu}_2),$$

i.e., the cutpoint for discriminating between π_1 and π_2 is halfway between $E(\underline{b}'\underline{x}|\pi_1)$ and $E(\underline{b}'\underline{x}|\pi_2)$. When $p \neq \frac{1}{2}$, the cutpoint is modified to account for the unequal prior probabilities of the two populations.

Thus the discriminant rule using the robust regression discriminant function becomes: "Assign $\underline{x}$ to π_1 if $\underline{b}_q'\underline{x} \geq c$; otherwise assign $\underline{x}$ to π_2," where

$$c = \frac{(\underline{b}_q'\underline{\mu}_1)^2 - (\underline{b}_q'\underline{\mu}_2)^2 + 2\,\underline{b}_q'\,\underline{b}_q\,\ell n(\frac{1-p}{p})}{2\,\underline{b}_q'(\underline{\mu}_1 - \underline{\mu}_2)}$$

and $\underline{b}_q$ is the vector which minimizes the sum of q^{th} power absolute deviations: $\sum_i |y_i - \underline{b}'\underline{x}_i|^q$.

3.2.4. The Type III Robust Linear Discriminant Function: Trimmed LDF (TLDF15, TLDF25)

The third type of linear discriminant function proposed as an alternative to Fisher's LDF when the initial samples are contaminated is the "a% trimmed LDF." It is constructed in a manner similar to the usual sample LDF. However, those observations which are poor for discrimination are eliminated before forming the LDF. The procedure is outlined below.

(1) Draw a sample of size n_1 from π_1 and a sample of size n_2 from π_2. (The samples actually obtained are from π_1^c and π_2^c.)

(2) For each of the n_1 individuals in sample 1, compute the usual sample LDF: $D_s(\underline{x}) = \{\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2)\}' S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$.

(3) Rank the $D_s(\underline{x})$ computed in (2) and eliminate the observations corresponding to the a% smallest ranks and the observations corresponding to the a% largest ranks.

(4) Repeat steps (2) and (3) for the n_2 individuals in sample 2.

(5) Compute a new linear discriminant function based on those observations not eliminated in steps (3) and (4):

$D_{RS}(\underline{x}) = \{\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1^T + \bar{\underline{x}}_2^T)\}' (S^T)^{-1} (\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)$ where $\bar{\underline{x}}_1^T$ and $\bar{\underline{x}}_2^T$ are the sample means of the "trimmed samples" and S^T is the pooled covariance matrix.

Those observations in the tails of the two samples will be eliminated. If the sample is scale contaminated, these are the observations most likely to have come from the contaminating distribution. Two trimmed LDFs will be evaluated in this study: the 15% trimmed LDF and the 25% trimmed LDF.

We thus have a total of nineteen procedures to be evaluated in addition to the usual linear discriminant function. Our criterion for evaluating each of the nineteen procedures will be the probability of misclassification.

3.3. Probabilities of Misclassification Corresponding to the Three Types of Robust LDFs

The evaluation of the performance of the three types of alternatives to Fisher's linear discriminant function will be made through a series of computer sampling experiments. Samples will be generated from π_1^c and π_2^c and the probability of misclassification conditional on the particular pair of samples will be calculated for each method being evaluated. These probabilities will be compared over the twenty methods.

3.3.1. Probability of Misclassification for Fisher's LDF

The probability of misclassification using Fisher's linear discriminant function to discriminate between π_1 and π_2 is derived below.

Conditional on $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$, and S, the sample LDF $D_s(\underline{x})$ = $\{\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2)\}' S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$ is normally distributed with the following means and variances:

$$E(D_s|\pi_1;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S) = [\underline{\mu}_1 - \tfrac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2)]'\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$$

$$E(D_s|\pi_2;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S) = [\underline{\mu}_2 - \tfrac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2)]'\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)$$

$$V(D_s|\pi_1;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S) = V(D_s|\pi_2;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S) = V(D_s|\bar{\underline{x}}_1,\bar{\underline{x}}_2,S)$$

$$= (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)'\ S^{-1}\ I\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2).$$

Thus

$$P_1 = P(D_s(\underline{x}) < \ln(\tfrac{1-p}{p})\,|\pi_1;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S)$$

$$= P\left(\frac{D_s(\underline{x}) - E(D_s(\underline{x})|\pi_1;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S)}{\sqrt{V(D_s|\bar{\underline{x}}_1,\bar{\underline{x}}_2,S)}} < \frac{\ln(\frac{1-p}{p}) - E(D_s(\underline{x})|\pi_1;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S)}{\sqrt{V(D_s|\bar{\underline{x}}_1,\bar{\underline{x}}_2,S)}}\right)$$

$$= \Phi\left(\frac{\ln(\frac{1-p}{p}) - (\underline{\mu}_1 - \frac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2))'\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}{\sqrt{(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)'\ S^{-1}\ I\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}}\right)$$

$$P_2 = P(D_s(\underline{x}) \geq \ln(\tfrac{1-p}{p})\,|\pi_2;\ \bar{\underline{x}}_1,\bar{\underline{x}}_2,S)$$

$$= 1 - \Phi\left(\frac{\ln(\frac{1-p}{p}) - (\underline{\mu}_2 - \frac{1}{2}(\bar{\underline{x}}_1 + \bar{\underline{x}}_2))'\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}{\sqrt{(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)'\ S^{-1}\ I\ S^{-1}(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)}}\right)$$

$$P = pP_1 + (1-p)P_2.$$

These are the exact probabilities of misclassification for Fisher's LDF, conditional on $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$, and S.

3.3.2. Probability of Misclassification for the Type I Robust Linear Discriminant Function

The probability of misclassification due to using a Type I robust LDF: $D_{RS}(\underline{x}) = \{\underline{x} - \frac{1}{2}(\tilde{\underline{\mu}}_1 + \tilde{\underline{\mu}}_2)\}'\ \tilde{\Sigma}^{-1}(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)$ is derived in the same manner as the probability for Fisher's LDF. The estimates $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$, and S are replaced by $\tilde{\underline{\mu}}_1$, $\tilde{\underline{\mu}}_2$, and $\tilde{\Sigma}$ in section 3.3.1 giving

$$P_1 = P(A_2|\pi_1;\ \tilde{\underline{\mu}}_1,\tilde{\underline{\mu}}_2,\tilde{\Sigma}) = \Phi\left(\frac{\ln(\frac{1-p}{p})-(\underline{\mu}_1 - \frac{1}{2}(\tilde{\underline{\mu}}_1 + \tilde{\underline{\mu}}_2))'\ \tilde{\Sigma}^{-1}(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)}{\sqrt{(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)'\ \tilde{\Sigma}^{-1}\ I\ \tilde{\Sigma}^{-1}(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)}}\right)$$

$$P_2 = P(A_1|\pi_2;\ \tilde{\underline{\mu}}_1,\tilde{\underline{\mu}}_2,\tilde{\Sigma}) = 1 - \Phi\left(\frac{\ln(\frac{1-p}{p})-(\underline{\mu}_2 - \frac{1}{2}(\tilde{\underline{\mu}}_1 + \tilde{\underline{\mu}}_2))'\ \tilde{\Sigma}^{-1}(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)}{\sqrt{(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)'\ \tilde{\Sigma}^{-1}\ I\ \tilde{\Sigma}^{-1}(\tilde{\underline{\mu}}_1 - \tilde{\underline{\mu}}_2)}}\right)$$

$$P = pP_1 = (1-p)P_2.$$

3.3.3. Probability of Misclassification for the Type II Robust Linear Discriminant Function

The Type II robust linear discriminant function is the robust regression LDF. An individual with measurements $\underline{x}$ is assigned to π_1 if $\underline{b}_q'\underline{x} \geq c$ where $\underline{b}_q$ and c are defined in section 3.2.3; otherwise the individual is assigned to π_2. The probabilities of misclassification were derived in section 3.2.3 in the process of determining the appropriate value of the cutoff point c:

$$P_1 = \Phi\left(\frac{c - \underline{b}_q'\underline{\mu}_1}{\sqrt{\underline{b}_q'\underline{b}_q}}\right)$$

$$P_2 = \Phi\left(\frac{\underline{b}_q'\underline{\mu}_2 - c}{\sqrt{\underline{b}_q'\underline{b}_q}}\right)$$

$$P = pP_1 + (1-p)P_2.$$

The cutoff point c and the arguments of Φ in P_1 and P_2 involve the parameters $\underline{\mu}_1$ and $\underline{\mu}_2$. Since the parameters are assumed to be unknown, the Robust Regression procedure requires estimates of $\underline{\mu}_1$ and $\underline{\mu}_2$ as well as estimates of the regression coefficients $\underline{b}_q$. These estimates will be obtained in the same manner as the estimate of $\underline{b}_q$, i.e., by minimizing the sum of q^{th} power deviations:

$\sum_{i=1}^{n_1} |x_{ij}^{(1)} - \hat{\mu}_j^{(1)}|^q$, $j = 1,\ldots,m$, $\sum_{i=1}^{n_2} |x_{ij}^{(2)} - \hat{\mu}_j^{(2)}|^q$, $j = 1,\ldots,m$, where $x_{ij}^{(k)}$ denotes the measurement of the j^{th} variable on the i^{th} individual in sample k, $k = 1,2$.

3.3.4. Probability of Misclassification for the Type III Robust Linear Discriminant Function

Recall that the trimmed LDF may be written as

$$D_{RS}(\underline{x}) = \{\underline{x} - \frac{1}{2}(\bar{\underline{x}}_1^T + \bar{\underline{x}}_2^T)\}' \ (S^T)^{-1} \ (\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)$$

where $\bar{\underline{x}}_1^T$, $\bar{\underline{x}}_2^T$, and S^T are the sample means and pooled covariance matrix for the trimmed samples. Thus the conditional probabilities of misclassification (conditional on the given samples) are the same as those given in section 3.3.1 with $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$, and S replaced by $\bar{\underline{x}}_1^T$, $\bar{\underline{x}}_2^T$, and S^T:

$$P_1 = P(A_2|\pi_1;\ \bar{\underline{x}}_1^T, \bar{\underline{x}}_2^T, S^T)$$

$$= \Phi\left(\frac{\ln(\frac{1-p}{p}) - (\underline{\mu}_1 - \frac{1}{2}(\bar{\underline{x}}_1^T + \bar{\underline{x}}_2^T))'(S^T)^{-1}(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)}{\sqrt{(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)'(S^T)^{-1}(S^T)^{-1}(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)}}\right)$$

$$P_2 = P(A_1|\pi_2;\ \bar{\underline{x}}_1^T, \bar{\underline{x}}_2^T, S^T)$$

$$= 1 - \Phi\left(\frac{\ln(\frac{1-p}{p}) - (\underline{\mu}_2 - \frac{1}{2}(\bar{\underline{x}}_1^T + \bar{\underline{x}}_2^T))'(S^T)^{-1}(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)}{\sqrt{(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)(S^T)^{-1}(S^T)^{-1}(\bar{\underline{x}}_1^T - \bar{\underline{x}}_2^T)}}\right)$$

$$P = pP_1 + (1-p)P_2.$$

3.4. Methodology for the Small Sample Studies

3.4.1. General Procedure for the Simulation

In order to evaluate the small sample performance of Fisher's linear discriminant function and of the nineteen proposed alternatives to Fisher's LDF when the initial samples are scale contaminated, a series of computer sampling experiments was performed. Samples were drawn from the two contaminated distributions π_1^c and π_2^c:

$$\pi_1^c: \quad (1-\alpha)\ N_m(\underline{0},I) + \alpha G_m(\underline{0},\underline{\theta})$$

$$\pi_2^c: \quad (1-\beta)\ N_m((\delta,0,\ldots,0),I) + \beta G_m((\delta,0,\ldots,0),\underline{\theta})$$

$$0 \le \alpha \le 1 \qquad 0 \le \beta \le 1 \qquad \delta > 0 \quad .$$

for the following combinations of the parameters $m,\alpha,\beta,\delta,n_1,n_2,G$, and $\underline{\theta}$: $m = 1,4$; $\delta = 1,3$; $n_1 = n_2 = 12,25$; $(\alpha,\beta) = (0,0)$, $(0,0.10)$, $(0,0.25)$, $(0.10,0.10)$, $(0.25,0.25)$; $G(\underline{x}) = N_m(\underline{\mu}_i,9I)$, $G(\underline{x}) = N_m(\underline{\mu}_i,25I)$, $G(\underline{x}) = N(\underline{\mu}_i,100I)$, $G(x_j) = \frac{1}{\pi}\frac{1}{1+(x_j-\mu_{ij})^2}$, $i = 1,2 \quad j = 1,\ldots,m, \qquad \mathrm{cov}(x_j,x_{j'}) = 0$. To give some idea of the relationships among the four contaminating distributions $G(\underline{x})$, Table 3.4.1 gives cumulative percentage points for the contaminated distributions $(1-\alpha)\ N(0,1) + \alpha G(0,\theta)$.

Ten pairs of samples (one from π_1^c and one from π_2^c) were generated for each combination of parameters. For each pair of samples, the twenty discriminant functions (Fisher's LDF, sixteen Type I LDFs, one Type II LDF, two Type III LDFs) were determined and the corresponding probabilities of misclassification were computed. Within each combination of parameters, each of the ten pairs of samples can be thought of as a block and the twenty discrimination procedures as treatments.

After the twenty probabilities of misclassification (P) were calculated for each of the ten pairs of samples, the mean and variance of P (over the ten samples) were computed for each procedure within

TABLE 3.4.1

CUMULATIVE PERCENTAGE POINTS FOR SCALE CONTAMINATED DISTRIBUTIONS

Distribution	$P_{.75}$	$P_{.95}$	$P_{.975}$	$P_{.995}$
N(0,1)	0.68	1.65	1.96	2.58
0.90 N(0,1) + 0.10 N(0,9)	0.73	1.93	2.53	4.93
0.90 N(0,1) + 0.10 N(0,25)	0.74	2.10	3.42	8.22
0.90 N(0,1) + 0.10 N(0,100)	0.75	2.32	6.74	16.44
0.90 N(0,1) + 0.10 Cauchy	0.69	1.78	2.23	6.31
0.75 N(0,1) + 0.25 N(0,9)	0.83	2.65	3.85	6.16
0.75 N(0,1) + 0.25 N(0,25)	0.88	4.21	6.41	10.27
0.75 N(0,1) + 0.25 N(0,100)	0.92	8.29	12.81	20.52
0.75 N(0,1) + 0.25 Cauchy	0.73	2.07	3.16	15.89

each combination of parameters. The average values of P were then compared across the various procedures.

3.4.2. The Computer Simulation Program

The computer program used to carry out the simulation study is a FORTRAN program consisting of a main program and thirty subroutines. The subroutines are used to generate random samples, sort the data, compute robust estimates of location and scale, determine Fisher's LDF and the three types of robust LDFs, and calculate probabilities of misclassification. The programs used to compute the robust estimates of location were adapted from those given by Andrews, *et al*. [3]. All other programs were written by the author.

The sampling experiments require the generation of random samples from the contaminated population $\pi^c = (1-\alpha)\ N(\underline{\mu},I) + \alpha G(\underline{\mu},\underline{\theta})$. The remainder of this section describes the procedure used for the generation of these random samples.

In order to generate a sample of size n from the contaminated distribution π^c, first determine how many of the n observations are to be drawn from $N(\underline{\mu},I)$ and how many are to be drawn from $G(\underline{\mu},\underline{\theta})$. The probability that an observation in the sample comes from $N(\underline{\mu},I)$ is $1-\alpha$ and the probability that the observation comes from $G(\underline{\mu},\underline{\theta})$ is α. Thus we begin by drawing a sample of size 1 from a binomial distribution with parameters n and α. This tells how many observations to draw from $G(\underline{\mu},\underline{\theta})$. Denote this number by n^c. We then proceed to draw n^c observations $\underline{x}$ from $G(\underline{\mu},\underline{\theta})$ and $n-n^c$ observations from $N(\underline{\mu},I)$. Each observation is an m-dimensional vector. Since the variables are independent and the observations are independent, the n^c or $n-n^c$ m-dimensional vectors can be drawn by calling the random number generating program $n^c m$ or $(n-n^c)m$ times.

To carry out the sampling experiments, random samples from a $N(\mu,\sigma^2)$ distribution and a Cauchy distribution centered at μ are needed in addition to the binomial observation. The binomial observation is generated in the following way: n observations are drawn from a uniform (0,1) distribution. The binomial observation is the number of these uniform observations which lie between 0 and α. The IBM Scientific Subroutines Package Subroutine RANDU was used to generate uniform random numbers.

Samples from a $N(\mu,\sigma^2)$ distribution are generated by drawing samples from the N(0,1) distribution, multiplying the observation drawn by σ and adding μ. N(0,1) observations were generated by Marsaglia and Bray's adaptation of the Box Müller technique [28].

Cauchy observations were obtained by taking the ratio of a N(0,1) random number to a half-normal random number, i.e., $x = r/s$ where $r \sim N(0,1)$ and $F(s) = P(S \leq s) = 2(\Phi(s) - 0.5)$, $s > 0$; $f(s) = 2\phi(s)$, $s > 0$.

3.5. Summary

In Chapter II it was shown that for large samples scale contamination of the initial samples has no effect on our ability to discriminate between π_1: $N(\underline{0},I)$ and π_2: $N((\delta,0,\ldots,0),I)$ when the prior probabilities of the two populations are equal. We have described in this chapter the methodology to be used in carrying out a sequence of sampling experiments to see if scale contamination has an effect on our ability to make correct classifications when the initial samples are small. If small sample scale contamination does have a harmful effect on our ability to discriminate, we have proposed several alternatives to Fisher's linear discriminant function. These alternatives will be evaluated along with Fisher's linear discriminant function in the simulation experiments. In addition, we shall see how the robust procedures perform when there is no contamination, i.e., $\alpha = \beta = 0$. In Chapters IV and V we present the results of the small sample studies. Chapter IV presents the results for $m = 1$; Chapter V presents the results for $m = 4$.

CHAPTER IV

SMALL SAMPLE RESULTS FOR THE UNIVARIATE STUDY

4.1. Introduction

In this chapter, we shall examine the results of the small sample simulation experiments for the univariate case. We wish to evaluate the performance of Fisher's linear discriminant function when the initial samples are scale contaminated and to compare its performance to that of the three types of alternatives described in Chapter III. Note that when m = 1, we have only eight Type I discriminant functions instead of sixteen. This is due to the fact that changing the value of s^2 in $D_s(x) = [(x - \frac{1}{2}(\bar{x}_1 + \bar{x}_2))'(\bar{x}_1 - \bar{x}_2)]/s^2$ multiples $D_s(x)$ by a constant and thus has no effect on the probability of misclassification, i.e., a Type I LDF with a = b = 0.05 in the Gnanadesikan Kettenring variance estimate gives the same probabilities of misclassification as the corresponding LDF with a = b = 0.10. Thus there are a total of eleven (instead of nineteen) robust linear discriminant functions to be considered in the univariate study.

As indicated in Chapter III, the following parameter combinations $(n,\delta,(\alpha,\beta),G(x))$ will be considered:

$$n = 12,\ 25$$

$$\delta = 1,\ 3$$

$$(\alpha,\beta) = (0,0.10),\ (0,0.25),\ (0.10,0.10),\ (0.25,0.25)$$

$$G(x) = N(\mu,9),\ N(\mu,25),\ N(\mu,100),\ \frac{1}{\pi}\frac{1}{(x-\mu)^2}$$

and

$$n = 12,\ 25$$

$$\delta = 1,\ 3$$

$$(\alpha,\beta) = (0,0),$$

giving a total of sixty-four contaminated parameter combinations to be considered, along with four uncontaminated parameter combinations. For each parameter combination, ten pairs of samples were generated, one from π_1^c: $(1-\alpha)\ N(0,1) + \alpha G(0,\theta)$, one from π_2^c: $(1-\beta)\ N(\delta,1) + \beta G(\delta,\theta)$. For each pair of samples, Fisher's LDF and the eleven robust LDFs were determined, along with the corresponding probabilities of misclassification. The twelve discriminant functions were evaluated by (a) comparing (within each parameter combination) the average probabilities of misclassification for each procedure, (b) performing analyses of variance on the contaminated probability of misclassification P^c, the relative increase in probability of misclassification due to using a linear discriminant function based on contaminated initial samples rather than the optimal discriminant function $(P^c-P)/P$, and on the two transformed variables $\ln(P^c/P)$ and $\ln((1-P^c)/P^c)$, (c) investigating the relative performance of the twelve procedures by looking at the rank of each procedure within each parameter combination.

From these comparisons we wish to see how well Fisher's linear discriminant function performs with contaminated initial samples, how other variables of interest (sample size, distance between the two underlying populations (δ), amount of contamination (α,β), and type of contamination $G(x)$) affect the probabilities of misclassification, and which of the eleven robust discriminant functions perform best when the initial samples are scale contaminated.

4.2. Average Probabilities of Misclassification for Each Procedure (m=1)

For each of the sixty-eight parameter combinations (sample size by distance between π_1 and π_2 by amount of contamination by type of contamination), the average probability of misclassification for each discriminant procedure was calculated. The averages are presented in Tables 4.2.1 and 4.2.2. Table 4.2.1 gives the results for $\delta = 1$; Table 4.2.2 gives the results for $\delta = 3$. The standard deviations of the probability of misclassification were also computed for each method and are presented in Tables 4.2.3 and 4.2.4. Each mean and standard deviation in Tables 4.2.1 through 4.2.4 is based on ten pairs of samples. Comparisons across methods are felt to be justified in that each pair of samples acts as a block, i.e., within each parameter combination, all methods are performed on the same ten pairs of samples. Care should be taken in making other comparisons (across n, δ, (α,β), or $G(x)$), however, since in this case each mean and standard deviation is based on different sets of samples.

In general, those methods which gave larger mean probabilities of misclassification also showed more variability from sample to sample. These methods which have relatively larger variances perform poorly with samples having extreme observations but perform well with more normal samples. Such methods are not recommended if we are looking for a robust procedure. Thus, conclusions drawn on the basis of the variability of the procedures will be the same as those based on the average probabilities of misclassification.

Based on Tables 4.2.1 through 4.2.4, the following conclusions can be drawn:

TABLE 4.2.1

AVERAGE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE (m=1, δ=1)
(OPTIMAL PROBABILITY = 0.3085)

			Discriminant Procedure											
G(x)	(α,β)	n	I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
--	(0,0)	12	0.311	0.312	0.311	0.311	0.311	0.311	0.311	0.311	0.311	0.311	0.312	0.311
		25	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.312
N(μ,9)	(0,0.10)	12	0.350	0.351	0.350	0.350	0.351	0.351	0.350	0.350	0.350	0.350	0.351	0.351
		25	0.310	0.310	0.310	0.310	0.310	0.310	0.311	0.311	0.310	0.310	0.310	0.310
N(μ,9)	(0,0.25)	12	0.320	0.319	0.320	0.321	0.321	0.322	0.320	0.319	0.358	0.321	0.319	0.360
		25	0.312	0.312	0.312	0.312	0.312	0.312	0.311	0.311	0.312	0.312	0.312	0.314
N(μ,9)	(0.10,0.10)	12	0.314	0.314	0.313	0.314	0.314	0.314	0.314	0.314	0.353	0.314	0.314	0.354
		25	0.311	0.311	0.311	0.310	0.310	0.310	0.311	0.312	0.311	0.311	0.311	0.312
N(μ,9)	(0.25,0.25)	12	0.318	0.315	0.318	0.318	0.318	0.317	0.316	0.317	0.318	0.322	0.315	0.327
		25	0.310	0.310	0.310	0.311	0.311	0.310	0.310	0.311	0.311	0.311	0.310	0.313
N(μ,9)	--	--	0.318	0.318	0.318	0.318	0.319	0.318	0.318	0.318	0.328	0.319	0.318	0.330
N(μ,25)	(0,0.10)	12	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.313	0.314
		25	0.311	0.311	0.311	0.311	0.311	0.311	0.312	0.311	0.311	0.311	0.311	0.312
N(μ,25)	(0,0.25)	12	0.313	0.312	0.313	0.313	0.312	0.311	0.313	0.313	0.352	0.316	0.312	0.355
		25	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.311	0.312	0.312	0.312	0.313
N(μ,25)	(0.10,0.10)	12	0.314	0.313	0.314	0.316	0.315	0.314	0.313	0.314	0.316	0.315	0.313	0.322
		25	0.314	0.315	0.314	0.314	0.312	0.312	0.315	0.316	0.316	0.314	0.315	0.322
N(μ,25)	(0.25,0.25)	12	0.318	0.318	0.318	0.318	0.317	0.317	0.318	0.318	0.395	0.320	0.318	0.399
		25	0.351	0.352	0.351	0.352	0.314	0.314	0.353	0.350	0.390	0.351	0.352	0.391
N(μ,25)	--	--	0.318	0.318	0.318	0.319	0.313	0.313	0.319	0.318	0.338	0.319	0.318	0.341
N(μ,100)	(0,0.10)	12	0.311	0.312	0.312	0.312	0.312	0.312	0.312	0.312	0.350	0.311	0.312	0.352
		25	0.310	0.311	0.310	0.310	0.310	0.311	0.311	0.311	0.425	0.310	0.311	0.425
N(μ,100)	(0,0.25)	12	0.318	0.310	0.312	0.312	0.311	0.311	0.310	0.316	0.314	0.324	0.310	0.331
		25	0.348	0.310	0.311	0.311	0.310	0.310	0.310	0.311	0.388	0.349	0.310	0.396
N(μ,100)	(0.10,0.10)	12	0.357	0.354	0.353	0.352	0.351	0.352	0.353	0.356	0.431	0.359	0.354	0.443
		25	0.311	0.311	0.311	0.311	0.310	0.310	0.311	0.311	0.312	0.311	0.311	0.322
N(μ,100)	(0.25,0.25)	12	0.368	0.371	0.369	0.371	0.407	0.373	0.370	0.369	0.438	0.405	0.371	0.434
		25	0.318	0.313	0.316	0.315	0.312	0.311	0.313	0.312	0.359	0.322	0.313	0.380
N(μ,100)	--	--	0.330	0.324	0.324	0.324	0.328	0.324	0.324	0.325	0.377	0.337	0.324	0.386
Cauchy	(0,0.10)	12	0.315	0.316	0.315	0.315	0.314	0.314	0.315	0.315	0.342	0.316	0.316	0.332
		25	0.309	0.309	0.309	0.309	0.309	0.309	0.310	0.310	0.309	0.309	0.309	0.310
Cauchy	(0,0.25)	12	0.352	0.354	0.316	0.316	0.316	0.317	0.354	0.353	0.348	0.315	0.354	0.334
		25	0.311	0.311	0.311	0.311	0.311	0.311	0.311	0.311	0.387	0.311	0.311	0.385
Cauchy	(0.10,0.10)	12	0.316	0.316	0.315	0.315	0.314	0.314	0.316	0.316	0.316	0.316	0.316	0.320
		25	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.313
Cauchy	(0.25,0.25)	12	0.314	0.314	0.314	0.314	0.315	0.315	0.314	0.314	0.372	0.353	0.314	0.392
		25	0.312	0.312	0.312	0.312	0.312	0.312	0.313	0.312	0.351	0.311	0.312	0.371
Cauchy	--	--	0.317	0.318	0.313	0.313	0.313	0.313	0.318	0.318	0.342	0.318	0.318	0.345
--	--	--	0.321	0.319	0.318	0.319	0.318	0.317	0.320	0.320	0.346	0.323	0.319	0.350

TABLE 4.2.2

AVERAGE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE (m=1, δ=3)
(OPTIMAL PROBABILITY = 0.0668)

			Discriminant Procedure											
			I	I	I	I	I	I	I	I	II	III	III	
G(x)	(α,β)	n	TRIM15	TRIM25	HUB	HAMHUB	HAMP	SIN	GAST	TRI	ROBREG	TLDF15	TLDF25	FISHER
—	(0,0)	12	0.070	0.070	0.070	0.070	0.070	0.071	0.070	0.070	0.070	0.070	0.070	0.070
		25	0.068	0.068	0.068	0.068	0.068	0.068	0.069	0.069	0.068	0.068	0.068	0.068
N(μ,9)	(0,0.10)	12	0.074	0.075	0.074	0.074	0.073	0.074	0.075	0.074	0.074	0.074	0.075	0.074
		25	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070
N(μ,9)	(0,0.25)	12	0.077	0.076	0.076	0.077	0.076	0.075	0.076	0.077	0.078	0.077	0.076	0.082
		25	0.070	0.071	0.070	0.070	0.071	0.071	0.072	0.072	0.070	0.070	0.071	0.070
N(μ,9)	(0.10,0.10)	12	0.070	0.071	0.070	0.070	0.070	0.071	0.071	0.070	0.071	0.071	0.071	0.071
		25	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.069	0.069	0.069	0.070	0.069
N(μ,9)	(0.25,0.25)	12	0.076	0.075	0.075	0.075	0.075	0.075	0.075	0.076	0.076	0.076	0.075	0.082
		25	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070
N(μ,9)	—	—	0.072	0.072	0.072	0.072	0.072	0.072	0.073	0.072	0.072	0.072	0.072	0.074
N(μ,25)	(0,0.10)	12	0.070	0.071	0.071	0.071	0.071	0.071	0.072	0.070	0.071	0.071	0.071	0.072
		25	0.071	0.071	0.070	0.070	0.070	0.070	0.072	0.070	0.072	0.071	0.071	0.074
N(μ,25)	(0,0.25)	12	0.078	0.079	0.078	0.079	0.078	0.077	0.079	0.077	0.080	0.080	0.079	0.087
		25	0.071	0.071	0.071	0.072	0.071	0.070	0.072	0.072	0.073	0.071	0.071	0.077
N(μ,25)	(0.10,0.10)	12	0.073	0.072	0.073	0.073	0.072	0.072	0.072	0.072	0.159	0.073	0.072	0.160
		25	0.072	0.073	0.072	0.072	0.072	0.073	0.072	0.073	0.071	0.072	0.073	0.070
N(μ,25)	(0.25,0.25)	12	0.080	0.075	0.080	0.077	0.075	0.075	0.078	0.078	0.166	0.085	0.075	0.180
		25	0.073	0.072	0.072	0.073	0.074	0.072	0.072	0.071	0.073	0.073	0.072	0.079
N(μ,25)	—	—	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.096	0.074	0.073	0.100
N(μ,100)	(0,0.10)	12	0.074	0.071	0.071	0.071	0.070	0.070	0.071	0.073	0.074	0.078	0.071	0.103
		25	0.071	0.071	0.070	0.070	0.069	0.069	0.071	0.072	0.072	0.070	0.071	0.085
N(μ,100)	(0,0.25)	12	0.092	0.080	0.085	0.084	0.081	0.077	0.081	0.088	0.089	0.106	0.080	0.113
		25	0.074	0.071	0.073	0.074	0.073	0.071	0.070	0.073	0.076	0.077	0.071	0.104
N(μ,100)	(0.10,0.10)	12	0.072	0.071	0.070	0.070	0.071	0.072	0.070	0.072	0.072	0.074	0.071	0.100
		25	0.069	0.069	0.070	0.070	0.070	0.070	0.069	0.070	0.070	0.070	0.069	0.086
N(μ,100)	(0.25,0.25)	12	0.114	0.090	0.110	0.100	0.081	0.075	0.099	0.105	0.200	0.146	0.090	0.268
		25	0.070	0.072	0.070	0.070	0.072	0.072	0.073	0.072	0.070	0.070	0.071	0.074
N(μ,100)	--	--	0.080	0.074	0.077	0.076	0.073	0.072	0.075	0.078	0.090	0.086	0.074	0.117
Cauchy	(0,0.10)	12	0.071	0.071	0.071	0.070	0.071	0.071	0.071	0.071	0.071	0.070	0.071	0.087
		25	0.068	0.068	0.068	0.068	0.068	0.068	0.068	0.069	0.068	0.068	0.068	0.068
Cauchy	(0,0.25)	12	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.072	0.156	0.072	0.072	0.118
		25	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.071	0.073	0.070	0.070	0.122
Cauchy	(0.10,0.10)	12	0.074	0.075	0.074	0.074	0.074	0.075	0.075	0.075	0.074	0.073	0.075	0.126
		25	0.070	0.070	0.070	0.070	0.069	0.069	0.070	0.070	0.155	0.070	0.070	0.133
Cauchy	(0.25,0.25)	12	0.079	0.076	0.076	0.076	0.075	0.075	0.077	0.076	0.084	0.099	0.076	0.122
		25	0.070	0.070	0.070	0.070	0.070	0.069	0.070	0.069	0.157	0.070	0.070	0.151
Cauchy	--	--	0.072	0.072	0.071	0.071	0.071	0.071	0.072	0.071	0.105	0.074	0.072	0.116
—	--	--	0.074	0.073	0.074	0.073	0.072	0.072	0.073	0.074	0.091	0.077	0.073	0.101

TABLE 4.2.3

STANDARD DEVIATIONS OF THE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE (m=1, δ=1)

			Discriminant Procedure											
G(x)	(α,β)	n	I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
--	(0,0)	12	0.006	0.006	0.006	0.005	0.005	0.005	0.006	0.005	0.004	0.005	0.006	0.004
		25	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
N(μ,9)	(0,0.10)	12	0.118	0.117	0.118	0.118	0.118	0.118	0.117	0.117	0.118	0.118	0.117	0.119
		25	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.001
N(μ,9)	(0,0.25)	12	0.015	0.013	0.015	0.016	0.018	0.019	0.014	0.014	0.118	0.016	0.013	0.117
		25	0.003	0.003	0.004	0.004	0.004	0.004	0.003	0.001	0.003	0.004	0.003	0.005
N(μ,9)	(0.10,0.10)	12	0.007	0.007	0.007	0.008	0.007	0.006	0.007	0.007	0.119	0.008	0.007	0.119
		25	0.003	0.003	0.002	0.002	0.002	0.002	0.003	0.004	0.002	0.002	0.003	0.003
N(μ,9)	(0.25,0.25)	12	0.011	0.008	0.010	0.011	0.010	0.010	0.008	0.009	0.011	0.016	0.008	0.021
		25	0.002	0.001	0.002	0.003	0.003	0.002	0.001	0.002	0.003	0.002	0.001	0.007
N(μ,25)	(0,0.10)	12	0.007	0.009	0.007	0.007	0.007	0.008	0.009	0.008	0.008	0.007	0.009	0.006
		25	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.004
N(μ,25)	(0,0.25)	12	0.006	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.118	0.009	0.004	0.109
		25	0.003	0.003	0.003	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.007
N(μ,25)	(0.10,0.10)	12	0.006	0.005	0.006	0.008	0.008	0.006	0.004	0.006	0.008	0.008	0.005	0.016
		25	0.006	0.006	0.006	0.006	0.003	0.004	0.007	0.008	0.008	0.006	0.006	0.014
N(μ,25)	(0.25,0.25)	12	0.022	0.022	0.021	0.020	0.019	0.020	0.022	0.022	0.156	0.023	0.022	0.151
		25	0.120	0.120	0.120	0.119	0.008	0.009	0.119	0.120	0.159	0.120	0.120	0.159
N(μ,100)	(0,0.10)	12	0.005	0.005	0.005	0.005	0.006	0.006	0.006	0.006	0.119	0.005	0.005	0.114
		25	0.002	0.003	0.002	0.002	0.002	0.003	0.003	0.003	0.183	0.002	0.003	0.169
N(μ,100)	(0,0.25)	12	0.024	0.002	0.004	0.006	0.003	0.003	0.002	0.021	0.014	0.037	0.002	0.051
		25	0.118	0.002	0.003	0.004	0.002	0.002	0.002	0.003	0.156	0.115	0.002	0.138
N(μ,100)	(0.10,0.10)	12	0.118	0.118	0.118	0.118	0.120	0.120	0.119	0.118	0.175	0.118	0.118	0.149
		25	0.003	0.004	0.004	0.004	0.004	0.003	0.004	0.003	0.005	0.004	0.003	0.013
N(μ,100)	(0.25,0.25)	12	0.112	0.113	0.112	0.112	0.148	0.115	0.113	0.113	0.159	0.147	0.113	0.140
		25	0.017	0.006	0.012	0.010	0.004	0.004	0.006	0.005	0.118	0.022	0.006	0.115
Cauchy	(0,0.10)	12	0.009	0.009	0.009	0.009	0.007	0.007	0.008	0.008	0.091	0.010	0.009	0.060
		25	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
Cauchy	(0,0.25)	12	0.116	0.117	0.011	0.011	0.011	0.013	0.117	0.117	0.106	0.010	0.117	0.059
		25	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.159	0.002	0.002	0.136
Cauchy	(0.10,0.10)	12	0.015	0.015	0.016	0.016	0.014	0.014	0.015	0.016	0.017	0.015	0.015	0.021
		25	0.001	0.002	0.001	0.002	0.001	0.001	0.002	0.002	0.002	0.001	0.002	0.010
Cauchy	(0.25,0.25)	12	0.005	0.005	0.005	0.005	0.005	0.005	0.004	0.004	0.127	0.119	0.005	0.120
		25	0.003	0.004	0.003	0.003	0.003	0.004	0.004	0.004	0.119	0.003	0.003	0.104

TABLE 4.2.4

STANDARD DEVIATIONS OF THE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE ($m=1$, $\delta=3$)

			Discriminant Procedure											
$G(x)$	(α,β)	n	I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I CAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
—	(0,0)	12	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.004	0.003
		25	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002
$N(\mu,9)$	(0,0.10)	12	0.011	0.013	0.011	0.011	0.010	0.011	0.014	0.012	0.012	0.011	0.013	0.012
		25	0.003	0.003	0.003	0.002	0.002	0.002	0.004	0.003	0.003	0.002	0.003	0.003
$N(\mu,9)$	(0,0.25)	12	0.012	0.012	0.012	0.011	0.010	0.010	0.011	0.011	0.011	0.011	0.012	0.014
		25	0.005	0.007	0.005	0.005	0.005	0.005	0.009	0.006	0.006	0.015	0.007	0.005
$N(\mu,9)$	(0.10,0.10)	12	0.005	0.006	0.005	0.004	0.005	0.006	0.006	0.005	0.005	0.006	0.006	0.006
		25	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.003
$N(\mu,9)$	(0.25,0.25)	12	0.018	0.014	0.018	0.018	0.018	0.016	0.014	0.017	0.017	0.019	0.014	0.026
		25	0.004	0.003	0.004	0.004	0.004	0.003	0.003	0.005	0.004	0.004	0.003	0.005
$N(\mu,25)$	(0,0.10)	12	0.002	0.005	0.004	0.004	0.004	0.005	0.005	0.003	0.004	0.003	0.005	0.005
		25	0.004	0.005	0.004	0.004	0.004	0.005	0.006	0.004	0.005	0.004	0.005	0.007
$N(\mu,25)$	(0,0.25)	12	0.007	0.009	0.007	0.007	0.010	0.010	0.008	0.007	0.008	0.010	0.009	0.015
		25	0.004	0.004	0.004	0.004	0.003	0.003	0.004	0.006	0.004	0.004	0.004	0.007
$N(\mu,25)$	(0.10,0.10)	12	0.006	0.007	0.006	0.005	0.006	0.007	0.007	0.006	0.268	0.006	0.007	0.258
		25	0.005	0.005	0.005	0.005	0.005	0.006	0.005	0.007	0.004	0.005	0.006	0.003
$N(\mu,25)$	(0.25,0.25)	12	0.015	0.007	0.015	0.009	0.010	0.008	0.010	0.010	0.267	0.028	0.007	0.260
		25	0.008	0.008	0.008	0.009	0.010	0.008	0.009	0.006	0.010	0.011	0.008	0.013
$N(\mu,100)$	(0,0.10)	12	0.010	0.005	0.005	0.004	0.006	0.006	0.005	0.008	0.007	0.020	0.005	0.037
		25	0.004	0.003	0.004	0.004	0.003	0.002	0.004	0.004	0.005	0.004	0.003	0.021
$N(\mu,100)$	(0,0.25)	12	0.031	0.013	0.021	0.022	0.026	0.023	0.015	0.026	0.021	0.046	0.013	0.040
		25	0.009	0.005	0.009	0.010	0.011	0.008	0.004	0.008	0.009	0.012	0.005	0.029
$N(\mu,100)$	(0.10,0.10)	12	0.009	0.007	0.005	0.005	0.008	0.008	0.006	0.008	0.007	0.015	0.007	0.039
		25	0.003	0.003	0.003	0.003	0.004	0.003	0.003	0.004	0.003	0.003	0.003	0.036
$N(\mu,100)$	(0.25,0.25)	12	0.074	0.037	0.076	0.051	0.020	0.010	0.059	0.062	0.263	0.100	0.037	0.235
		25	0.003	0.009	0.005	0.005	0.008	0.009	0.011	0.007	0.005	0.003	0.008	0.011
Cauchy	(0,0.10)	12	0.004	0.005	0.004	0.004	0.003	0.004	0.005	0.004	0.004	0.004	0.005	0.041
		25	0.001	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.002
Cauchy	(0,0.25)	12	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.264	0.009	0.009	0.135
		25	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.006	0.009	0.003	0.003	0.134
Cauchy	(0.10,0.10)	12	0.009	0.010	0.009	0.009	0.009	0.010	0.010	0.009	0.008	0.008	0.010	0.119
		25	0.003	0.002	0.003	0.003	0.002	0.002	0.002	0.003	0.268	0.003	0.002	0.149
Cauchy	(0.25,0.25)	12	0.015	0.009	0.010	0.010	0.011	0.010	0.011	0.010	0.028	0.072	0.009	0.127
		25	0.004	0.004	0.004	0.004	0.003	0.003	0.004	0.003	0.272	0.004	0.004	0.211

(1) When there is no contamination, i.e., $\alpha = \beta = 0$, all methods are equivalent. Thus, in the uncontaminated situation, any of the eleven robust procedures is as good as Fisher's LDF.

(2) For mild single sample contamination, e.g., 10% of $N(\mu,9)$ or $N(\mu,25)$, Fisher's LDF performs as well as the robust procedures. For samples of size 25, Fisher's LDF performs well for any amount of $N(\mu,9)$ contamination, i.e., $(\alpha,\beta) = (0,0.10), (0,0.25), (0.10,0.10), (0.25,0.25)$.

In general, however, Fisher's linear discriminant function does not perform as well as the robust procedures when the initial samples are scale contaminated.

(3) The Type II robust LDF, the Robust Regression discriminant function, does not perform well when the initial samples are scale contaminated. In general, when Fisher's LDF performs poorly, so does the Type II discriminant function.

(4) For mild contaminating distributions,e.g., $N(\mu,9)$,the other robust discriminant functions (Type I and Type III) are equivalent, i.e., they give essentially the same probabilities of misclassification.

(5) For moderate contaminating distributions, e.g., $N(\mu,25)$, most of the Type I and Type III procedures remain similar to each other. The Type I Hampel and Sin LDFs tend to perform better than the other procedures when $n = 25$.

(6) For heavier contamination, e.g., $N(\mu,100)$, a few other differences appear, especially for $n = 12$. The 15% Trimmed LDF tends not to perform as well as some of the other robust procedures. The Type I 15% Trimmed Mean LDF and the Type I Trimean LDF also

tend not to perform as well as the other procedures. The Type I Sin estimate again tends to perform better than the other methods in many cases.

(7) For Cauchy contamination, the Type I and Type III discriminant functions perform quite similarly to one another for most parameter combinations. Note, however, the following exceptions. For $(\alpha,\beta) = (0,0.25)$, $n = 12$, $\delta = 1$, the Type I Huber, Hampel Huber, Hampel and Sin LDFs perform better than the other procedures. For $(\alpha,\beta) = (0.25,0.25)$, $n = 12$, $\delta = 1$ and 3, the 15% Trimmed LDF does not perform as well as the other procedures.

(8) Differences among the methods decrease as the sample size increases, i.e., there are smaller differences among the methods when $n = 25$ than when $n = 12$. This is to be expected from the asymptotic results presented in Chapter II. When n goes to infinity, Fisher's LDF becomes optimal. With $n = 25$, there are still many situations, however, where Fisher's linear discriminant function does not perform well.

Overall, the Type I and Type III LDFs are very similar to one another. Fisher's linear discriminant function and the Type II Robust Regression discriminant function can perform very poorly when the initial samples are scale contaminated.

4.3. The Effects of Sample Size, Distance Between Populations, Amount of Contamination, Type of Contamination, and Procedure on the Probability of Misclassification (m=1)

To assess the effects the control variables (namely, sample size, distance between π_1 and π_2, amount of contamination, type of contamination, and procedure used for discrimination) have on the probability of misclassification, an analysis of variance was run on the simulated probabilities. The model assumed was

$$
\begin{aligned}
P^{c}_{ijk\ell mn} = \mu &+ n_i + d_j + c_k + g_\ell + m_m + r_{n(ijk\ell)} \\
&+ (nd)_{ij} + (nc)_{ik} + (ng)_{i\ell} + (nm)_{im} \\
&+ (dc)_{jk} + (dg)_{j\ell} + (dm)_{jm} \\
&+ (cg)_{k\ell} + (cm)_{km} + (gm)_{\ell m} + (mr)_{mn(ijk\ell)} \\
&+ \varepsilon_{ijk\ell mn}
\end{aligned}
$$

where

$P^{c}_{ijk\ell mn}$ is the probability of misclassification corresponding to the n^{th} replicate ($n = 1,\ldots,10$) for the i^{th} sample size ($i = 1,2$), j^{th} distance between populations ($j = 1,2$), k^{th} amount of contamination ($k = 1,2,3,4$), ℓ^{th} contaminating distribution ($\ell = 1,2,3,4$), m^{th} method ($m = 1,2,\ldots,12$)

μ is the overall mean probability of misclassification.

n_i is the effect of the i^{th} sample size ($i = 1$ for sample size 12; $i = 2$ for sample size 25).

d_j is the effect of the j^{th} distance between populations ($j = 1$ for $\delta = 1$; $j = 2$ for $\delta = 3$)

c_k is the effect of the k^{th} amount of contamination ($k = 1$ for $(\alpha,\beta) = (0,0.10)$; $k = 2$ for $(\alpha,\beta) = (0,0.25)$; $k = 3$ for $(\alpha,\beta) = (0.10,0.10)$; $k = 4$ for $(\alpha,\beta) = (0.25,0.25)$)

g_ℓ is the effect of the ℓ^{th} contaminating distribution (ℓ = 1 for N(0,9); ℓ = 2 for N(0,25); ℓ = 3 for N(0,100); ℓ = 4 for Cauchy)

m_m is the effect of the m^{th} discrimination procedure (m = 1 through m = 8 correspond to the Type I LDFs: m = 1 for the 15% Trimmed Mean; m = 2 for the 25% Trimmed Mean; m = 3 for the Huber LDF; m = 4 for Hampel's Huber; m = 5 for the Hampel LDF; m = 6 for the Type I Sin Estimate LDF; m = 7 for the Gastwirth LDF and m = 8 for the Trimean; m = 9 for the Type II Robust Regression discriminant function; m = 10 for the Type III 15% Trimmed LDF; m = 11 for the 25% Trimmed LDF and m = 12 for Fisher's LDF)

$r_{n(ijk\ell)}$ is the effect of the n^{th} replicate nested within the i^{th} sample size, j^{th} distance, k^{th} amount of contamination; ℓ^{th} contaminating distribution.

The other terms in the model correspond to the two factor interactions and an error term. Three factor and higher order interactions are assumed to be zero and are used to obtain the pooled error term.

Note that replicates (or blocks) are crossed with methods (each method is performed on each of the ten blocks) but are nested within the other factors. Sample size, distance, amount of contamination, and type of contamination are assumed to be fixed factors in the analysis; replicates are assumed to be random.

The results of the analysis are presented in Table 4.3.1. All main effects are significant. The probability of misclassification is

TABLE 4.3.1

ANALYSIS OF VARIANCE ON THE PROBABILITY OF MISCLASSIFICATION
(m=1)

Effect	F	df	p
Sample Size	14.64	(1, 609)	$p < .01$
Distance (δ)	8520.82	(1, 609)	$p < .01$
Amount of Contamination	4.68	(3, 609)	$p < .01$
Type of Contamination	3.90	(3, 609)	$p < .01$
Methods	39.47	(11, 6875)	$p < .01$
Sample Size × Distance	1.22	(1, 609)	$p > .10$
Sample Size × Amount of Contamination	0.86	(3, 609)	$p > .10$
Sample Size × Type of Contamination	1.92	(3, 609)	$p > .10$
Sample Size × Methods	1.80	(11, 6875)	$.05 < p < .10$
Distance × Amount of Contamination	1.74	(3, 609)	$p > .10$
Distance × Type of Contamination	2.55	(3, 609)	$.05 < p < .10$
Distance × Methods	5.26	(11, 6875)	$p < .01$
Amount × Type of Contamination	1.74	(9, 609)	$.05 < p < .10$
Amount of Contamination × Methods	2.73	(33, 6875)	$p < .01$
Type of Contamination × Methods	3.76	(33, 6875)	$p < .01$

lower with n = 25 than with n = 12; it is lower for $\delta = 3$ than for $\delta = 1$; as the amount of contamination increases, the average probability of misclassification increases; as the contamination becomes more severe, the average probability of misclassification increases; the twelve discriminant procedures differ significantly from one another.

Three of the two factor interactions are statistically significant: methods by distance, methods by amount of contamination, and methods by type of contamination. The methods by sample size interaction is only slightly above the five percent level of significance. Thus the relative performance of the twelve discrimination procedures differs for different levels of δ, for different amounts of contamination, for different types of contamination and to a lesser extent for different sample sizes. The presence of these interactions makes it difficult for us to make general statements about the comparison of methods without always qualifying our statements. We have observed this already in section 4.2 where our conclusions concerning the relative performance of the twelve discriminant procedures were dependent upon sample size, distance between populations, amount and type of contamination. The major difference is between the Fisher and the Robust Regression LDFs versus the Type I and Type III discriminant functions.

Table 4.3.2 gives the mean probability of misclassification for each method within each level of the main factors. The overall means and the means for each level of δ are given. The value of b given in parentheses in the left column is the number of blocks, i.e., the number of pairs of samples, involved in the mean. This table is based only on contaminated parameter combinations, i.e., $\alpha = \beta = 0$ is not included in the calculations.

TABLE 4.3.2

AVERAGE MARGINAL PROBABILITIES OF MISCLASSIFICATION

(m=1)

			Discriminant Procedure											
			I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
Overall		(b=640)	0.198	0.196	0.196	0.196	0.195	0.194	0.196	0.197	0.219	0.200	0.196	0.226
δ=1		(b=320)	0.321	0.319	0.318	0.319	0.318	0.317	0.320	0.320	0.346	0.323	0.319	0.350
δ=3		(b=320)	0.074	0.073	0.074	0.073	0.072	0.072	0.073	0.074	0.091	0.077	0.073	0.101
m=12	δ=1	(b=160)	0.326	0.325	0.323	0.323	0.325	0.323	0.325	0.326	0.354	0.329	0.325	0.357
m=25	δ=1	(b=160)	0.316	0.314	0.314	0.314	0.311	0.311	0.314	0.314	0.338	0.317	0.314	0.343
m=12	δ=3	(b=160)	0.078	0.075	0.077	0.076	0.074	0.074	0.076	0.077	0.100	0.083	0.075	0.115
m=25	δ=3	(b=160)	0.070	0.070	0.070	0.071	0.071	0.070	0.071	0.071	0.082	0.071	0.070	0.088
m=12		(b=320)	0.202	0.200	0.200	0.199	0.200	0.138	0.200	0.201	0.227	0.206	0.200	0.236
m=25		(b=320)	0.193	0.192	0.192	0.192	0.191	0.191	0.192	0.192	0.210	0.194	0.192	0.215
(α,β)=(0,0.10)	δ=1	(b=80)	0.316	0.317	0.316	0.316	0.317	0.317	0.317	0.316	0.339	0.316	0.317	0.338
(α,β)=(0,0.25)	δ=1	(b=80)	0.323	0.317	0.313	0.314	0.313	0.313	0.318	0.318	0.346	0.320	0.318	0.348
(α,β)=(0.10,0.10)	δ=1	(b=80)	0.318	0.318	0.318	0.318	0.317	0.317	0.318	0.319	0.333	0.319	0.318	0.339
(α,β)=(0.25,0.25)	δ=1	(b=80)	0.326	0.326	0.326	0.326	0.326	0.321	0.326	0.326	0.367	0.337	0.326	0.376
(α,β)=(0,0.10)	δ=3	(b=60)	0.071	0.071	0.071	0.071	0.070	0.070	0.071	0.071	0.072	0.071	0.071	0.079
(α,β)=(0,0.25)	δ=3	(b=80)	0.075	0.074	0.074	0.075	0.074	0.073	0.074	0.075	0.087	0.078	0.074	0.097
(α,β)=(0.10,0.10)	δ=3	(b=80)	0.071	0.071	0.071	0.071	0.071	0.072	0.071	0.071	0.093	0.071	0.071	0.102
(α,β)=(0.25,0.25)	δ=3	(b=80)	0.079	0.075	0.078	0.076	0.074	0.073	0.077	0.077	0.112	0.086	0.075	0.128
(α,β)=(0,0.10)		(b=160)	0.194	0.194	0.193	0.194	0.193	0.193	0.194	0.194	0.205	0.194	0.194	0.209
(α,β)=(0,0.25)		(b=160)	0.199	0.196	0.194	0.194	0.194	0.193	0.196	0.197	0.217	0.199	0.196	0.222
(α,β)=(0.10,0.10)		(b=160)	0.195	0.195	0.194	0.194	0.194	0.194	0.195	0.195	0.213	0.195	0.195	0.220
(α,β)=(0.25,0.25)		(b=160)	0.203	0.200	0.202	0.201	0.200	0.197	0.201	0.201	0.239	0.212	0.200	0.252
G(x)=N(μ,9)	δ=1	(b=80)	0.318	0.318	0.318	0.318	0.319	0.318	0.318	0.318	0.328	0.319	0.318	0.330
G(x)=N(μ,25)	δ=1	(b=80)	0.318	0.318	0.318	0.319	0.313	0.313	0.319	0.318	0.338	0.319	0.318	0.341
G(x)=N(μ,100)	δ=1	(b=80)	0.330	0.324	0.324	0.324	0.328	0.324	0.324	0.325	0.377	0.337	0.324	0.336
G(x)=Cauchy	δ=1	(b=80)	0.317	0.318	0.313	0.313	0.313	0.313	0.318	0.318	0.342	0.318	0.318	0.345
G(x)=N(μ,9)	δ=3	(b=80)	0.072	0.072	0.072	0.072	0.072	0.072	0.073	0.072	0.072	0.072	0.072	0.074
G(x)=N(μ,25)	δ=3	(b=80)	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.073	0.096	0.074	0.073	0.100
G(x)=N(μ,100)	δ=3	(b=80)	0.080	0.074	0.077	0.076	0.073	0.072	0.075	0.078	0.090	0.086	0.074	0.117
G(x)=Cauchy	δ=3	(b=80)	0.072	0.072	0.071	0.071	0.071	0.071	0.072	0.071	0.105	0.074	0.072	0.116
G(x)=N(μ,9)		(b=160)	0.195	0.195	0.195	0.195	0.195	0.195	0.195	0.195	0.200	0.195	0.195	0.202
G(x)=N(μ,25)		(b=160)	0.196	0.196	0.196	0.196	0.193	0.193	0.196	0.196	0.217	0.197	0.196	0.220
G(x)=N(μ,100)		(b=160)	0.205	0.199	0.201	0.200	0.201	0.198	0.200	0.201	0.234	0.211	0.199	0.251
G(x)=Cauchy		(b=160)	0.194	0.195	0.192	0.192	0.192	0.192	0.195	0.195	0.223	0.196	0.195	0.230

Overall, Fisher's LDF gives the largest probability of misclassification, i.e., Fisher's LDF does not perform as well as the robust discriminant functions when the initial samples are scale contaminated. Among the robust procedures, the Robust Regression LDF does not perform quite as well as the others. The Type I and Type III discriminant functions are very similar to one another. Among the ten Type I and Type III discriminant functions, the Sin LDF gives the smallest probability of misclassification and the 15% Trimmed LDF gives the largest probability. The difference between the two, however, is only 0.006.

The same analysis of variance model was used to analyze (1) the relative increase in probability of misclassification due to using a linear discriminant function based on contaminated initial samples rather than the optimal discriminant function, i.e., $(P^c-P)/P$, (2) $\ell n(P^c/P)$, and (3) $\ell n((1-P^c)/P^c)$. The results of these analyses were similar to those for P^c, with a few additional significant two factor interactions. All main factors were again significant ($p \leq 0.05$); all two factor interactions involving methods were significant; the sample size by distance and amount of contamination by type of contamination interactions were significant. In addition, for $\ell n((1-P^c)/P^c)$, the distance by amount of contamination and distance by type of contamination interactions were statistically significant.

Table 4.3.3 gives the marginal mean values of $(P^c-P)/P$ for each method within each level of the main factors. The general conclusions concerning the relative performance of the twelve discriminant procedures based on $(P^c-P)/P$ are the same as those based on P^c. Fisher's LDF may not perform well when the initial samples are scale contaminated. Overall, Fisher's LDF gives a 32% greater probability of misclassification

TABLE 4.3.3

AVERAGE MARGINAL VALUES OF $(P^c-P)/P$
$(m=1)$

			Discriminant Procedure											
			I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
Overall		(b=640)	0.073	0.060	0.064	0.062	0.055	0.050	0.064	0.068	0.238	0.095	0.059	0.324
$\delta=1$		(b=320)	0.039	0.034	0.030	0.031	0.030	0.026	0.034	0.035	0.121	0.045	0.034	0.133
$\delta=3$		(b=320)	0.105	0.085	0.097	0.092	0.081	0.074	0.094	0.101	0.355	0.145	0.085	0.515
$m=12$ $\delta=1$		(b=160)	0.054	0.052	0.045	0.046	0.052	0.045	0.052	0.054	0.146	0.066	0.052	0.157
$m=25$ $\delta=1$		(b=160)	0.024	0.015	0.016	0.016	0.007	0.007	0.016	0.016	0.095	0.025	0.015	0.110
$m=12$ $\delta=3$		(b=160)	0.161	0.119	0.143	0.131	0.109	0.099	0.132	0.144	0.488	0.235	0.119	0.722
$m=25$ $\delta=3$		(b=160)	0.052	0.051	0.051	0.053	0.053	0.048	0.056	0.057	0.222	0.055	0.051	0.307
$m=12$		(b=320)	0.107	0.086	0.094	0.089	0.080	0.072	0.092	0.099	0.317	0.150	0.086	0.440
$m=25$		(b=320)	0.038	0.033	0.034	0.034	0.030	0.028	0.036	0.036	0.159	0.040	0.033	0.208
$(\alpha,\beta)=(0,0.10)$	$\delta=1$	(b=80)	0.024	0.025	0.024	0.024	0.024	0.024	0.025	0.024	0.097	0.024	0.025	0.095
$(\alpha,\beta)=(0,0.25)$	$\delta=1$	(b=80)	0.046	0.027	0.014	0.015	0.014	0.013	0.028	0.030	0.121	0.036	0.027	0.127
$(\alpha,\beta)=(0.10,0.10)$	$\delta=1$	(b=80)	0.030	0.029	0.028	0.029	0.026	0.026	0.029	0.031	0.078	0.032	0.029	0.096
$(\alpha,\beta)=(0.25,0.25)$	$\delta=1$	(b=80)	0.056	0.054	0.055	0.056	0.054	0.040	0.055	0.054	0.187	0.090	0.054	0.216
$(\alpha,\beta)=(0,0.10)$	$\delta=3$	(b=80)	0.058	0.060	0.054	0.053	0.050	0.052	0.065	0.062	0.068	0.067	0.060	0.180
$(\alpha,\beta)=(0,0.25)$	$\delta=3$	(b=80)	0.127	0.100	0.112	0.115	0.105	0.090	0.104	0.124	0.296	0.163	0.100	0.440
$(\alpha,\beta)=(0.10,0.10)$	$\delta=3$	(b=80)	0.063	0.065	0.060	0.059	0.063	0.068	0.062	0.065	0.383	0.064	0.065	0.523
$(\alpha,\beta)=(0.25,0.25)$	$\delta=3$	(b=80)	0.170	0.116	0.164	0.141	0.105	0.086	0.145	0.152	0.673	0.285	0.116	0.915
$(\alpha,\beta)=(0,0.10)$		(b=160)	0.041	0.042	0.039	0.039	0.037	0.038	0.045	0.043	0.083	0.045	0.042	0.137
$(\alpha,\beta)=(0,0.25)$		(b=160)	0.086	0.064	0.063	0.065	0.059	0.052	0.066	0.077	0.209	0.099	0.064	0.284
$(\alpha,\beta)=(0.10,0.10)$		(b=160)	0.046	0.047	0.044	0.044	0.045	0.047	0.045	0.048	0.230	0.048	0.047	0.309
$(\alpha,\beta)=(0.25,0.25)$		(b=160)	0.117	0.085	0.109	0.099	0.079	0.063	0.100	0.103	0.430	0.188	0.085	0.566
$G(x)=N(\mu,9)$	$\delta=1$	(b=80)	0.030	0.029	0.030	0.030	0.031	0.030	0.029	0.029	0.061	0.032	0.029	0.068
$G(x)=N(\mu,25)$	$\delta=1$	(b=80)	0.030	0.030	0.030	0.031	0.014	0.013	0.031	0.030	0.094	0.033	0.030	0.103
$G(x)=N(\mu,100)$	$\delta=1$	(b=80)	0.069	0.048	0.049	0.050	0.061	0.048	0.048	0.051	0.221	0.089	0.048	0.248
$G(x)=$Cauchy	$\delta=1$	(b=80)	0.027	0.028	0.012	0.012	0.012	0.012	0.028	0.028	0.106	0.028	0.028	0.115
$G(x)=N(\mu,9)$	$\delta=3$	(b=80)	0.075	0.077	0.075	0.075	0.074	0.074	0.082	0.079	0.080	0.077	0.077	0.099
$G(x)=N(\mu,25)$	$\delta=3$	(b=80)	0.094	0.089	0.096	0.094	0.091	0.085	0.097	0.092	0.426	0.109	0.089	0.486
$G(x)=N(\mu,100)$	$\delta=3$	(b=80)	0.188	0.108	0.156	0.137	0.961	0.750	0.126	0.165	0.349	0.290	0.108	0.741
$G(x)=$Cauchy	$\delta=3$	(b=80)	0.068	0.068	0.063	0.062	0.062	0.061	0.071	0.066	0.565	0.103	0.068	0.733
$G(x)=N(\mu,9)$		(b=160)	0.052	0.053	0.052	0.053	0.052	0.052	0.056	0.054	0.071	0.054	0.053	0.084
$G(x)=N(\mu,25)$		(b=160)	0.062	0.059	0.063	0.062	0.053	0.049	0.064	0.061	0.260	0.071	0.059	0.295
$G(x)=N(\mu,100)$		(b=160)	0.123	0.078	0.102	0.093	0.079	0.061	0.087	0.108	0.285	0.189	0.078	0.494
$G(x)=$Cauchy		(b=160)	0.047	0.048	0.038	0.037	0.037	0.037	0.050	0.047	0.336	0.065	0.048	0.424

than the optimal discriminant procedure, as compared to the best of the robust discriminant functions, i.e., the Sin LDF, which gives a 5% greater probability of misclassification than the optimal procedure. The Type II LDF (the Robust Regression discriminant function) does not perform well compared to the other robust procedures. The Type III 15% Trimmed LDF also does not perform as well as the other procedures. (It is much better than the Robust Regression procedure, however.)

4.4. The Relative Ranking of the Twelve Discriminant Procedures (m=1)

Another approach to evaluating the relative performance of the twelve discriminant procedures when the initial samples are scale contaminated is to consider the relative rank of each procedure. Recall that within each parameter combination $(n,\delta,(\alpha,\beta),G(x))$ ten pairs of samples were drawn, one sample from π_1^c, the other from π_2^c. Each of the twelve discriminant procedures was applied to each of the ten pairs of samples and the corresponding probabilities of misclassification were calculated. In sections 4.2 and 4.3 we examined the values of these probabilities in order to evaluate the relative performance of the twelve discriminant procedures. In this section, we look at the relative ranking of each procedure rather than the actual numeric value of the probabilities of misclassification. For each block (pair of samples) within each parameter combination, the twelve procedures were assigned ranks from 1 to 12. If two or more procedures gave the same probability of misclassification, they were each assigned the average of the ranks that would have otherwise been assigned.

Tables 4.4.1 and 4.4.2 give, for each parameter combination, the average rank of each discrimination procedure. Table 4.4.3 is a summary

TABLE 4.4.1

AVERAGE RANK OF EACH DISCRIMINANT PROCEDURE BASED
ON THE PROBABILITY OF MISCLASSIFICATION
($m=1$, $\delta=1$)

			Discriminant Procedure											
G(x)	(α,β)	n	I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
—	(0,0)	12	6.5	7.6	6.5	6.2	6.1	6.8	7.4	6.4	5.8	5.3	7.6	5.6
		25	6.6	7.2		5.7	5.7	6.8	7.3	7.3	5.8	6.6	7.2	5.3
		—	6.5	7.4	6.7	5.9	5.9	6.8	7.3	6.9	5.8	5.9	7.4	5.4
N(μ,9)	(0,0.10)	12	4.9	6.2	5.6	7.3	8.2	7.4	6.2	4.1	7.0	6.3	6.2	8.9
		25	6.5	6.3	6.3	6.6	6.6	6.5	6.8	6.2	6.7	6.9	6.3	6.8
N(μ,9)	(0,0.25)	12	6.2	6.1	6.6	6.6	6.3	6.4	6.8	4.8	7.0	6.7	6.1	8.7
		25	6.7	5.7	6.8	8.0	6.9	6.0	5.0	5.4	7.4	6.4	5.7	8.3
N(μ,9)	(0.10,0.10)	12	6.6	6.3	5.4	5.7	6.9	6.4	6.9	5.7	7.5	5.7	6.3	8.8
		25	6.5	7.0	6.3	5.2	5.7	4.9	8.0	8.4	6.4	6.3	7.0	6.5
N(μ,9)	(0.25,0.25)	12	7.1	4.4	6.9	6.7	6.7	5.2	5.5	6.6	6.4	8.6	4.4	9.7
		25	6.2	5.8	6.4	6.7	6.3	5.8	6.3	8.0	7.1	6.3	5.8	7.7
N(μ,9)	—	--	6.3	5.9	6.3	6.6	6.7	6.1	6.4	6.1	6.9	6.6	5.9	8.1
N(μ,25)	(0,0.10)	12	6.5	6.7	5.6	6.2	6.8	5.7	6.5	5.9	6.8	6.4	6.7	8.3
		25	6.0	7.3	5.5	5.8	5.8	6.4	8.0	6.3	7.0	5.9	7.3	6.8
N(μ,25)	(0,0.25)	12	6.4	5.1	6.1	7.1	5.5	4.4	6.1	7.2	8.4	8.2	5.1	8.7
		25	6.8	6.5	6.2	6.9	6.2	5.8	6.4	6.4	6.7	6.7	6.7	6.8
N(μ,25)	(0.10,0.10)	12	6.1	4.6	7.1	8.0	7.4	6.6	5.5	5.9	6.8	7.8	4.6	7.8
		25	6.5	7.5	5.9	5.4	3.1	3.3	7.0	7.3	8.5	6.0	7.3	10.4
N(μ,25)	(0.25,0.25)	12	6.0	7.0	5.5	5.9	5.6	5.2	6.2	6.0	8.3	6.5	7.0	9.0
		25	6.0	6.9	5.6	6.2	5.2	5.2	8.2	4.6	8.5	6.7	6.8	8.3
N(μ,25)	--	--	6.3	6.4	5.9	6.4	5.7	5.3	6.7	6.2	7.6	6.8	6.4	8.2
N(μ,100)	(0,0.10)	12	4.2	5.5	6.3	7.7	8.0	7.3	6.6	6.9	7.1	5.1	5.5	8.0
		25	5.2	6.8	5.2	4.9	6.0	5.8	6.6	6.9	9.6	5.2	6.2	9.8
N(μ,100)	(0,0.25)	12	7.6	5.2	7.2	5.5	6.9	6.9	4.5	5.8	5.5	9.2	5.2	8.9
		25	7.3	4.9	6.1	6.5	6.0	5.4	4.4	6.6	8.3	8.4	5.4	9.1
N(μ,100)	(0.10,0.10)	12	6.2	6.7	6.1	5.3	6.7	6.8	5.7	5.5	8.3	5.7	6.7	8.7
		25	5.3	6.2	5.7	6.8	4.0	4.4	7.6	7.2	8.5	6.2	6.2	10.4
N(μ,100)	(0.25,0.25)	12	4.7	6.7	5.9	6.5	7.8	7.8	7.2	6.1	6.9	6.2	6.7	5.8
		25	7.7	5.6	6.7	7.3	4.7	3.9	5.1	4.9	9.1	8.3	6.2	8.9
N(μ,100)	--	--	6.0	5.9	6.1	6.3	6.2	6.0	5.9	6.2	7.9	6.8	6.0	8.7
Cauchy	(0,0.10)	12	6.2	8.3	6.5	6.4	5.2	5.7	6.5	5.4	6.4	7.1	8.3	6.2
		25	5.8	6.3	5.8	6.3	6.5	6.5	7.2	7.4	6.7	5.8	6.3	7.7
Cauchy	(0,0.25)	12	6.9	8.1	6.1	5.5	5.7	5.9	8.4	7.8	4.1	5.8	8.1	6.0
		25	5.1	6.0	6.1	6.8	7.0	6.2	6.5	4.9	8.8	6.1	6.0	8.9
Cauchy	(0.10,0.10)	12	7.6	7.5	5.7	5.3	4.3	4.7	6.8	7.6	8.2	6.1	7.5	7.1
		25	6.1	6.8	6.7	5.9	5.5	5.5	6.7	7.2	7.4	5.5	6.6	8.3
Cauchy	(0.25,0.25)	12	5.8	5.7	5.6	6.7	6.6	6.0	5.5	4.5	8.9	8.1	5.7	9.2
		25	5.2	5.9	5.1	5.4	7.0	7.1	6.7	7.2	7.2	5.0	5.9	10.5
Cauchy	--	--	6.1	6.8	5.9	6.0	5.9	5.9	6.8	6.5	7.2	6.2	6.8	7.9

TABLE 4.4.2

AVERAGE RANK OF EACH DISCRIMINANT PROCEDURE BASED ON THE PROBABILITY OF MISCLASSIFICATION

($m=1$, $\delta=3$)

			Discriminant Procedure											
			I	I	I	I	I	I	I	I	II	III	III	
G(x)	(α,β)	n	TRIM15	TRIM25	HUB	HAMHUB	HAMP	SIN	GAST	TRI	ROBREG	TLDF15	TLDF25	FISHER
--	(0,0)	12	5.8	7.0	5.8	6.8	7.0	8.1	6.0	5.2	4.7	7.4	7.0	7.5
		25	6.5	6.7	6.2	6.3	5.7	6.8	8.1	7.7	6.2	6.2	6.7	5.2
		--	6.1	6.9	6.0	6.5	6.3	7.4	7.1	6.4	5.4	6.8	6.9	6.3
N(μ,9)	(0,0.10)	12	6.1	6.8	5.7	6.8	6.5	5.7	7.9	6.6	7.2	6.0	6.8	6.2
		25	6.2	6.5	7.5	6.6	6.8	6.4	7.3	6.2	5.9	6.1	6.4	6.4
N(μ,9)	(0,0.25)	12	6.5	4.3	6.4	7.0	5.6	4.2	6.3	7.0	8.5	8.3	4.3	9.7
		25	5.6	6.7	5.9	6.3	7.0	7.3	8.5	8.5	5.3	5.3	6.5	5.5
N(μ,9)	(0.10,0.10)	12	5.8	6.8	5.5	5.7	5.9	7.8	7.9	6.1	6.9	6.1	6.8	7.0
		25	6.6	7.5	7.2	6.9	6.7	6.7	7.6	5.1	5.5	5.1	7.5	5.9
N(μ,9)	(0.25,0.25)	12	6.3	6.2	5.9	6.2	5.8	5.9	6.9	7.5	7.0	6.2	6.2	8.2
		25	7.0	6.1	7.0	7.2	6.9	6.1	6.5	7.1	5.5	6.6	5.9	6.5
N(μ,9)	--	--	6.2	6.3	6.3	6.6	6.4	6.2	7.3	6.7	6.4	6.2	6.3	6.9
N(μ,25)	(0,0.10)	12	5.4	6.0	5.6	6.7	7.1	6.4	7.2	5.7	7.8	6.4	6.0	8.0
		25	6.1	7.0	6.1	5.9	5.1	5.1	7.6	6.3	8.0	5.8	6.9	8.3
N(μ,25)	(0,0.25)	12	6.0	6.1	6.5	6.3	6.2	5.5	7.2	6.5	7.0	6.8	6.1	8.1
		25	5.4	5.9	5.9	7.5	5.4	4.3	7.2	7.5	8.9	5.8	5.9	8.4
N(μ,25)	(0.10,0.10)	12	6.1	5.5	6.3	6.9	6.5	6.5	4.9	5.8	8.0	6.5	5.5	9.8
		25	6.3	7.8	6.3	5.5	7.8	8.3	7.2	7.8	4.5	5.2	8.0	3.7
N(μ,25)	(0.25,0.25)	12	7.2	4.7	7.5	6.0	4.8	4.6	6.6	6.8	8.9	7.1	4.7	9.5
		25	6.5	5.3	5.9	7.7	6.7	4.4	6.3	5.8	7.5	7.3	5.3	9.5
N(μ,25)	--	--	6.1	6.0	6.2	6.5	6.2	5.6	6.8	6.5	7.6	6.3	6.0	8.1
N(μ,100)	(0,0.10)	12	6.6	7.2	5.5	5.4	4.5	5.2	6.4	7.7	7.6	6.5	7.2	8.5
		25	6.1	6.4	5.9	5.3	4.5	4.5	7.5	8.4	8.5	6.2	6.5	8.5
N(μ,100)	(0,0.25)	12	7.7	6.3	5.9	5.6	4.8	3.1	5.2	6.8	8.3	8.6	6.3	9.5
		25	7.3	4.0	6.5	7.6	5.3	4.2	3.9	7.7	8.1	9.2	4.4	10.2
N(μ,100)	(0.10,0.10)	12	6.3	6.6	5.8	5.8	6.6	7.5	5.6	5.8	7.4	5.4	6.6	8.8
		25	6.0	5.3	6.5	6.2	7.8	7.3	4.6	5.9	6.5	7.6	5.3	9.3
N(μ,100)	(0.25,0.25)	12	8.1	4.7	6.8	6.3	3.9	4.1	4.5	5.8	8.2	9.5	4.7	11.7
		25	6.2	6.6	5.4	5.9	6.2	6.6	7.6	7.2	5.6	6.4	6.2	8.5
N(μ,100)	--	--	6.8	5.9	6.0	6.0	5.4	5.3	5.6	6.9	7.5	7.4	5.9	9.4
Cauchy	(0,0.10)	12	5.7	6.7	6.6	5.9	5.8	6.8	7.6	6.4	6.4	5.9	6.7	7.9
		25	6.3	6.7	6.3	5.2	5.6	4.6	7.7	8.5	7.0	6.3	6.7	7.6
Cauchy	(0,0.25)	12	6.4	7.2	5.8	5.6	6.5	6.7	7.3	5.5	6.0	6.1	7.2	8.1
		25	5.7	5.5	5.8	6.9	7.7	6.1	6.3	6.8	6.8	5.8	5.5	9.3
Cauchy	(0.10,0.10)	12	5.5	7.7	5.9	5.4	6.1	6.7	8.1	6.9	6.5	4.2	7.7	7.7
		25	5.6	6.5	6.2	6.3	4.9	4.9	5.6	6.3	8.8	6.2	6.7	10.3
Cauchy	(0.25,0.25)	12	5.7	6.7	5.7	6.2	6.8	5.8	6.3	5.2	6.4	6.9	6.7	9.9
		25	6.2	6.2	6.0	6.2	6.0	5.3	7.1	5.7	8.1	6.5	6.0	8.9
Cauchy	--	--	5.9	6.6	6.0	5.9	6.2	5.8	7.0	6.4	7.0	6.0	6.6	8.7

TABLE 4.4.3

AVERAGE MARGINAL RANK OF EACH DISCRIMINANT PROCEDURE
(a=1)

		I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
		Discriminant Procedure											
Overall		6.2	6.2	6.1	6.3	6.1	5.8	6.6	6.4	7.3	6.5	6.2	8.3
$\delta=1$		6.2	6.3	6.1	6.3	6.1	5.8	6.5	6.2	7.4	6.6	6.3	8.2
$\delta=3$		6.2	6.2	6.1	6.3	6.0	5.7	6.7	6.6	7.1	6.5	6.2	8.3
$n=12$ $\delta=1$		6.2	6.2	6.1	6.4	6.5	6.1	6.3	6.0	7.1	6.8	6.2	8.1
$n=25$ $\delta=1$		6.2	6.3	6.0	6.3	5.8	5.5	6.6	6.5	7.7	6.3	6.3	8.4
$n=12$ $\delta=3$		6.3	6.2	6.1	6.1	5.8	5.8	6.6	6.4	7.4	6.6	6.3	8.6
$n=25$ $\delta=3$		6.2	6.2	6.2	6.4	6.2	5.7	6.8	6.9	6.9	6.3	6.2	7.9
$n=12$		6.2	6.2	6.1	6.3	6.2	5.9	6.4	6.2	7.2	6.7	6.2	8.4
$n=25$		6.2	6.3	6.1	6.4	6.0	5.6	6.7	6.7	7.3	6.3	6.3	8.2
$(\alpha,\beta)=(0,0.10)$	$\delta=1$	5.7	6.7	5.8	6.4	6.6	6.4	6.8	6.1	7.2	6.1	6.6	7.8
$(\alpha,\beta)=(0,0.25)$	$\delta=1$	6.6	5.9	6.4	6.6	6.3	5.8	6.0	6.1	7.0	7.2	6.0	8.1
$(\alpha,\beta)=(0.10,0.10)$	$\delta=1$	6.3	6.6	6.1	5.9	5.4	5.3	6.8	6.8	7.7	6.1	6.5	8.5
$(\alpha,\beta)=(0.25,0.25)$	$\delta=1$	6.1	6.0	5.9	6.4	6.2	5.7	6.3	6.0	7.8	6.9	6.1	8.6
$(\alpha,\beta)=(0,0.10)$	$\delta=3$	6.0	6.6	6.1	6.0	5.7	5.6	7.4	7.0	7.3	6.1	6.6	7.7
$(\alpha,\beta)=(0,0.25)$	$\delta=3$	6.3	5.7	6.1	6.6	6.1	5.2	6.5	7.0	7.3	7.0	5.8	8.6
$(\alpha,\beta)=(0.10,0.10)$	$\delta=3$	6.0	6.7	6.2	6.1	6.5	5.9	6.4	6.2	6.7	5.8	6.7	7.8
$(\alpha,\beta)=(0.25,0.25)$	$\delta=3$	6.6	5.8	6.2	6.4	5.8	5.2	6.4	6.4	7.1	7.0	5.7	9.1
$(\alpha,\beta)=(0,0.10)$		5.8	6.6	6.0	6.2	6.2	6.0	7.1	6.5	7.2	6.1	6.6	7.7
$(\alpha,\beta)=(0,0.25)$		6.5	5.8	6.2	6.6	6.2	5.5	6.2	6.6	7.2	7.1	5.9	8.4
$(\alpha,\beta)=(0.10,0.10)$		6.2	6.6	6.1	6.0	6.0	6.1	6.6	6.5	7.2	5.9	6.6	8.1
$(\alpha,\beta)=(0.25,0.25)$		6.3	5.9	6.1	6.4	6.0	5.5	6.4	6.2	7.5	7.0	5.9	8.8
$G(x)=N(\mu,9)$	$\delta=1$	6.3	5.9	6.3	6.6	6.7	6.1	6.4	6.1	6.9	6.6	5.9	8.1
$G(x)=N(\mu,25)$	$\delta=1$	6.3	6.4	5.9	6.4	5.7	5.3	6.7	6.2	7.6	6.8	6.4	8.2
$G(x)=N(\mu,100)$	$\delta=1$	6.0	5.9	5.1	6.3	6.2	6.0	5.9	6.2	7.9	6.8	6.0	8.7
$G(x)$=Cauchy	$\delta=1$	6.1	6.8	5.9	6.0	5.9	5.9	6.8	6.5	7.2	6.2	6.8	7.9
$G(x)=N(\mu,9)$	$\delta=3$	6.2	6.3	6.3	6.6	6.4	6.2	7.3	6.7	6.4	6.2	6.3	6.9
$G(x)=N(\mu,25)$	$\delta=3$	6.1	6.0	6.2	6.5	6.2	5.6	6.8	6.5	7.6	6.3	6.0	8.1
$G(x)=N(\mu,100)$	$\delta=3$	6.8	5.9	6.0	6.0	5.4	5.3	5.6	6.9	7.5	7.4	5.9	9.4
$G(x)$=Cauchy	$\delta=3$	5.9	6.6	6.0	5.9	6.2	5.8	7.0	6.4	7.0	6.0	6.6	8.7
$G(x)=N(\mu,9)$		6.3	6.1	6.3	6.6	6.5	6.1	6.9	6.4	6.7	6.4	6.1	7.5
$G(x)=N(\mu,25)$		6.2	6.2	6.1	6.5	5.9	5.5	6.7	6.3	7.6	6.5	6.2	8.2
$G(x)=N(\mu,100)$		6.4	5.9	6.1	6.1	5.8	5.7	5.8	6.5	7.7	7.1	5.9	9.0
$G(x)$=Cauchy		6.0	6.7	6.0	6.0	6.0	5.9	6.9	6.4	7.1	6.1	6.7	8.3

table giving the average rank for each level of the factors considered in section 4.3. The latter table is based only on contaminated parameter combinations, i.e., it does not include $\alpha = \beta = 0$.

Although the rank approach does give an indication of how the twelve procedures perform relative to one another, the results can be misleading. Suppose, for example, that for one particular block, method 1 gave a probability of misclassification of 0.068 while methods 2-12 gave probabilities of misclassification of 0.067. The ranks for methods 2-12 would be 6; the rank for method 1 would be 12. Thus a small difference in probability can be heavily weighted in the rank approach.

The general conclusions concerning the relative performance of the twelve discriminant procedures when the initial samples are scale contaminated are the same whether we consider the actual values of the probabilities of misclassification or the relative ranks. Based on the relative rank of each discriminant function, the Type I Sin LDF ranks best (5.8) followed by the Type I Huber and Hampel (6.1). Fisher's LDF and the Type II Robust Regression LDF are again the poorest procedures (ranks 8.3 and 7.3 respectively).

When there is no contamination, i.e., $\alpha = \beta = 0$, Fisher's LDF and the Type II LDF perform best. The Type I 25% Trimmed Mean LDF, the 25% Trimmed LDF, and the Type I Gastwirth LDF have the largest ranks. The Sin LDF has a large rank for $\delta = 3$. Although the rank approach suggests some systematic differences among the twelve procedures in the uncontaminated situation, the differences among the procedures are actually very small. Table 4.4.4 gives the actual probabilities of misclassification for each of the ten pairs of samples drawn within each

TABLE 4.4.4

PROBABILITIES OF MISCLASSIFICATION FOR INDIVIDUAL PAIRS OF SAMPLES WHEN THERE IS NO CONTAMINATION (m=1)

		Discriminant Procedure											
	Sample Number	I TRIM15	I TRIM25	I HUB	I HAMHUB	I HAMP	I SIN	I GAST	I TRI	II ROBREG	III TLDF15	III TLDF25	FISHER
	1	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	2	0.326	0.328	0.326	0.326	0.325	0.326	0.328	0.325	0.322	0.325	0.328	0.322
	3	0.310	0.310	0.310	0.309	0.309	0.309	0.310	0.310	0.309	0.309	0.310	0.309
	4	0.312	0.315	0.312	0.313	0.312	0.313	0.314	0.313	0.314	0.311	0.315	0.311
n=12	5	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
δ=1	6	0.309	0.309	0.309	0.310	0.310	0.310	0.309	0.309	0.309	0.309	0.309	0.309
	7	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	8	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310	0.310
	9	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	10	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.310
	1	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	2	0.321	0.319	0.322	0.322	0.322	0.321	0.318	0.316	0.320	0.321	0.319	0.322
	3	0.318	0.319	0.318	0.317	0.317	0.318	0.319	0.322	0.318	0.318	0.319	0.315
	4	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.310
n=25	5	0.312	0.311	0.312	0.311	0.311	0.312	0.310	0.311	0.310	0.312	0.311	0.310
δ=1	6	0.313	0.315	0.313	0.312	0.312	0.312	0.316	0.313	0.314	0.313	0.315	0.312
	7	0.309	0.310	0.309	0.309	0.309	0.310	0.310	0.310	0.309	0.309	0.310	0.309
	8	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	9	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309
	10	0.313	0.313	0.313	0.313	0.313	0.313	0.314	0.314	0.313	0.313	0.313	0.312
	1	0.075	0.077	0.075	0.076	0.075	0.076	0.077	0.077	0.076	0.074	0.077	0.073
	2	0.073	0.073	0.074	0.073	0.073	0.073	0.073	0.071	0.072	0.074	0.073	0.074
	3	0.070	0.070	0.070	0.070	0.070	0.070	0.069	0.068	0.068	0.070	0.070	0.070
	4	0.074	0.074	0.074	0.073	0.074	0.076	0.074	0.076	0.073	0.073	0.074	0.070
n=12	5	0.067	0.068	0.067	0.068	0.068	0.068	0.067	0.067	0.068	0.068	0.068	0.067
δ=3	6	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.068	0.068	0.067	0.067	0.068
	7	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067
	8	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.068
	9	0.074	0.074	0.073	0.074	0.074	0.075	0.073	0.073	0.073	0.074	0.074	0.075
	10	0.067	0.067	0.067	0.068	0.068	0.068	0.067	0.067	0.067	0.068	0.067	0.068
	1	0.069	0.069	0.069	0.069	0.069	0.069	0.069	0.068	0.069	0.069	0.069	0.069
	2	0.067	0.067	0.067	0.067	0.067	0.068	0.067	0.067	0.067	0.067	0.067	0.067
	3	0.069	0.069	0.069	0.069	0.069	0.069	0.070	0.070	0.069	0.069	0.069	0.068
	4	0.067	0.067	0.067	0.068	0.068	0.067	0.067	0.067	0.067	0.067	0.067	0.068
n=25	5	0.074	0.073	0.073	0.072	0.072	0.072	0.073	0.076	0.072	0.073	0.073	0.072
δ=3	6	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.067
	7	0.067	0.068	0.067	0.067	0.067	0.068	0.069	0.067	0.068	0.067	0.068	0.067
	8	0.068	0.068	0.068	0.068	0.067	0.068	0.068	0.068	0.068	0.068	0.068	0.067
	9	0.067	0.067	0.067	0.067	0.067	0.067	0.068	0.068	0.067	0.067	0.067	0.067
	10	0.067	0.067	0.067	0.067	0.067	0.067	0.067	0.068	0.067	0.067	0.067	0.067

uncontaminated parameter combination. The results indicate that there is very little difference among the twelve procedures when $\alpha = \beta = 0$.

4.5. Summary

In this chapter, we have examined the results of the small sample simulation experiments for the univariate case. We have seen that when there is no contamination in the initial samples, i.e., $\alpha = \beta = 0$, all twelve procedures (eight Type I LDFs, one Type II LDF, two Type III LDFs and Fisher's LDF) perform equally well. For mild contamination, Fisher's LDF performs as well as the robust procedures. In general, however, Fisher's linear discriminant function does not perform as well as the robust procedures when the initial samples are scale contaminated. Among the robust discriminant functions considered in this study, the Type II Robust Regression LDF does not perform well. When Fisher's LDF performs poorly, the Type II LDF also performs poorly. Among the Type I robust discriminant functions, the M-estimate LDFs seem to do best, although all the Type I and Type III discriminant functions are remarkably similar to each other. The Type I Sin LDF is overall the best procedure. The 15% Trimmed LDF and the Type I Trimean LDF do not tend to perform quite as well as the other Type I and Type III discriminant functions.

It should be noted that differences among the methods decrease as the sample size increases. Fisher's discriminant function performs better when n = 25 than when n = 12. This supports the asymptotic results derived in Chapter II: as n becomes large, Fisher's LDF becomes optimal.

In the next chapter, we shall investigate the results of the small sample simulation experiments for the four variable case.

CHAPTER V

SMALL SAMPLE RESULTS FOR THE FOUR VARIABLE STUDY

5.1. Introduction

In this chapter, we continue our study of the small sample performance of Fisher's linear discriminant function and of the three types of discriminant functions proposed as alternatives to Fisher's LDF when the initial samples are scale contaminated. We turn to the case where we have four independent variables $\underline{x} = (x_1, x_2, x_3, x_4)$ to be used as a basis for classifying individuals into one of the two populations π_1: $N(\underline{0}, I)$ or π_2: $N((\delta,0,0,0), I)$. There are a total of twenty discriminant functions to be considered in the four variable model: Fisher's LDF, eight Type I LDFs with a 5% Gnanadesikan-Kettenring variance estimate, eight Type I LDFs with a 10% Gnanadesikan-Kettenring variance estimate, the Type II Robust Regression LDF, and the two Type III LDFs.

The same parameter combinations $(n, \delta, (\alpha, \beta), G(x))$ considered in the univariate case will be considered here:

$$n = 12,\ 25$$

$$\delta = 1,\ 3$$

$$(\alpha, \beta) = (0, 0.10),\ (0, 0.25),\ (0.10, 0.10),\ (0.25, 0.25)$$

$$G(\underline{x}) = N(\underline{\mu}, 9I),\ N(\underline{\mu}, 25I),\ N(\underline{\mu}, 100I),\ \text{Cauchy}$$

and

$n = 12,\ 25$

$\delta = 1,\ 3$

$(\alpha,\beta) = (0,0)$.

The fourth contaminating distribution, $G(\underline{x})$ = Cauchy, refers to the distribution $G(x_i) = \frac{1}{(x-\mu_i)^2}$, $i = 1,2,3,4$; $\text{cov}(x_i,x_j) = 0 \quad i \neq j$. For each of the sixty-eight parameter combinations, ten pairs of samples were generated, one from π_1^c: $(1-\alpha)\ N(\underline{0},I) + \alpha G(\underline{0},\underline{\theta})$, one from π_2^c: $(1-\beta)\ N((\delta,0,0,0),I) + \beta G((\delta,0,0,0),\underline{\theta})$. For each pair of samples, all LDFs were determined, along with the corresponding probabilities of misclassification. The twenty discriminant functions were evaluated by comparing average probabilities of misclassification for each procedure, by performing several analyses of variance, and by investigating the relative rank of each procedure.

In the univariate case, we found that Fisher's discriminant function can perform very poorly when the initial samples are scale contaminated. This is also true for m = 4. As in the univariate case, we will see in section 5.3 that sample size, distance between the two underlying populations, amount of contamination, and type of contamination all affect the probabilities of misclassification. Finally, in comparing the twenty discriminant procedures, we shall see that as in the univariate study, the Robust Regression LDF may perform very poorly. The Type I LDFs will prove to be quite similar to one another. The Type III LDF will be seen not to perform as well for m = 4 as it did for m = 1.

5.2. Average Probability of Misclassification for Each Procedure (m=4)

Tables 5.2.1 and 5.2.2 give the average probability of misclassification for each discriminant procedure within each parameter combination. Table 5.2.1 gives the results for $\delta = 1$; Table 5.2.2 gives the results for $\delta = 3$. The standard deviations of the probability of misclassification are presented in Tables 5.2.3 and 5.2.4.

As in the univeriate case, those methods which gave larger mean probabilities of misclassification generally gave larger standard deviations.

Based on Tables 5.2.1 and 5.2.2, the following conclusions can be drawn.

(1) Consider first the uncontaminated situation, i.e., $\alpha = \beta = 0$. When n = 25, the twenty procedures are very similar to one another. The robust procedures perform as well as Fisher's LDF. When n = 12 and $\delta = 1$, the Type I and Type III discriminant functions give probabilities of misclassification 0.01 to 0.02 higher than the Fisher and Robust Regression LDFs. The Type I Sin LDF gives the highest probability of misclassification (0.375 for the 5% Sin LDF, 0.372 for the 10% Sin LDF), as compared to Fisher's LDF (0.357). The best of the Type I discriminant functions in this case is the 15% Trimmed Mean LDF (0.363, 0.364). When n = 12 and $\delta = 3$, the Fisher LDF and the Type I LDFs give essentially the same probabilities of misclassification. The Type II and Type III LDFs give slightly higher probabilities.

(2) For single sample $N(\underline{\mu}, 9I)$ contamination, Fisher's LDF performs

TABLE 5.2.1

AVERAGE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE ($m=4$, $\delta=1$)
(OPTIMAL PROBABILITY = 0.3085)

			Discriminant Procedure							
G(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
—	(0,0)	12	0.363	0.366	0.365	0.368	0.369	0.375	0.365	0.365
		25	0.337	0.338	0.337	0.336	0.336	0.336	0.340	0.336
N(μ,9I)	(0,0.10)	12	0.415	0.410	0.420	0.418	0.420	0.416	0.421	0.422
		25	0.342	0.345	0.342	0.343	0.345	0.345	0.346	0.346
N(μ,9I)	(0,0.25)	12	0.374	0.373	0.373	0.373	0.371	0.369	0.377	0.375
		25	0.336	0.337	0.336	0.336	0.337	0.336	0.338	0.338
N(μ,9I)	(0.10,0.10)	12	0.370	0.370	0.369	0.371	0.371	0.371	0.370	0.368
		25	0.357	0.355	0.356	0.357	0.356	0.355	0.356	0.361
N(μ,9I)	(0.25,0.25)	12	0.382	0.372	0.380	0.377	0.372	0.367	0.384	0.378
		25	0.347	0.346	0.348	0.351	0.346	0.343	0.344	0.348
N(μ,9I)	--	--	0.365	0.363	0.366	0.366	0.365	0.363	0.367	0.367
N(μ,25I)	(0,0.10)	12	0.385	0.384	0.386	0.391	0.390	0.390	0.383	0.385
		25	0.348	0.353	0.348	0.347	0.349	0.349	0.356	0.356
N(μ,25I)	(0,0.25)	12	0.415	0.412	0.414	0.417	0.409	0.408	0.410	0.411
		25	0.340	0.341	0.340	0.345	0.349	0.344	0.340	0.346
N(μ,25I)	(0.10,0.10)	12	0.371	0.377	0.370	0.371	0.370	0.372	0.375	0.371
		25	0.346	0.346	0.347	0.348	0.345	0.344	0.343	0.344
N(μ,25I)	(0.25,0.25)	12	0.409	0.410	0.412	0.408	0.408	0.406	0.414	0.409
		25	0.354	0.359	0.355	0.355	0.358	0.362	0.361	0.356
N(μ,25I)	--	--	0.371	0.373	0.371	0.373	0.372	0.372	0.373	0.372
N(μ,100I)	(0,0.10)	12	0.378	0.380	0.376	0.378	0.385	0.383	0.376	0.374
		25	0.340	0.341	0.340	0.339	0.335	0.335	0.340	0.342
N(μ,100I)	(0,0.25)	12	0.387	0.398	0.390	0.397	0.411	0.407	0.394	0.391
		25	0.366	0.361	0.363	0.364	0.359	0.360	0.364	0.371
N(μ,100I)	(0.10,0.10)	12	0.382	0.391	0.380	0.380	0.383	0.377	0.381	0.382
		25	0.378	0.373	0.378	0.377	0.365	0.365	0.375	0.380
N(μ,100I)	(0.25,0.25)	12	0.444	0.446	0.446	0.441	0.446	0.438	0.443	0.455
		25	0.380	0.373	0.378	0.375	0.376	0.373	0.370	0.376
N(μ,100I)	--	--	0.382	0.383	0.381	0.382	0.382	0.380	0.380	0.384
Cauchy	(0,0.10)	12	0.363	0.367	0.362	0.361	0.357	0.360	0.366	0.366
		25	0.340	0.340	0.339	0.339	0.339	0.340	0.340	0.337
Cauchy	(0,0.25)	12	0.374	0.378	0.372	0.370	0.374	0.373	0.381	0.376
		25	0.335	0.337	0.335	0.335	0.336	0.336	0.341	0.336
Cauchy	(0.10,0.10)	12	0.387	0.390	0.381	0.382	0.370	0.369	0.381	0.381
		25	0.332	0.333	0.332	0.332	0.332	0.331	0.333	0.333
Cauchy	(0.25,0.25)	12	0.398	0.398	0.398	0.397	0.391	0.391	0.399	0.399
		25	0.354	0.357	0.353	0.353	0.349	0.348	0.356	0.350
Cauchy	--	—	0.360	0.362	0.359	0.359	0.356	0.356	0.362	0.360
--	--	--	0.370	0.370	0.369	0.370	0.369	0.368	0.371	0.371

TABLE 5.2.1 (Continued)

			Discriminant Procedure											
$G(\underline{x})$	(α,β)	n	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
--	(0,0)	12	0.364	0.365	0.366	0.369	0.369	0.372	0.368	0.368	0.354	0.365	0.362	0.357
		25	0.339	0.340	0.339	0.337	0.337	0.337	0.343	0.338	0.340	0.344	0.343	0.340
$N(\underline{\mu},9I)$	(0,0.10)	12	0.419	0.420	0.422	0.422	0.425	0.423	0.424	0.424	0.423	0.423	0.430	0.418
		25	0.344	0.346	0.344	0.345	0.348	0.348	0.348	0.348	0.331	0.335	0.333	0.331
$N(\underline{\mu},9I)$	(0,0.25)	12	0.377	0.377	0.375	0.373	0.370	0.371	0.381	0.379	0.371	0.374	0.368	0.371
		25	0.336	0.336	0.336	0.336	0.337	0.337	0.339	0.339	0.343	0.329	0.330	0.340
$N(\underline{\mu},9I)$	(0.10,0.10)	12	0.365	0.367	0.370	0.371	0.367	0.368	0.369	0.366	0.385	0.383	0.377	0.379
		25	0.353	0.353	0.352	0.352	0.350	0.349	0.352	0.358	0.385	0.372	0.383	0.386
$N(\underline{\mu},9I)$	(0.25,0.25)	12	0.369	0.357	0.367	0.363	0.358	0.354	0.367	0.363	0.419	0.422	0.423	0.420
		25	0.349	0.345	0.350	0.352	0.348	0.345	0.344	0.354	0.369	0.357	0.357	0.367
$N(\underline{\mu},9I)$	--	--	0.364	0.363	0.365	0.364	0.363	0.362	0.365	0.367	0.378	0.374	0.375	0.376
$N(\underline{\mu},25I)$	(0,0.10)	12	0.379	0.376	0.379	0.381	0.378	0.376	0.375	0.378	0.387	0.391	0.387	0.385
		25	0.340	0.344	0.340	0.340	0.337	0.336	0.347	0.344	0.361	0.350	0.356	0.359
$N(\underline{\mu},25I)$	(0,0.25)	12	0.397	0.401	0.398	0.399	0.396	0.395	0.399	0.395	0.421	0.409	0.424	0.419
		25	0.339	0.342	0.341	0.348	0.351	0.348	0.344	0.346	0.361	0.355	0.357	0.365
$N(\underline{\mu},25I)$	(0.10,0.10)	12	0.378	0.381	0.376	0.376	0.374	0.377	0.380	0.376	0.394	0.379	0.391	0.388
		25	0.342	0.342	0.343	0.344	0.338	0.336	0.338	0.338	0.382	0.362	0.372	0.375
$N(\underline{\mu},25I)$	(0.25,0.25)	12	0.419	0.415	0.420	0.416	0.403	0.401	0.419	0.416	0.447	0.440	0.435	0.445
		25	0.355	0.360	0.356	0.356	0.361	0.363	0.363	0.355	0.341	0.341	0.338	0.347
$N(\underline{\mu},25I)$	--	--	0.369	0.370	0.369	0.370	0.367	0.367	0.379	0.379	0.382	0.380	0.382	0.378
$N(\underline{\mu},100I)$	(0,0.10)	12	0.377	0.376	0.369	0.376	0.376	0.376	0.371	0.370	0.380	0.376	0.375	0.380
		25	0.338	0.339	0.338	0.337	0.332	0.333	0.339	0.338	0.372	0.362	0.367	0.369
$N(\underline{\mu},100I)$	(0,0.25)	12	0.386	0.393	0.387	0.390	0.395	0.393	0.387	0.387	0.397	0.399	0.397	0.397
		25	0.384	0.367	0.376	0.375	0.358	0.360	0.370	0.387	0.417	0.394	0.378	0.419
$N(\underline{\mu},100I)$	(0.10,0.10)	12	0.380	0.396	0.382	0.379	0.393	0.388	0.388	0.384	0.425	0.403	0.419	0.419
		25	0.366	0.364	0.365	0.369	0.351	0.352	0.366	0.370	0.458	0.407	0.399	0.456
$N(\underline{\mu},100I)$	(0.25,0.25)	12	0.439	0.441	0.456	0.454	0.463	0.453	0.443	0.454	0.467	0.435	0.407	0.458
		25	0.361	0.354	0.362	0.362	0.356	0.351	0.349	0.359	0.401	0.387	0.376	0.413
$N(\underline{\mu},100I)$	--	--	0.379	0.379	0.379	0.380	0.378	0.376	0.377	0.381	0.415	0.395	0.390	0.414
Cauchy	(0,0.10)	12	0.364	0.369	0.360	0.360	0.357	0.360	0.368	0.367	0.374	0.360	0.367	0.368
		25	0.337	0.339	0.336	0.337	0.336	0.336	0.341	0.336	0.373	0.352	0.365	0.368
Cauchy	(0,0.25)	12	0.367	0.370	0.369	0.365	0.371	0.372	0.375	0.370	0.397	0.404	0.380	0.392
		25	0.335	0.335	0.335	0.336	0.336	0.335	0.337	0.336	0.401	0.369	0.357	0.404
Cauchy	(0.10,0.10)	12	0.375	0.386	0.380	0.379	0.372	0.374	0.380	0.377	0.417	0.400	0.408	0.409
		25	0.332	0.332	0.331	0.332	0.332	0.331	0.333	0.331	0.341	0.333	0.334	0.342
Cauchy	(0.25,0.25)	12	0.397	0.396	0.397	0.395	0.391	0.390	0.398	0.396	0.434	0.420	0.419	0.433
		25	0.347	0.347	0.346	0.346	0.339	0.338	0.347	0.345	0.410	0.349	0.358	0.411
Cauchy	--	--	0.357	0.359	0.357	0.356	0.354	0.354	0.360	0.357	0.393	0.373	0.374	0.391
--	--	--	0.367	0.368	0.367	0.368	0.366	0.365	0.368	0.368	0.393	0.380	0.380	0.392

TABLE 5.2.2

AVERAGE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE ($m=4$, $\delta=3$)
(OPTIMAL PROBABILITY = 0.0668)

			Discriminant Procedure							
G(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
--	(0,0)	12	0.096	0.096	0.096	0.096	0.096	0.097	0.095	0.095
		25	0.083	0.082	0.083	0.083	0.084	0.084	0.082	0.082
N(μ,9I)	(0,0.10)	12	0.089	0.091	0.090	0.090	0.090	0.089	0.091	0.089
		25	0.081	0.081	0.081	0.081	0.081	0.082	0.081	0.082
N(μ,9I)	(0,0.25)	12	0.109	0.106	0.110	0.110	0.109	0.110	0.107	0.108
		25	0.079	0.079	0.078	0.078	0.079	0.079	0.078	0.079
N(μ,9I)	(0.10,0.10)	12	0.098	0.100	0.098	0.098	0.097	0.098	0.098	0.097
		25	0.079	0.080	0.079	0.079	0.079	0.080	0.081	0.080
N(μ,9I)	(0.25,0.25)	12	0.104	0.102	0.102	0.105	0.105	0.104	0.103	0.103
		25	0.088	0.089	0.089	0.089	0.090	0.090	0.090	0.091
N(μ,9I)	--	--	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091
N(μ,25I)	(0,0.10)	12	0.090	0.090	0.089	0.090	0.090	0.090	0.090	0.089
		25	0.076	0.076	0.077	0.076	0.076	0.076	0.075	0.075
N(μ,25I)	(0,0.25)	12	0.108	0.111	0.107	0.107	0.108	0.110	0.110	0.110
		25	0.096	0.098	0.096	0.096	0.099	0.098	0.101	0.097
N(μ,25I)	(0.10,0.10)	12	0.109	0.106	0.103	0.102	0.098	0.099	0.105	0.103
		25	0.089	0.089	0.089	0.090	0.088	0.088	0.090	0.090
N(μ,25I)	(0.25,0.25)	12	0.123	0.118	0.124	0.125	0.123	0.123	0.119	0.123
		25	0.119	0.116	0.119	0.122	0.129	0.130	0.116	0.116
N(μ,25I)	--	--	0.101	0.101	0.101	0.101	0.101	0.101	0.101	0.100
N(μ,100I)	(0,0.10)	12	0.118	0.117	0.118	0.119	0.119	0.119	0.117	0.116
		25	0.098	0.098	0.098	0.099	0.101	0.099	0.098	0.098
N(μ,100I)	(0,0.25)	12	0.150	0.149	0.151	0.150	0.157	0.155	0.150	0.151
		25	0.109	0.107	0.107	0.107	0.108	0.107	0.108	0.113
N(μ,100I)	(0.10,0.10)	12	0.116	0.118	0.115	0.115	0.117	0.117	0.118	0.115
		25	0.090	0.090	0.089	0.090	0.089	0.088	0.090	0.091
N(μ,100I)	(0.25,0.25)	12	0.251	0.230	0.249	0.242	0.240	0.230	0.243	0.248
		25	0.129	0.127	0.128	0.128	0.131	0.129	0.126	0.129
N(μ,100I)	--	--	0.133	0.129	0.132	0.131	0.133	0.131	0.131	0.133
Cauchy	(0,0.10)	12	0.101	0.105	0.101	0.100	0.099	0.101	0.103	0.100
		25	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.081
Cauchy	(0,0.25)	12	0.125	0.126	0.125	0.125	0.124	0.124	0.125	0.125
		25	0.077	0.077	0.077	0.078	0.080	0.078	0.078	0.077
Cauchy	(0.10,0.10)	12	0.096	0.098	0.096	0.095	0.095	0.096	0.099	0.098
		25	0.084	0.085	0.084	0.084	0.085	0.085	0.085	0.085
Cauchy	(0.25,0.25)	12	0.133	0.131	0.136	0.139	0.141	0.140	0.132	0.133
		25	0.086	0.086	0.086	0.085	0.085	0.085	0.087	0.086
Cauchy	--	--	0.098	0.098	0.098	0.098	0.098	0.099	0.099	0.098
--	--	--	0.106	0.105	0.105	0.105	0.106	0.106	0.106	0.106

TABLE 5.2.2 (Continued)

			Discriminant Procedure											
$G(\underline{x})$	(α,β)	n	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I CAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
--	(0,0)	12	0.094	0.095	0.094	0.096	0.094	0.096	0.093	0.094	0.098	0.099	0.104	0.095
		25	0.082	0.083	0.082	0.082	0.083	0.083	0.083	0.083	0.077	0.080	0.083	0.077
$N(\underline{\mu}, 9I)$	(0,0.10)	12	0.089	0.092	0.089	0.089	0.090	0.091	0.090	0.089	0.101	0.100	0.112	0.101
		25	0.082	0.083	0.082	0.081	0.081	0.081	0.084	0.084	0.083	0.079	0.082	0.085
$N(\underline{\mu}, 9I)$	(0,0.25)	12	0.106	0.108	0.108	0.109	0.109	0.108	0.108	0.107	0.118	0.109	0.110	0.118
		25	0.078	0.078	0.078	0.078	0.078	0.078	0.077	0.077	0.079	0.079	0.081	0.081
$N(\underline{\mu}, 9I)$	(0.10,0.10)	12	0.101	0.101	0.100	0.100	0.099	0.100	0.100	0.101	0.107	0.099	0.106	0.107
		25	0.076	0.077	0.076	0.076	0.076	0.076	0.078	0.079	0.078	0.079	0.084	0.080
$N(\underline{\mu}, 9I)$	(0.25,0.25)	12	0.110	0.111	0.110	0.111	0.110	0.111	0.109	0.109	0.111	0.122	0.117	0.112
		25	0.084	0.084	0.084	0.084	0.084	0.085	0.084	0.086	0.101	0.093	0.098	0.107
$N(\underline{\mu}, 9I)$	—	--	0.091	0.092	0.091	0.091	0.091	0.092	0.091	0.092	0.097	0.095	0.099	0.099
$N(\underline{\mu}, 25I)$	(0,0.10)	12	0.097	0.097	0.097	0.097	0.097	0.096	0.097	0.099	0.103	0.099	0.112	0.101
		25	0.074	0.074	0.074	0.073	0.074	0.074	0.075	0.075	0.094	0.086	0.093	0.099
$N(\underline{\mu}, 25I)$	(0,0.25)	12	0.105	0.109	0.106	0.106	0.109	0.111	0.107	0.109	0.113	0.121	0.114	0.116
		25	0.091	0.094	0.093	0.097	0.096	0.095	0.096	0.091	0.101	0.094	0.099	0.110
$N(\underline{\mu}, 25I)$	(0.10,0.10)	12	0.107	0.111	0.105	0.105	0.101	0.105	0.110	0.107	0.109	0.125	0.116	0.112
		25	0.081	0.082	0.080	0.081	0.080	0.080	0.084	0.082	0.099	0.095	0.098	0.100
$N(\underline{\mu}, 25I)$	(0.25,0.25)	12	0.120	0.127	0.120	0.120	0.129	0.127	0.129	0.130	0.105	0.103	0.109	0.114
		25	0.102	0.101	0.102	0.104	0.109	0.109	0.100	0.098	0.133	0.125	0.126	0.146
$N(\underline{\mu}, 25I)$	--	--	0.097	0.099	0.097	0.098	0.099	0.100	0.100	0.099	0.107	0.106	0.108	0.112
$N(\underline{\mu}, 100I)$	(0,0.10)	12	0.103	0.095	0.104	0.101	0.098	0.097	0.099	0.100	0.145	0.112	0.111	0.140
		25	0.090	0.090	0.090	0.090	0.090	0.089	0.092	0.088	0.104	0.091	0.103	0.111
$N(\underline{\mu}, 100I)$	(0,0.25)	12	0.139	0.137	0.139	0.138	0.149	0.149	0.140	0.139	0.139	0.121	0.133	0.141
		25	0.117	0.113	0.116	0.117	0.118	0.116	0.114	0.116	0.174	0.119	0.139	0.207
$N(\underline{\mu}, 100I)$	(0.10,0.10)	12	0.119	0.123	0.118	0.118	0.121	0.123	0.119	0.117	0.179	0.155	0.170	0.183
		25	0.100	0.100	0.101	0.101	0.098	0.098	0.103	0.101	0.119	0.104	0.113	0.135
$N(\underline{\mu}, 100I)$	(0.25,0.25)	12	0.234	0.224	0.233	0.226	0.224	0.220	0.232	0.226	0.260	0.239	0.207	0.295
		25	0.126	0.124	0.125	0.125	0.119	0.117	0.125	0.121	0.140	0.133	0.127	0.151
$N(\underline{\mu}, 100I)$	--	--	0.129	0.126	0.128	0.127	0.127	0.126	0.128	0.126	0.158	0.134	0.138	0.170
Cauchy	(0,0.10)	12	0.101	0.105	0.101	0.100	0.099	0.101	0.103	0.100	0.095	0.092	0.093	0.092
		25	0.086	0.087	0.086	0.086	0.086	0.086	0.088	0.086	0.083	0.084	0.085	0.086
Cauchy	(0,0.25)	12	0.088	0.090	0.088	0.087	0.085	0.086	0.089	0.089	0.123	0.096	0.117	0.120
		25	0.079	0.079	0.079	0.079	0.081	0.080	0.081	0.079	0.182	0.086	0.127	0.178
Cauchy	(0.10,0.10)	12	0.092	0.094	0.093	0.092	0.091	0.092	0.095	0.094	0.096	0.097	0.097	0.094
		25	0.087	0.087	0.086	0.087	0.087	0.087	0.086	0.087	0.143	0.133	0.143	0.151
Cauchy	(0.25,0.25)	12	0.101	0.096	0.102	0.105	0.105	0.104	0.098	0.101	0.136	0.089	0.127	0.161
		25	0.083	0.084	0.083	0.083	0.082	0.083	0.084	0.083	0.132	0.083	0.099	0.138
Cauchy	--	--	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.124	0.095	0.111	0.128
--	--	--	0.101	0.101	0.101	0.101	0.101	0.101	0.102	0.101	0.121	0.107	0.114	0.127

TABLE 5.2.3

STANDARD DEVIATIONS OF THE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE ($m=4$, $\delta=1$)

			Discriminant Procedure							
C(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
--	(0,0)	12	0.049	0.045	0.049	0.051	0.052	0.050	0.047	0.047
		25	0.020	0.023	0.020	0.019	0.019	0.019	0.025	0.019
N(μ,9I)	(0,0.10)	12	0.084	0.068	0.093	0.091	0.090	0.078	0.085	0.086
		25	0.026	0.026	0.027	0.029	0.032	0.032	0.026	0.027
N(μ,9I)	(0,0.25)	12	0.047	0.038	0.047	0.045	0.048	0.044	0.046	0.047
		25	0.019	0.019	0.019	0.019	0.021	0.020	0.019	0.018
N(μ,9I)	(0.10,0.10)	12	0.034	0.034	0.033	0.033	0.032	0.032	0.035	0.034
		25	0.053	0.049	0.052	0.052	0.050	0.051	0.049	0.053
N(μ,9I)	(0.25,0.25)	12	0.041	0.027	0.040	0.031	0.023	0.025	0.040	0.037
		25	0.031	0.026	0.030	0.032	0.034	0.031	0.027	0.034
N(μ,25I)	(0,0.10)	12	0.086	0.082	0.092	0.096	0.094	0.095	0.084	0.083
		25	0.028	0.030	0.028	0.029	0.030	0.028	0.030	0.033
N(μ,25I)	(0,0.25)	12	0.076	0.077	0.074	0.074	0.072	0.075	0.077	0.080
		25	0.030	0.030	0.030	0.029	0.037	0.031	0.032	0.029
N(μ,25I)	(0.10,0.10)	12	0.050	0.058	0.047	0.041	0.041	0.044	0.057	0.048
		25	0.035	0.039	0.034	0.035	0.035	0.035	0.034	0.035
N(μ,25I)	(0.25,0.25)	12	0.072	0.069	0.071	0.070	0.066	0.064	0.069	0.071
		25	0.022	0.021	0.022	0.023	0.018	0.016	0.021	0.023
N(μ,100I)	(0,0.10)	12	0.081	0.085	0.079	0.081	0.078	0.079	0.083	0.081
		25	0.020	0.021	0.019	0.017	0.015	0.016	0.021	0.022
N(μ,100I)	(0,0.25)	12	0.068	0.073	0.058	0.071	0.085	0.075	0.060	0.067
		25	0.032	0.034	0.032	0.032	0.031	0.033	0.034	0.030
N(μ,100I)	(0.10,0.10)	12	0.053	0.062	0.054	0.053	0.050	0.045	0.050	0.052
		25	0.048	0.046	0.049	0.050	0.044	0.042	0.042	0.049
N(μ,100I)	(0.25,0.25)	12	0.055	0.060	0.056	0.064	0.067	0.071	0.055	0.053
		25	0.059	0.053	0.056	0.052	0.053	0.051	0.054	0.053
Cauchy	(0,0.10)	12	0.036	0.039	0.035	0.035	0.036	0.037	0.037	0.036
		25	0.025	0.029	0.025	0.023	0.023	0.024	0.030	0.023
Cauchy	(0,0.25)	12	0.052	0.057	0.052	0.052	0.051	0.050	0.051	0.053
		25	0.026	0.026	0.026	0.027	0.029	0.028	0.029	0.026
Cauchy	(0.10,0.10)	12	0.047	0.048	0.051	0.054	0.045	0.044	0.047	0.044
		25	0.018	0.017	0.019	0.020	0.020	0.019	0.018	0.019
Cauchy	(0.25,0.25)	12	0.061	0.063	0.063	0.063	0.054	0.056	0.063	0.063
		25	0.038	0.042	0.038	0.039	0.036	0.037	0.041	0.040

TABLE 5.2.3 (Continued)

			Discriminant Procedure											
$G(\underset{\sim}{x})$	(α,β)	n	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
—	(0,0)	12	0.059	0.057	0.059	0.061	0.061	0.059	0.060	0.060	0.040	0.049	0.039	0.039
		25	0.024	0.026	0.023	0.021	0.020	0.021	0.030	0.024	0.020	0.020	0.020	0.021
$N(\underset{\sim}{\mu},9I)$	(0,0.10)	12	0.095	0.081	0.096	0.097	0.096	0.089	0.090	0.091	0.072	0.077	0.072	0.090
		25	0.026	0.025	0.027	0.028	0.029	0.030	0.025	0.025	0.018	0.017	0.017	0.019
$N(\underset{\sim}{\mu},9I)$	(0,0.25)	12	0.059	0.044	0.058	0.056	0.056	0.053	0.053	0.060	0.060	0.061	0.061	0.058
		25	0.020	0.023	0.019	0.023	0.022	0.020	0.029	0.023	0.039	0.014	0.012	0.030
$N(\underset{\sim}{\mu},9I)$	(0.10,0.10)	12	0.028	0.028	0.035	0.034	0.025	0.026	0.030	0.030	0.047	0.043	0.043	0.044
		25	0.050	0.051	0.049	0.049	0.048	0.049	0.050	0.052	0.050	0.047	0.048	0.050
$N(\underset{\sim}{\mu},9I)$	(0.25,0.25)	12	0.032	0.028	0.034	0.027	0.031	0.034	0.031	0.029	0.084	0.082	0.093	0.079
		25	0.025	0.021	0.026	0.027	0.031	0.028	0.022	0.029	0.032	0.024	0.032	0.021
$N(\underset{\sim}{\mu},25I)$	(0,0.10)	12	0.087	0.081	0.092	0.094	0.090	0.091	0.082	0.094	0.085	0.073	0.089	0.086
		25	0.022	0.023	0.023	0.025	0.020	0.018	0.025	0.023	0.048	0.028	0.040	0.045
$N(\underset{\sim}{\mu},25I)$	(0,0.25)	12	0.081	0.081	0.080	0.079	0.077	0.078	0.083	0.092	0.088	0.087	0.088	0.089
		25	0.024	0.023	0.023	0.027	0.033	0.027	0.024	0.026	0.032	0.027	0.034	0.032
$N(\underset{\sim}{\mu},25I)$	(0.10,0.10)	12	0.044	0.049	0.041	0.039	0.042	0.045	0.049	0.042	0.037	0.034	0.032	0.033
		25	0.033	0.036	0.033	0.035	0.031	0.030	0.028	0.032	0.057	0.060	0.052	0.050
$N(\underset{\sim}{\mu},25I)$	(0.25,0.25)	12	0.082	0.075	0.080	0.077	0.076	0.078	0.077	0.091	0.074	0.065	0.085	0.067
		25	0.020	0.019	0.020	0.020	0.019	0.019	0.018	0.021	0.016	0.023	0.017	0.019
$N(\underset{\sim}{\mu},100I)$	(0,0.10)	12	0.066	0.069	0.065	0.067	0.068	0.064	0.067	0.064	0.082	0.078	0.075	0.075
		25	0.021	0.023	0.020	0.017	0.013	0.015	0.023	0.021	0.036	0.038	0.036	0.037
$N(\underset{\sim}{\mu},100I)$	(0,0.25)	12	0.079	0.093	0.076	0.083	0.084	0.083	0.080	0.080	0.064	0.073	0.060	0.059
		25	0.058	0.040	0.046	0.042	0.035	0.035	0.036	0.059	0.074	0.070	0.032	0.054
$N(\underset{\sim}{\mu},100I)$	(0.10,0.10)	12	0.057	0.074	0.055	0.057	0.062	0.052	0.062	0.062	0.087	0.093	0.082	0.085
		25	0.036	0.037	0.036	0.039	0.033	0.031	0.035	0.035	0.086	0.060	0.054	0.086
$N(\underset{\sim}{\mu},100I)$	(0.25,0.25)	12	0.072	0.082	0.075	0.075	0.083	0.084	0.079	0.075	0.067	0.062	0.057	0.052
		25	0.057	0.045	0.059	0.057	0.049	0.045	0.046	0.053	0.081	0.088	0.065	0.076
Cauchy	(0,0.10)	12	0.038	0.035	0.038	0.037	0.035	0.035	0.036	0.036	0.064	0.050	0.047	0.060
		25	0.018	0.019	0.019	0.017	0.016	0.015	0.021	0.018	0.060	0.018	0.044	0.058
Cauchy	(0,0.25)	12	0.053	0.053	0.053	0.052	0.054	0.054	0.054	0.054	0.078	0.070	0.063	0.075
		25	0.030	0.027	0.030	0.030	0.032	0.030	0.028	0.030	0.073	0.069	0.048	0.073
Cauchy	(0.10,0.10)	12	0.047	0.061	0.053	0.055	0.050	0.054	0.053	0.050	0.092	0.070	0.063	0.075
		25	0.013	0.012	0.014	0.015	0.015	0.014	0.013	0.014	0.051	0.021	0.025	0.051
Cauchy	(0.25,0.25)	12	0.069	0.069	0.070	0.069	0.064	0.068	0.070	0.070	0.113	0.097	0.094	0.104
		25	0.041	0.041	0.040	0.040	0.034	0.033	0.044	0.038	0.105	0.054	0.057	0.100

TABLE 5.2.4

STANDARD DEVIATIONS OF THE PROBABILITY OF MISCLASSIFICATION FOR EACH PROCEDURE ($m=4$, $\delta=3$)

			Discriminant Procedure							
G(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I CAST -05	I TRI -05
--	(0,0)	12	0.026	0.023	0.027	0.026	0.027	0.024	0.022	0.023
		25	0.008	0.010	0.008	0.007	0.007	0.008	0.010	0.009
$N(\mu,9I)$	(0,0.10)	12	0.020	0.020	0.019	0.019	0.019	0.019	0.019	0.020
		25	0.009	0.009	0.010	0.010	0.011	0.011	0.009	0.010
$N(\mu,9I)$	(0,0.25)	12	0.031	0.031	0.030	0.030	0.029	0.030	0.032	0.032
		25	0.007	0.007	0.007	0.007	0.007	0.006	0.007	0.007
$N(\mu,9I)$	(0.10,0.10)	12	0.015	0.017	0.015	0.014	0.014	0.015	0.016	0.015
		25	0.019	0.020	0.019	0.018	0.018	0.019	0.021	0.019
$N(\mu,9I)$	(0.25,0.25)	12	0.024	0.025	0.024	0.025	0.028	0.028	0.027	0.024
		25	0.014	0.014	0.014	0.014	0.014	0.014	0.015	0.015
$N(\mu,25I)$	(0,0.10)	12	0.018	0.018	0.018	0.019	0.018	0.017	0.018	0.018
		25	0.006	0.006	0.007	0.006	0.006	0.006	0.005	0.005
$N(\mu,25I)$	(0,0.25)	12	0.021	0.021	0.020	0.021	0.022	0.022	0.022	0.022
		25	0.016	0.020	0.016	0.015	0.015	0.015	0.022	0.017
$N(\mu,25I)$	(0.10,0.10)	12	0.018	0.020	0.015	0.016	0.014	0.014	0.020	0.016
		25	0.013	0.012	0.013	0.014	0.011	0.011	0.012	0.013
$N(\mu,25I)$	(0.25,0.25)	12	0.053	0.048	0.055	0.056	0.054	0.054	0.050	0.053
		25	0.043	0.044	0.044	0.046	0.065	0.067	0.042	0.038
$N(\mu,100I)$	(0,0.10)	12	0.095	0.094	0.094	0.092	0.091	0.092	0.095	0.094
		25	0.034	0.033	0.034	0.034	0.035	0.034	0.033	0.033
$N(\mu,100I)$	(0,0.25)	12	0.086	0.083	0.086	0.086	0.093	0.095	0.085	0.087
		25	0.031	0.030	0.030	0.032	0.032	0.031	0.030	0.039
$N(\mu,100I)$	(0.10,0.10)	12	0.059	0.060	0.058	0.058	0.059	0.060	0.061	0.060
		25	0.021	0.021	0.020	0.021	0.016	0.016	0.020	0.020
$N(\mu,100I)$	(0.25,0.25)	12	0.090	0.085	0.090	0.087	0.058	0.062	0.092	0.084
		25	0.034	0.034	0.033	0.033	0.032	0.032	0.035	0.032
Cauchy	(0,0.10)	12	0.035	0.040	0.034	0.032	0.031	0.033	0.040	0.033
		25	0.010	0.010	0.010	0.009	0.008	0.009	0.011	0.010
Cauchy	(0,0.25)	12	0.130	0.129	0.130	0.130	0.130	0.130	0.129	0.130
		25	0.007	0.006	0.007	0.007	0.011	0.008	0.007	0.008
Cauchy	(0.10,0.10)	12	0.025	0.024	0.025	0.025	0.025	0.025	0.026	0.025
		25	0.012	0.010	0.012	0.013	0.012	0.012	0.009	0.011
Cauchy	(0.25,0.25)	12	0.107	0.107	0.108	0.107	0.109	0.110	0.106	0.108
		25	0.017	0.017	0.017	0.015	0.015	0.016	0.019	0.015

TABLE 5.2.4 (Continued)

			Discriminant Procedure											
			I	I	I	I	I	I	I	I	II	III	III	
C(x)	(α,β)	n	TRIM15 -10	TRIM25 -10	HUB -10	HAMHUB -10	HAMP -10	SIN -10	GAST -10	TRI -10	ROBREG	TLDF15	TLDF25	FISHER
—	(0,0)	12	0.023	0.020	0.023	0.022	0.023	0.021	0.018	0.021	0.040	0.040	0.047	0.031
		25	0.007	0.009	0.007	0.007	0.008	0.008	0.009	0.009	0.004	0.007	0.009	0.004
N(μ,9I)	(0,0.10)	12	0.021	0.024	0.020	0.021	0.019	0.020	0.020	0.019	0.026	0.026	0.031	0.024
		25	0.010	0.012	0.009	0.009	0.010	0.010	0.013	0.011	0.014	0.004	0.014	0.021
N(μ,9I)	(0,0.25)	12	0.030	0.033	0.029	0.030	0.030	0.030	0.033	0.031	0.062	0.039	0.051	0.065
		25	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.005	0.005	0.006	0.006
N(μ,9I)	(0.10,0.10)	12	0.021	0.021	0.021	0.022	0.021	0.021	0.019	0.020	0.022	0.022	0.026	0.027
		25	0.012	0.013	0.012	0.011	0.011	0.011	0.014	0.014	0.011	0.007	0.017	0.015
N(μ,9I)	(0.25,0.25)	12	0.032	0.029	0.031	0.031	0.032	0.031	0.032	0.034	0.028	0.030	0.030	0.028
		25	0.020	0.020	0.020	0.020	0.021	0.020	0.020	0.020	0.026	0.016	0.017	0.030
N(μ,25I)	(0,0.10)	12	0.021	0.021	0.021	0.024	0.024	0.022	0.021	0.025	0.029	0.026	0.031	0.029
		25	0.004	0.003	0.004	0.004	0.004	0.003	0.003	0.005	0.021	0.017	0.019	0.030
N(μ,25I)	(0,0.25)	12	0.031	0.032	0.031	0.029	0.032	0.033	0.030	0.034	0.035	0.031	0.034	0.042
		25	0.019	0.024	0.019	0.019	0.017	0.017	0.027	0.019	0.018	0.015	0.018	0.030
N(μ,25I)	(0.10,0.10)	12	0.027	0.032	0.027	0.028	0.027	0.030	0.032	0.029	0.027	0.028	0.032	0.026
		25	0.009	0.009	0.009	0.011	0.009	0.009	0.011	0.012	0.036	0.019	0.024	0.036
N(μ,25I)	(0.25,0.25)	12	0.051	0.051	0.051	0.051	0.047	0.045	0.051	0.052	0.020	0.025	0.021	0.024
		25	0.020	0.016	0.021	0.022	0.039	0.040	0.013	0.016	0.070	0.065	0.065	0.058
N(μ,100I)	(0,0.10)	12	0.060	0.040	0.054	0.040	0.031	0.031	0.049	0.054	0.090	0.039	0.034	0.093
		25	0.027	0.029	0.027	0.026	0.026	0.026	0.028	0.026	0.026	0.018	0.026	0.035
N(μ,100I)	(0,0.25)	12	0.064	0.060	0.065	0.063	0.070	0.072	0.063	0.066	0.054	0.043	0.047	0.060
		25	0.036	0.032	0.036	0.038	0.037	0.035	0.029	0.040	0.117	0.041	0.048	0.125
N(μ,100I)	(0.10,0.10)	12	0.054	0.057	0.052	0.053	0.058	0.057	0.057	0.054	0.106	0.100	0.088	0.105
		25	0.033	0.033	0.034	0.035	0.031	0.031	0.036	0.033	0.035	0.025	0.028	0.060
N(μ,100I)	(0.25,0.25)	12	0.070	0.061	0.071	0.061	0.044	0.064	0.062	0.066	0.098	0.094	0.069	0.102
		25	0.046	0.046	0.046	0.045	0.041	0.037	0.046	0.037	0.053	0.053	0.052	0.063
Cauchy	(0,0.10)	12	0.022	0.026	0.022	0.021	0.021	0.018	0.023	0.022	0.027	0.018	0.023	0.026
		25	0.019	0.020	0.019	0.017	0.016	0.017	0.020	0.017	0.018	0.017	0.021	0.030
Cauchy	(0,0.25)	12	0.018	0.017	0.017	0.017	0.014	0.015	0.019	0.019	0.128	0.026	0.105	0.126
		25	0.008	0.009	0.008	0.009	0.012	0.011	0.012	0.011	0.166	0.014	0.065	0.163
Cauchy	(0.10,0.10)	12	0.018	0.019	0.017	0.016	0.017	0.018	0.021	0.020	0.030	0.033	0.022	0.027
		25	0.017	0.015	0.017	0.018	0.017	0.017	0.012	0.016	0.111	0.106	0.110	0.116
Cauchy	(0.25,0.25)	12	0.023	0.016	0.024	0.026	0.028	0.027	0.018	0.023	0.076	0.023	0.045	0.117
		25	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.127	0.020	0.028	0.121

as well as the other procedures. In general, however, Fisher's linear discriminant function does not perform as well as the robust procedures when the initial samples are scale contaminated.

(3) As in the univariate case, the Type II Robust Regression discriminant function does not perform well. The Type II discriminant function and Fisher's LDF give very similar results.

(4) For mild contaminating distributions, e.g., $N(\underline{\mu}, 9I)$, with both samples contaminated, the Type I LDFs perform better than the others. With 10% contamination, the Type I discriminant functions are very similar. For $\delta = 3$ and 25% contamination, there is little difference among the Type I functions. For $\delta = 1$ and 25% contamination, there is a little more variation among the Type I LDFs. The Sin LDF performs best followed by the 25% Trimmed Mean LDF and Hampel's LDF.

(5) For moderate contaminating distributions, e.g., $N(\underline{\mu}, 25I)$, the Type I LDFs again perform better than the other discriminant functions. There is generally a difference of 0.01 to 0.02 in probability of misclassification between the best and worst of the Type I LDFs. The Sin LDF and the Trimmed Mean LDFs tend to perform very well in this case.

In the following special cases Fisher's LDF and the Type II and III LDFs performed very well:

$(\delta=1, (\alpha,\beta)=(0.25,0.25), n=25)$ and $(\delta=3, (\alpha,\beta)=(0.25,0.25), n=12)$.

(6) For heavier contaminating distributions, e.g., $N(\underline{\mu}, 100I)$, the situation is very similar to that of moderate contamination. The Type I LDFs perform better than the others with the

following exceptions: (δ=1,(α,β)=(0.25,0.25),n=12), (δ=3,(α,β)=(0,0.25),n=12), (δ=3,(α,β)=(0.25,0.25),n=12). There is generally a difference of 0.01 to 0.03 in probability of misclassification between the best and worst of the Type I LDFs. The Hampel, Sin, and Trimmed Mean LDFs often give the smallest probabilities of misclassification.

(7) Cauchy contamination gave results similar to those for moderate and heavy contamination. A Type I LDF is generally the safest procedure in this situation, with the Hampel and Sin discriminant functions usually performing best. There were several cases, however, where Fisher's LDF or one of the Type III LDFs performed as well as or better than the Type I discriminant functions: for (δ=1,(α,β)=(0,0.10),n=12) the 15% Trimmed LDF performed well; for (δ=1,(α,β)=(0.10,0.10),n=25), the Type III LDFs performed well; for (δ=3,(α,β)=(0,0.10),n=12) Fisher's LDF and the Type II and Type III LDFs gave smaller probabilities of misclassification than the Type I LDFs; for (δ=3,(α,β)=(0.25,0.25),n=12 and n=25), the 15% Trimmed LDF performed as well as or better than the Type I discriminant functions.

(8) As in the univariate study, differences among the methods decrease as the sample size increases, supporting once again our asymptotic results which indicate that for large samples, Fisher's linear discriminant function is optimal even when the initial samples are scale contaminated. With samples of size 25, however, Fisher's LDF may still perform very poorly with contaminated samples.

Overall, the Type I LDFs are very similar to each other.

The Type III Trimmed discriminant functions usually do not perform as well as the Type I LDFs but perform considerably better than the Robust Regression LDF or Fisher's LDF.

5.3. The Effects of Sample Size, Distance Between Populations, Amount of Contamination, Type of Contamination, and Procedure on the Probability of Misclassification (m=4)

To evaluate the effects the control variables, i.e., sample size, distance between π_1 and π_2, amount of contamination, type of contamination, and procedure used for discrimination, have on the probability of misclassification when $m = 4$, the same analyses of variance described in section 4.3 were run on the four variable simulated probabilities. In this case, the number of discriminant procedures to be evaluated is 20 rather than 12.

The results of the analysis on P^c are presented in Table 5.3.1. As in the univariate case, all main effects are highly significant. Methods interact with all other factors, making it difficult once again to make general statements about the relative performance of the twenty discriminant functions without always qualifying our statements. In addition to the highly significant interaction of all factors with methods, the amount by type of contamination interaction, the sample size by distance interaction and the sample size by amount of contamination interaction are significant. This makes it difficult to make conclusions about the effects of any factors without careful qualifications.

Table 5.3.2 gives the mean probability of misclassification for each method within each level of the main factors. The overall means and the means for each level of δ are given. Recall that b is the number of blocks, i.e., the number of pairs of samples, included in the

TABLE 5.3.1

ANALYSIS OF VARIANCE ON THE PROBABILITY OF MISCLASSIFICATION
(m=4)

Effect	F	df	p
Sample Size	94.38	(1, 609)	$p < .01$
Distance (δ)	6811.45	(1, 609)	$p < .01$
Amount of Contamination	13.11	(3, 609)	$p < .01$
Type of Contamination	18.37	(3, 609)	$p < .01$
Methods	43.54	(19, 11989)	$p < .01$
Sample Size × Distance	4.40	(1, 609)	$.01 < p < .05$
Sample Size × Amount of Contamination	2.95	(3, 609)	$.01 < p < .05$
Sample Size × Type of Contamination	0.79	(3, 609)	$p > .10$
Sample Size × Methods	3.77	(19, 11989)	$p < .01$
Distance × Amount of Contamination	0.45	(3, 609)	$p > .10$
Distance × Type of Contamination	1.94	(3, 609)	$p > .10$
Distance × Methods	4.35	(19, 11989)	$p < .01$
Amount × Type of Contamination	3.27	(9, 609)	$p < .01$
Amount of Contamination × Methods	2.33	(57, 11989)	$p < .01$
Type of Contamination × Methods	3.59	(57, 11989)	$p < .01$

TABLE 5.3.2

AVERAGE MARGINAL PROBABILITIES OF MISCLASSIFICATION
(m=4)

			Discriminant Procedure							
			I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
Overall		(b=640)	0.238	0.238	0.237	0.237	0.237	0.237	0.238	0.238
δ=1		(b=320)	0.370	0.370	0.369	0.370	0.369	0.368	0.371	0.371
δ=3		(b=320)	0.106	0.105	0.105	0.105	0.106	0.106	0.106	0.106
n=12 δ=1		(b=160)	0.389	0.391	0.389	0.389	0.389	0.387	0.391	0.390
n=25 δ=1		(b=160)	0.350	0.350	0.349	0.350	0.348	0.348	0.350	0.351
n=12 δ=3		(b=160)	0.120	0.119	0.120	0.119	0.120	0.119	0.120	0.119
n=25 δ=3		(b=160)	0.091	0.091	0.091	0.091	0.092	0.092	0.092	0.092
n=12		(b=320)	0.255	0.255	0.254	0.254	0.254	0.253	0.255	0.255
n=25		(b=320)	0.220	0.220	0.220	0.221	0.220	0.220	0.221	0.221
(α,β)=(0,0.10)	δ=1	(b=80)	0.364	0.365	0.364	0.365	0.365	0.365	0.366	0.366
(α,β)=(0,0.25)	δ=1	(b=80)	0.366	0.367	0.365	0.367	0.368	0.367	0.368	0.368
(α,β)=(0.10,0.10)	δ=1	(b=80)	0.365	0.367	0.364	0.365	0.362	0.361	0.364	0.365
(α,β)=(0.25,0.25)	δ=1	(b=80)	0.383	0.383	0.384	0.382	0.381	0.378	0.384	0.384
(α,β)=(0,0.10)	δ=3	(b=80)	0.092	0.092	0.092	0.092	0.092	0.092	0.092	0.091
(α,β)=(0,0.25)	δ=3	(b=80)	0.107	0.107	0.106	0.106	0.108	0.108	0.107	0.108
(α,β)=(0.10,0.10)	δ=3	(b=80)	0.095	0.096	0.094	0.094	0.093	0.094	0.096	0.095
(α,β)=(0.25,0.25)	δ=3	(b=80)	0.129	0.125	0.129	0.129	0.130	0.129	0.127	0.129
(α,β)=(0,0.10)		(b=160)	0.228	0.229	0.228	0.228	0.228	0.228	0.229	0.229
(α,β)=(0,0.25)		(b=160)	0.236	0.237	0.236	0.237	0.238	0.237	0.238	0.238
(α,β)=(0.10,0.10)		(b=160)	0.230	0.231	0.229	0.229	0.228	0.227	0.230	0.230
(α,β)=(0.25,0.25)		(b=160)	0.256	0.254	0.257	0.256	0.256	0.254	0.255	0.256
$G(\underline{x})=N(\underline{\mu},9I)$	δ=1	(b=80)	0.365	0.363	0.366	0.366	0.365	0.363	0.367	0.367
$G(\underline{x})=N(\underline{\mu},25I)$	δ=1	(b=80)	0.371	0.373	0.371	0.373	0.372	0.372	0.373	0.372
$G(\underline{x})=N(\underline{\mu},100I)$	δ=1	(b=80)	0.382	0.383	0.381	0.382	0.382	0.380	0.380	0.384
$G(\underline{x})$=Cauchy	δ=1	(b=80)	0.360	0.362	0.359	0.359	0.356	0.356	0.362	0.360
$G(\underline{x})=N(\underline{\mu},9I)$	δ=3	(b=80)	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091
$G(\underline{x})=N(\underline{\mu},25I)$	δ=3	(b=80)	0.101	0.101	0.101	0.101	0.101	0.101	0.101	0.100
$G(\underline{x})=N(\underline{\mu},100I)$	δ=3	(b=80)	0.133	0.129	0.132	0.131	0.133	0.131	0.131	0.133
$G(\underline{x})$=Cauchy	δ=3	(b=80)	0.098	0.098	0.098	0.098	0.098	0.099	0.099	0.098
$G(\underline{x})=N(\underline{\mu},9I)$		(b=160)	0.228	0.227	0.728	0.228	0.228	0.227	0.229	0.229
$G(\underline{x})=N(\underline{\mu},25I)$		(b=160)	0.236	0.237	0.236	0.237	0.237	0.237	0.237	0.236
$G(\underline{x})=N(\underline{\mu},100I)$		(b=160)	0.257	0.256	0.257	0.256	0.258	0.255	0.256	0.258
$G(\underline{x})$=Cauchy		(b=160)	0.229	0.230	0.229	0.228	0.227	0.227	0.230	0.229

TABLE 5.3.2 (Continued)

			Discriminant Procedure											
			I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
Overall		(b=640)	0.234	0.235	0.234	0.234	0.234	0.233	0.235	0.235	0.257	0.244	0.247	0.260
$\delta=1$		(b=320)	0.367	0.368	0.367	0.368	0.366	0.365	0.368	0.368	0.393	0.380	0.380	0.392
$\delta=3$		(b=320)	0.101	0.101	0.101	0.101	0.101	0.101	0.102	0.101	0.121	0.107	0.114	0.127
$n=12$ $\delta=1$		(b=160)	0.387	0.389	0.388	0.387	0.387	0.386	0.389	0.388	0.409	0.401	0.401	0.405
$n=25$ $\delta=1$		(b=160)	0.347	0.347	0.347	0.348	0.344	0.344	0.347	0.349	0.378	0.360	0.360	0.378
$n=12$ $\delta=3$		(b=160)	0.113	0.113	0.113	0.112	0.113	0.113	0.114	0.113	0.128	0.117	0.122	0.132
$n=25$ $\delta=3$		(b=160)	0.090	0.090	0.090	0.090	0.090	0.090	0.091	0.090	0.115	0.098	0.106	0.123
$n=12$		(b=320)	0.250	0.251	0.250	0.250	0.250	0.249	0.251	0.250	0.268	0.259	0.261	0.268
$n=25$		(b=320)	0.219	0.218	0.218	0.219	0.217	0.217	0.219	0.219	0.247	0.229	0.233	0.251
$(\alpha,\beta)=(0,0.10)$	$\delta=1$	(b=80)	0.362	0.364	0.361	0.362	0.361	0.361	0.364	0.363	0.375	0.368	0.373	0.372
$(\alpha,\beta)=(0,0.25)$	$\delta=1$	(b=80)	0.365	0.365	0.365	0.365	0.364	0.364	0.366	0.367	0.389	0.379	0.374	0.388
$(\alpha,\beta)=(0.10,0.10)$	$\delta=1$	(b=80)	0.361	0.365	0.362	0.363	0.360	0.360	0.363	0.363	0.398	0.380	0.386	0.394
$(\alpha,\beta)=(0.25,0.25)$	$\delta=1$	(b=80)	0.380	0.377	0.382	0.381	0.377	0.374	0.379	0.380	0.411	0.394	0.389	0.412
$(\alpha,\beta)=(0,0.10)$	$\delta=3$	(b=80)	0.089	0.089	0.089	0.089	0.089	0.088	0.090	0.089	0.101	0.093	0.099	0.102
$(\alpha,\beta)=(0,0.25)$	$\delta=3$	(b=80)	0.100	0.101	0.101	0.102	0.103	0.103	0.101	0.101	0.129	0.103	0.115	0.134
$(\alpha,\beta)=(0.10,0.10)$	$\delta=3$	(b=80)	0.095	0.097	0.095	0.095	0.094	0.095	0.097	0.096	0.116	0.111	0.116	0.120
$(\alpha,\beta)=(0.25,0.25)$	$\delta=3$	(b=80)	0.120	0.119	0.120	0.120	0.120	0.119	0.120	0.119	0.140	0.123	0.126	0.153
$(\alpha,\beta)=(0,0.10)$		(b=160)	0.226	0.226	0.225	0.225	0.225	0.224	0.227	0.226	0.238	0.231	0.236	0.237
$(\alpha,\beta)=(0,0.25)$		(b=160)	0.233	0.233	0.233	0.233	0.234	0.234	0.234	0.234	0.259	0.241	0.244	0.261
$(\alpha,\beta)=(0.10,0.10)$		(b=160)	0.228	0.231	0.229	0.229	0.227	0.227	0.230	0.229	0.257	0.245	0.251	0.257
$(\alpha,\beta)=(0.25,0.25)$		(b=160)	0.250	0.248	0.251	0.250	0.249	0.247	0.249	0.250	0.275	0.259	0.258	0.282
$G(\underline{x})=N(\underline{\mu},9I)$	$\delta=1$	(b=80)	0.364	0.363	0.365	0.364	0.363	0.362	0.365	0.367	0.378	0.374	0.375	0.376
$G(\underline{x})=N(\underline{\mu},25I)$	$\delta=1$	(b=80)	0.369	0.370	0.369	0.370	0.367	0.379	0.379	0.379	0.382	0.380	0.382	0.378
$G(\underline{x})=N(\underline{\mu},100I)$	$\delta=1$	(b=80)	0.379	0.379	0.379	0.380	0.378	0.376	0.377	0.381	0.415	0.395	0.390	0.414
$G(\underline{x})=$Cauchy	$\delta=1$	(b=80)	0.357	0.359	0.357	0.356	0.354	0.354	0.360	0.357	0.393	0.373	0.374	0.391
$G(\underline{x})=N(\underline{\mu},9I)$	$\delta=3$	(b=80)	0.091	0.092	0.091	0.091	0.091	0.092	0.091	0.092	0.097	0.095	0.099	0.099
$G(\underline{x})=N(\underline{\mu},25I)$	$\delta=3$	(b=80)	0.097	0.099	0.097	0.098	0.099	0.100	0.100	0.099	0.107	0.106	0.108	0.112
$G(\underline{x})=N(\underline{\mu},100I)$	$\delta=3$	(b=80)	0.129	0.126	0.128	0.127	0.127	0.126	0.128	0.126	0.158	0.134	0.138	0.170
$G(\underline{x})=$Cauchy	$\delta=3$	(b=80)	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.124	0.095	0.111	0.128
$G(\underline{x})=N(\underline{\mu},9I)$		(b=160)	0.227	0.227	0.228	0.228	0.227	0.227	0.228	0.229	0.238	0.235	0.237	0.238
$G(\underline{x})=N(\underline{\mu},25I)$		(b=160)	0.233	0.235	0.233	0.234	0.233	0.233	0.235	0.234	0.247	0.242	0.245	0.249
$G(\underline{x})=N(\underline{\mu},100I)$		(b=160)	0.254	0.252	0.254	0.254	0.253	0.251	0.252	0.254	0.286	0.265	0.264	0.292
$G(\underline{x})=$Cauchy		(b=160)	0.223	0.224	0.223	0.223	0.222	0.222	0.225	0.223	0.259	0.234	0.242	0.259

calculation of the mean.

Overall, Fisher's discriminant function again gives the largest probability of misclassification, viz., 0.02 to 0.03 higher than most of the other procedures. The Robust Regression LDF continues to behave very similarly to Fisher's LDF. The Type III Trimmed LDFs give a probability of misclassification about 0.01 higher than the Type I functions. The Type I LDFs with 5% trimmed variances give slightly higher probabilities of misclassification than the corresponding Type I LDFs with 10% trimmed variances. Overall, the Type I discriminant functions with 5% trimmed variances give essentially the same probabilities of misclassification; similarly the Type I estimates with 10% trimmed variances give very similar results.

The other analyses of variance (on the variables $((P^c-P)/P$, $\ell n(P^c/P)$ and $\ell n((1-P^c)/P^c))$ gave similar results with some additional significant two factor interactions. All two factor interactions except the sample size by amount of contamination and the sample size by type of contamination interactions were significant in these analyses.

Table 5.3.3 gives the marginal mean values of $(P^c-P)/P$ for each method within each level of the main factors. The conclusions concerning the relative performance of the twelve discriminant procedures are the same as those based on P^c. Overall, Fisher's LDF gives a 58% greater probability of misclassification than the optimal discriminant procedure, as compared to the best of the robust discriminant functions, i.e., the Sin LDF, which gives a 35% greater probability of misclassification than the optimal procedure. The Type II LDF does not perform well as compared to the other robust procedures. The Type III discriminant functions do not perform as well as the Type I LDFs.

TABLE 5.3.3

AVERAGE MARGINAL VALUES OF $(P^c-P)/P$
$(m=4)$

			Discriminant Procedure							
			I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
Overall		(b=640)	0.387	0.382	0.384	0.385	0.388	0.383	0.387	0.388
δ=1		(b=320)	0.196	0.199	0.195	0.196	0.194	0.190	0.199	0.200
δ=3		(b=320)	0.577	0.566	0.573	0.573	0.582	0.575	0.575	0.576
n=12 δ=1		(b=160)	0.260	0.265	0.260	0.260	0.260	0.254	0.266	0.263
n=25 δ=1		(b=160)	0.131	0.132	0.131	0.132	0.128	0.126	0.133	0.136
n=12 δ=3		(b=160)	0.791	0.770	0.786	0.782	0.784	0.777	0.784	0.782
n=25 δ=3		(b=160)	0.364	0.361	0.360	0.364	0.380	0.374	0.366	0.370
n=12		(b=320)	0.526	0.518	0.523	0.521	0.522	0.515	0.525	0.522
n=25		(b=320)	0.248	0.247	0.246	0.248	0.254	0.250	0.250	0.253
(α,β)=(0,0.10)	δ=1	(b=80)	0.177	0.181	0.178	0.180	0.181	0.181	0.185	0.184
(α,β)=(0,0.25)	δ=1	(b=80)	0.184	0.188	0.183	0.188	0.191	0.187	0.192	0.191
(α,β)=(0.10,0.10)	δ=1	(b=80)	0.182	0.188	0.178	0.180	0.170	0.168	0.179	0.181
(α,β)=(0.25,0.25)	δ=1	(b=80)	0.241	0.238	0.242	0.237	0.232	0.225	0.242	0.242
(α,β)=(0,0.10)	δ=3	(b=80)	0.370	0.377	0.370	0.369	0.374	0.374	0.374	0.363
(α,β)=(0,0.25)	δ=3	(b=80)	0.589	0.591	0.588	0.588	0.612	0.605	0.600	0.606
(α,β)=(0.10,0.10)	δ=3	(b=80)	0.421	0.430	0.405	0.402	0.394	0.400	0.429	0.416
(α,β)=(0.25,0.25)	δ=3	(b=80)	0.930	0.864	0.929	0.932	0.948	0.924	0.895	0.920
(α,β)=(0,0.10)		(b=160)	0.273	0.279	0.274	0.275	0.277	0.277	0.279	0.273
(α,β)=(0,0.25)		(b=160)	0.387	0.389	0.385	0.388	0.402	0.396	0.396	0.398
(α,β)=(0.10,0.10)		(b=160)	0.301	0.309	0.292	0.291	0.282	0.284	0.304	0.298
(α,β)=(0.25,0.25)		(b=160)	0.585	0.551	0.586	0.585	0.590	0.574	0.569	0.581
$G(\underline{x})=N(\underline{\mu},9I)$	δ=1	(b=80)	0.182	0.176	0.183	0.183	0.180	0.174	0.188	0.187
$G(\underline{x})=N(\underline{\mu},25I)$	δ=1	(b=80)	0.201	0.207	0.202	0.206	0.205	0.204	0.207	0.205
$G(\underline{x})=N(\underline{\mu},100I)$	δ=1	(b=80)	0.235	0.239	0.234	0.235	0.238	0.229	0.231	0.242
$G(\underline{x})$=Cauchy	δ=1	(b=80)	0.166	0.173	0.162	0.160	0.152	0.152	0.172	0.164
$G(\underline{x})=N(\underline{\mu},9I)$	δ=3	(b=80)	0.357	0.361	0.358	0.361	0.365	0.365	0.363	0.360
$G(\underline{x})=N(\underline{\mu},25I)$	δ=3	(b=80)	0.510	0.502	0.501	0.506	0.513	0.515	0.504	0.499
$G(\underline{x})=N(\underline{\mu},100I)$	δ=3	(b=80)	0.981	0.932	0.971	0.959	0.980	0.949	0.960	0.980
$G(\underline{x})$=Cauchy	δ=3	(b=80)	0.461	0.468	0.464	0.466	0.470	0.473	0.471	0.465
$G(\underline{x})=N(\underline{\mu},9I)$		(b=160)	0.270	0.268	0.270	0.272	0.272	0.270	0.276	0.274
$G(\underline{x})=N(\underline{\mu},25I)$		(b=160)	0.356	0.354	0.351	0.356	0.359	0.359	0.355	0.352
$G(\underline{x})=N(\underline{\mu},100I)$		(b=160)	0.608	0.585	0.603	0.597	0.609	0.589	0.596	0.611
$G(\underline{x})$=Cauchy		(b=160)	0.314	0.320	0.313	0.313	0.311	0.313	0.322	0.314

TABLE 5.3.3 (Continued)

			Discriminant Procedure											
			I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
Overall		(b=640)	0.350	0.352	0.350	0.351	0.349	0.347	0.358	0.353	0.543	0.417	0.466	0.584
δ=1		(b=320)	0.188	0.190	0.189	0.190	0.183	0.180	0.191	0.192	0.272	0.231	0.231	0.267
δ=3		(b=320)	0.512	0.515	0.511	0.512	0.515	0.514	0.524	0.513	0.813	0.604	0.701	0.901
n=12 δ=1		(b=160)	0.251	0.258	0.255	0.254	0.252	0.248	0.259	0.255	0.322	0.298	0.296	0.311
n=25 δ=1		(b=160)	0.124	0.122	0.123	0.126	0.115	0.112	0.124	0.130	0.223	0.164	0.165	0.224
n=12 δ=3		(b=160)	0.685	0.690	0.684	0.676	0.688	0.690	0.695	0.689	0.904	0.752	0.820	0.966
n=25 δ=3		(b=160)	0.340	0.339	0.338	0.347	0.342	0.339	0.353	0.337	0.722	0.457	0.582	0.836
n=12		(b=320)	0.468	0.474	0.470	0.465	0.470	0.469	0.477	0.472	0.613	0.525	0.558	0.638
n=25		(b=320)	0.232	0.230	0.230	0.236	0.228	0.226	0.239	0.233	0.472	0.310	0.374	0.530
(α,β)=(0,0.10)	δ=1	(b=80)	0.172	0.177	0.169	0.172	0.169	0.168	0.179	0.175	0.214	0.192	0.206	0.204
(α,β)=(0,0.25)	δ=1	(b=80)	0.182	0.182	0.180	0.181	0.179	0.178	0.186	0.188	0.257	0.227	0.210	0.257
(α,β)=(0.10,0.10)	δ=1	(b=80)	0.169	0.181	0.173	0.174	0.164	0.164	0.175	0.174	0.289	0.229	0.248	0.276
(α,β)=(0.25,0.25)	δ=1	(b=80)	0.228	0.219	0.235	0.232	0.221	0.211	0.226	0.231	0.330	0.275	0.259	0.333
(α,β)=(0,0.10)	δ=3	(b=80)	0.336	0.334	0.335	0.325	0.322	0.316	0.342	0.332	0.510	0.387	0.474	0.524
(α,β)=(0,0.25)	δ=3	(b=80)	0.498	0.504	0.506	0.516	0.539	0.536	0.514	0.507	0.918	0.538	0.394	0.405
(α,β)=(0.10,0.10)	δ=3	(b=80)	0.422	0.447	0.416	0.421	0.405	0.422	0.446	0.433	0.734	0.651	0.730	0.797
(α,β)=(0.25,0.25)	δ=3	(b=80)	0.794	0.774	0.788	0.785	0.793	0.783	0.794	0.780	1.090	0.841	0.884	1.280
(α,β)=(0,0.10)		(b=160)	0.254	0.255	0.252	0.248	0.246	0.242	0.260	0.254	0.362	0.290	0.340	0.364
(α,β)=(0,0.25)		(b=160)	0.340	0.343	0.343	0.349	0.359	0.357	0.350	0.348	0.588	0.383	0.463	0.628
(α,β)=(0.10,0.10)		(b=160)	0.295	0.314	0.294	0.298	0.285	0.293	0.311	0.303	0.512	0.440	0.489	0.537
(α,β)=(0.25,0.25)		(b=160)	0.511	0.497	0.512	0.508	0.507	0.497	0.510	0.506	0.709	0.558	0.572	0.809
$G(\underline{x})=N(\underline{\mu},9I)$	δ=1	(b=80)	0.178	0.173	0.180	0.179	0.174	0.171	0.183	0.186	0.224	0.211	0.215	0.218
$G(\underline{x})=N(\underline{\mu},25I)$	δ=1	(b=80)	0.193	0.198	0.194	0.197	0.189	0.226	0.226	0.227	0.235	0.230	0.238	0.224
$G(\underline{x})=N(\underline{\mu},100I)$	δ=1	(b=80)	0.226	0.226	0.227	0.230	0.224	0.216	0.219	0.233	0.342	0.279	0.261	0.340
$G(\underline{x})$=Cauchy	δ=1	(b=80)	0.154	0.162	0.155	0.153	0.146	0.147	0.164	0.156	0.272	0.208	0.209	0.264
$G(\underline{x})=N(\underline{\mu},9I)$	δ=3	(b=80)	0.355	0.370	0.355	0.358	0.355	0.366	0.363	0.366	0.450	0.415	0.473	0.474
$G(\underline{x})=N(\underline{\mu},25I)$	δ=3	(b=80)	0.449	0.481	0.448	0.462	0.482	0.486	0.490	0.478	0.600	0.581	0.616	0.677
$G(\underline{x})=N(\underline{\mu},100I)$	δ=3	(b=80)	0.918	0.878	0.916	0.898	0.896	0.883	0.911	0.880	1.353	1.001	1.058	1.544
$G(\underline{x})$=Cauchy	δ=3	(b=80)	0.327	0.329	0.325	0.329	0.326	0.323	0.334	0.328	0.849	0.419	0.657	0.909
$G(\underline{x})=N(\underline{\mu},9I)$		(b=160)	0.266	0.272	0.268	0.268	0.265	0.269	0.273	0.276	0.337	0.313	0.344	0.346
$G(\underline{x})=N(\underline{\mu},25I)$		(b=160)	0.321	0.340	0.321	0.330	0.335	0.336	0.344	0.335	0.426	0.403	0.427	0.462
$G(\underline{x})=N(\underline{\mu},100I)$		(b=160)	0.572	0.552	0.571	0.564	0.560	0.550	0.565	0.557	0.847	0.640	0.660	0.942
$G(\underline{x})$=Cauchy		(b=160)	0.241	0.246	0.240	0.241	0.236	0.235	0.249	0.242	0.561	0.314	0.433	0.567

5.4. The Relative Ranking of the Twenty Discriminant Procedures (m=4)

In this section, we evaluate the relative performance of the twenty discriminant procedures by considering the relative rank of each procedure. For each block, i.e., pair of samples, within each parameter combination, the twenty procedures were assigned ranks from 1 to 20. Tied procedures were each assigned the average of the ranks that would have otherwise been assigned.

Tables 5.4.1 and 5.4.2 give, for each parameter combination, the average rank of each discrimination procedure. Table 5.4.3 is a summary table giving the average rank for each level of the factors considered in section 5.3. Recall the problem with the rank approach described in Chapter IV. Relatively small differences in actual probabilities of misclassification were weighted as large differences. This continues to exist when m = 4, but the problem is not quite as severe, since there are fewer cases where several procedures give exactly the same probability of misclassification.

The general conclusions concerning the relative performance of the twenty discriminant procedures are the same whether we consider the actual values of the probabilities of misclassification or the relative ranks. The 10% Type I discriminant functions tend to give slightly lower ranks than the 5% Type I LDFs. The Hampel and Sin LDFs rank best (9.5) followed by the Huber (9.8). Fisher's LDF and the Type II Robust Regression LDF are again the poorest procedures (ranks 12.7 and 12.6 respectively). The Type III LDFs do not perform as well as the Type I discriminant functions.

When there is no contamination, the Robust Regression LDF ranks

TABLE 5.4.1

AVERAGE RANK OF EACH DISCRIMINANT PROCEDURE BASED ON THE PROBABILITY OF MISCLASSIFICATION
($m=4$, $\delta=1$)

			Discriminant Procedure							
G(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I CAST -05	I TRI -05
--	(0,0)	12	7.9	10.6	9.9	11.7	12.5	14.5	10.0	9.7
		25	10.6	9.5	10.0	7.9	8.3	7.2	12.0	8.7
		--	9.2	10.0	9.9	9.8	10.4	10.8	11.0	9.2
N(μ,9I)	(0,0.10)	12	8.0	7.5	9.7	8.7	9.0	8.1	10.3	10.6
		25	8.7	12.5	8.9	9.7	11.3	12.2	14.0	12.1
N(μ,9I)	(0,0.25)	12	10.5	11.4	9.6	10.8	10.2	8.8	12.7	10.7
		25	10.9	11.0	10.3	10.3	11.2	11.7	11.9	14.1
N(μ,9I)	(0.10,0.10)	12	9.6	8.6	8.4	9.5	9.9	10.6	9.8	8.1
		25	10.8	10.6	10.5	10.8	9.8	10.4	10.7	9.5
N(μ,9I)	(0.25,0.25)	12	12.1	10.8	11.9	11.7	10.9	11.4	12.6	9.9
		25	8.7	8.2	10.1	11.8	9.7	6.9	7.8	10.3
N(μ,9I)	--	--	9.9	10.1	9.9	10.4	10.3	10.0	11.2	10.7
N(μ,25I)	(0,0.10)	12	12.0	11.3	11.0	11.4	10.1	10.5	11.4	12.7
		25	11.2	12.1	11.1	12.0	11.9	11.8	13.4	13.9
N(μ,25I)	(0,0.25)	12	12.1	10.5	12.0	13.2	10.6	10.0	11.8	10.9
		25	9.8	9.6	8.5	9.2	8.7	8.1	9.3	11.6
N(μ,25I)	(0.10,0.10)	12	9.1	11.0	7.8	7.3	7.5	9.5	9.3	9.5
		25	10.7	10.3	11.0	11.5	10.4	11.2	9.8	10.8
N(μ,25I)	(0.25,0.25)	12	8.5	9.3	9.7	8.1	10.4	10.1	10.8	8.8
		25	9.1	12.6	10.1	9.8	11.3	12.9	13.5	11.5
N(μ,25I)	--	--	10.3	10.8	10.2	10.3	10.1	10.5	11.2	11.2
N(μ,100I)	(0,0.10)	12	10.4	10.8	9.5	9.7	9.5	9.7	9.5	8.3
		25	10.2	11.8	11.0	10.7	8.8	8.6	10.3	10.8
N(μ,100I)	(0,0.25)	12	8.6	11.1	10.3	12.1	11.9	12.8	11.4	10.4
		25	11.4	9.5	10.5	11.6	7.7	8.5	9.5	12.1
N(μ,100I)	(0.10,0.10)	12	9.9	7.8	9.2	9.0	10.7	7.4	7.6	8.9
		25	11.1	10.5	10.6	8.9	7.7	7.7	11.4	11.3
N(μ,100I)	(0.25,0.25)	12	10.6	8.6	11.3	11.4	11.7	8.5	9.4	11.2
		25	11.1	9.0	11.0	12.0	10.6	10.9	11.0	11.3
N(μ,100I)	--	--	10.4	9.9	10.4	10.7	9.8	9.3	10.0	10.5
Cauchy	(0,0.10)	12	9.2	11.0	8.8	8.4	7.6	10.1	12.0	10.6
		25	10.3	11.8	9.4	11.3	11.3	12.4	11.1	9.2
Cauchy	(0,0.25)	12	11.4	10.1	11.3	9.1	11.0	11.4	13.8	11.4
		25	8.7	10.4	9.4	9.3	8.4	8.5	12.1	10.7
Cauchy	(0.10,0.10)	12	12.3	13.4	11.5	11.0	8.2	9.0	13.1	10.9
		25	10.4	10.9	9.5	9.5	9.5	8.2	11.8	10.4
Cauchy	(0.25,0.25)	12	9.7	9.0	9.4	8.5	8.3	7.7	10.1	10.0
		25	11.7	10.1	11.6	9.8	9.7	8.8	9.7	11.1
Cauchy	--	--	10.5	10.8	10.1	9.6	9.3	9.5	11.7	10.5

TABLE 5.4.1 (Continued)

			Discriminant Procedure											
G(x)	(α,β)	n	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
—	(0,0)	12	9.0	10.7	11.2	12.5	12.7	13.9	11.5	11.4	5.9	9.6	8.1	7.2
		25	10.8	11.4	10.5	8.6	9.4	8.6	14.7	9.8	11.1	14.0	14.0	13.3
		—	9.9	11.0	10.9	10.6	11.0	11.2	13.1	10.6	8.5	11.7	11.0	10.2
N(μ,9I)	(0,0.10)	12	9.3	10.7	11.4	11.6	11.7	11.9	11.1	12.1	12.0	12.3	15.0	9.9
		25	10.6	12.5	10.9	11.4	14.0	14.8	13.0	13.4	4.9	5.8	5.7	4.0
N(μ,9I)	(0,0.25)	12	11.0	13.1	9.6	9.9	9.5	8.8	13.9	11.5	8.4	11.9	9.4	8.7
		25	9.2	9.1	9.1	9.0	10.7	10.7	10.4	11.9	10.1	9.9	9.8	9.2
N(μ,9I)	(0.10,0.10)	12	10.6	11.1	10.0	10.6	10.8	11.3	11.5	10.8	12.7	13.8	10.7	11.1
		25	8.4	10.2	6.9	7.5	7.0	5.7	9.2	8.6	16.5	13.9	16.0	17.4
N(μ,9I)	(0.25,0.25)	12	10.0	7.1	8.2	8.0	8.2	8.2	8.1	6.7	14.1	14.8	10.7	15.3
		25	10.9	9.3	11.4	12.9	9.0	6.7	7.9	12.9	15.6	12.4	12.2	15.9
N(μ,9I)	—	—	10.0	10.4	9.7	10.1	10.1	9.8	10.6	11.0	11.8	11.9	11.2	11.4
N(μ,25I)	(0,0.10)	12	9.6	9.5	9.1	9.1	8.7	8.8	8.7	10.0	11.0	13.4	11.9	9.2
		25	7.1	9.1	7.1	7.0	6.5	5.6	10.0	8.9	14.1	12.2	12.4	13.2
N(μ,25I)	(0,0.25)	12	9.6	10.1	9.7	10.3	8.1	8.2	10.7	8.6	11.9	10.1	11.1	11.1
		25	9.5	9.9	9.0	10.8	11.1	11.6	10.2	10.8	12.7	12.3	12.6	15.3
N(μ,25I)	(0.10,0.10)	12	11.5	12.6	9.6	10.8	9.0	10.4	11.4	10.2	14.4	12.2	13.9	13.5
		25	8.7	8.1	8.9	9.0	7.8	7.6	7.9	7.2	16.1	12.3	15.4	15.7
N(μ,25I)	(0.25,0.25)	12	9.6	9.1	11.0	9.9	8.6	8.4	10.7	9.2	16.9	13.4	12.4	15.7
		25	8.8	12.2	9.4	10.3	12.3	12.2	13.3	9.6	8.0	7.3	6.9	9.3
N(μ,25I)	—	—	9.3	10.1	9.2	9.7	9.0	9.1	10.4	9.3	13.1	11.7	12.1	12.9
N(μ,100I)	(0,0.10)	12	13.6	11.1	11.5	11.6	10.5	12.0	12.4	13.4	8.7	8.0	9.7	10.6
		25	8.3	9.4	8.7	8.1	5.6	5.6	8.5	9.1	17.6	14.7	16.0	16.9
N(μ,100I)	(0,0.25)	12	9.2	9.3	8.8	9.9	9.9	11.1	8.8	9.2	11.2	10.5	12.8	11.2
		25	12.6	9.7	12.0	12.4	5.2	5.9	10.5	10.7	14.0	10.4	10.3	15.9
N(μ,100I)	(0.10,0.10)	12	10.3	9.7	10.4	10.7	10.7	8.5	10.3	9.9	16.8	10.6	16.8	15.3
		25	10.6	8.9	9.7	9.9	4.4	4.7	10.8	10.6	17.1	14.4	13.5	16.8
N(μ,100I)	(0.25,0.25)	12	10.1	7.5	12.2	12.0	13.1	9.1	8.0	11.1	13.8	11.5	7.2	12.1
		25	8.2	8.2	9.0	10.1	9.4	8.9	7.4	8.8	14.3	12.4	11.8	15.1
N(μ,100I)	—	—	10.4	9.2	10.3	10.6	8.6	8.2	9.6	10.4	14.2	11.6	12.3	14.2
Cauchy	(0,0.10)	12	12.7	14.1	10.4	8.9	8.8	10.8	14.5	13.2	11.0	9.4	10.1	8.8
		25	6.7	10.6	7.0	9.9	9.2	9.6	11.5	7.3	14.3	14.1	12.2	11.3
Cauchy	(0,0.25)	12	8.6	10.0	8.6	6.7	8.9	10.2	11.0	9.8	11.5	13.8	11.4	10.4
		25	8.7	8.9	9.6	9.6	9.5	8.4	10.5	10.4	15.8	12.6	12.6	16.4
Cauchy	(0.10,0.10)	12	10.4	11.7	9.9	9.0	8.1	9.4	9.9	8.9	11.5	10.1	11.7	11.6
		25	11.7	11.8	10.6	11.2	11.5	10.1	13.1	11.9	8.7	10.8	9.2	9.6
Cauchy	(0.25,0.25)	12	12.9	12.4	12.7	11.5	8.9	8.2	12.4	11.5	12.0	11.9	10.4	12.4
		25	12.2	10.9	11.7	10.9	8.4	7.0	10.4	10.4	12.8	9.0	10.1	14.2
Cauchy	—	—	10.5	11.3	10.1	9.7	9.2	9.2	11.7	10.4	12.2	11.5	11.0	11.8

TABLE 5.4.2

AVERAGE RANK OF EACH DISCRIMINANT PROCEDURE BASED ON THE PROBABILITY OF MISCLASSIFICATION
($m=4$, $\delta=3$)

G(x)	(α,β)	n	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
			Discriminant Procedure							
--	(0,0)	12	12.5	11.3	11.7	11.2	10.2	11.5	10.8	11.3
		25	12.2	9.9	12.0	11.8	13.3	12.3	10.0	10.3
		--	12.3	10.6	11.8	11.5	11.7	11.9	10.4	10.8
$N(\mu,9I)$	(0,0.10)	12	9.6	10.0	10.9	9.9	9.1	8.1	12.1	9.9
		25	8.7	9.5	7.4	7.1	8.3	9.4	10.6	11.1
$N(\mu,9I)$	(0,0.25)	12	10.9	8.5	11.7	11.0	10.9	10.8	9.4	10.3
		25	9.9	10.3	9.4	9.2	11.1	11.7	8.8	9.8
$N(\mu,9I)$	(0.10,0.10)	12	10.6	12.1	10.5	10.7	10.0	9.7	9.8	9.9
		25	9.6	12.0	10.3	9.5	10.1	10.3	13.6	12.5
$N(\mu,9I)$	(0.25,0.25)	12	10.2	9.7	8.7	10.4	10.2	10.3	9.2	9.5
		25	9.2	11.0	8.9	8.7	10.7	11.4	11.2	12.2
$N(\mu,9I)$	--	--	9.8	10.4	9.7	9.6	10.1	10.2	10.6	10.7
$N(\mu,25I)$	(0,0.10)	12	7.1	9.1	7.0	7.4	6.4	7.1	8.6	7.8
		25	10.4	10.0	11.2	10.2	10.9	9.0	10.5	8.7
$N(\mu,25I)$	(0,0.25)	12	10.4	12.1	9.6	9.9	10.5	11.0	11.6	12.0
		25	10.7	11.5	10.5	11.0	12.3	11.6	12.8	11.3
$N(\mu,25I)$	(0.10,0.10)	12	14.2	12.5	11.9	10.2	8.7	8.3	12.4	12.7
		25	11.5	12.3	10.0	11.0	9.7	10.5	14.4	11.1
$N(\mu,25I)$	(0.25,0.25)	12	10.1	8.5	10.8	10.5	10.4	10.9	9.0	10.3
		25	10.2	8.2	10.5	12.2	12.6	13.0	8.8	9.4
$N(\mu,25I)$	--	--	10.6	10.5	10.2	10.3	10.2	10.2	11.0	10.4
$N(\mu,100I)$	(0,0.10)	12	8.8	8.6	8.4	9.1	10.6	10.3	7.8	7.1
		25	11.6	10.7	10.9	11.5	12.9	11.7	10.9	11.2
$N(\mu,100I)$	(0,0.25)	12	12.3	11.9	12.3	11.2	11.1	10.2	12.7	12.6
		25	10.0	9.6	10.2	9.8	9.5	9.4	10.4	11.2
$N(\mu,100I)$	(0.10,0.10)	12	9.2	11.6	8.2	8.1	9.7	10.1	11.8	8.6
		25	8.5	9.4	8.1	8.2	7.7	6.7	9.2	10.8
$N(\mu,100I)$	(0.25,0.25)	12	11.6	9.1	11.2	10.0	11.6	10.8	10.6	10.9
		25	12.3	10.2	11.5	11.9	12.0	10.1	10.2	13.2
$N(\mu,100I)$	--	--	10.5	10.1	10.1	10.0	10.6	9.9	10.5	10.7
Cauchy	(0,0.10)	12	10.7	13.4	11.5	11.4	11.3	13.1	11.6	10.2
		25	8.4	7.1	7.8	7.8	8.7	8.4	7.2	9.1
Cauchy	(0,0.25)	12	10.3	12.2	10.4	10.5	7.8	8.6	12.4	11.0
		25	6.8	7.5	7.2	8.1	9.9	7.3	9.0	7.9
Cauchy	(0.10,0.10)	12	9.9	12.0	8.3	8.4	7.8	9.6	12.5	11.1
		25	8.3	10.8	8.1	8.4	9.8	10.2	10.7	10.5
Cauchy	(0.25,0.25)	12	9.9	8.6	10.4	13.2	13.5	12.4	9.5	9.5
		25	11.3	11.2	10.5	10.1	8.6	9.7	11.6	10.9
Cauchy	--	--	9.5	10.4	9.3	9.7	9.7	9.9	10.6	10.0

TABLE 5.4.2 (Continued)

G(x)	(α,β)	n	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
			Discriminant Procedure											
--	(0,0)	12	9.2	12.2	9.5	11.8	9.5	11.0	10.9	10.1	8.0	9.0	11.6	8.9
		25	9.9	11.6	10.2	10.1	11.1	12.0	11.3	10.9	5.2	7.9	11.5	6.9
		--	9.5	11.9	9.8	10.9	10.3	11.5	11.1	10.5	6.6	8.4	11.5	7.9
N(μ,9I)	(0,0.10)	12	8.1	9.9	9.7	9.7	10.3	10.8	9.9	8.6	12.6	15.0	13.9	12.3
		25	12.5	13.2	11.6	10.1	10.8	11.8	12.9	14.6	10.1	8.6	9.9	12.3
N(μ,9I)	(0,0.25)	12	9.3	8.7	10.7	11.2	10.8	10.3	10.2	10.5	11.8	13.0	10.0	10.4
		25	12.0	9.5	11.1	10.3	10.8	11.6	8.4	7.8	11.7	11.3	13.1	12.6
N(μ,9I)	(0.10,0.10)	12	10.8	11.6	10.6	10.5	9.7	10.8	10.1	10.3	13.3	9.2	10.2	11.2
		25	7.0	10.0	7.0	7.0	8.3	8.7	10.5	11.7	11.9	11.5	15.3	13.6
N(μ,9I)	(0.25,0.25)	12	11.1	11.2	9.7	11.7	9.8	12.0	9.8	9.8	11.1	13.9	11.6	10.6
		25	7.9	8.5	7.7	6.9	8.7	9.1	8.2	10.9	14.4	13.7	15.4	15.9
N(μ,9I)	--	--	9.8	10.3	9.8	9.7	9.9	10.6	10.0	10.5	12.1	12.0	12.4	12.4
N(μ,25I)	(0,0.10)	12	11.8	11.8	11.8	11.3	11.9	10.3	12.5	13.2	13.8	12.4	17.5	11.7
		25	7.1	8.4	7.0	6.9	7.2	7.4	9.1	10.1	16.8	15.6	16.8	17.2
N(μ,25I)	(0,0.25)	12	8.1	9.8	8.2	9.1	10.1	10.5	8.4	10.2	12.4	12.9	12.2	11.7
		25	7.7	8.5	8.5	9.0	9.8	8.3	9.7	7.1	12.6	9.7	12.4	15.4
N(μ,25I)	(0.10,0.10)	12	9.8	11.1	8.4	8.2	4.9	6.8	10.9	9.7	10.1	15.1	13.8	10.8
		25	7.8	10.6	6.6	7.8	6.5	7.3	12.9	9.2	12.6	13.1	13.2	12.6
N(μ,25I)	(0.25,0.25)	12	9.5	11.5	9.9	10.6	12.8	13.2	11.7	11.4	8.6	9.0	10.9	10.0
		25	9.4	7.9	9.1	9.6	8.9	9.4	8.5	7.3	15.0	11.5	10.9	18.2
N(μ,25I)	--	--	8.9	10.0	8.7	9.1	9.0	9.2	10.5	9.8	12.7	12.4	13.5	13.5
N(μ,100I)	(0,0.10)	12	9.8	7.9	11.7	12.6	11.3	11.0	9.2	7.9	15.4	14.2	13.2	15.5
		25	9.1	7.5	9.4	9.2	9.6	8.8	8.7	5.9	14.9	9.1	12.1	14.6
N(μ,100I)	(0,0.25)	12	10.4	10.0	10.2	9.4	10.2	10.2	11.4	10.4	9.2	5.7	9.4	9.7
		25	9.8	7.5	9.5	10.5	10.6	8.7	8.9	9.5	14.2	9.6	13.9	17.6
N(μ,100I)	(0.10,0.10)	12	10.2	9.8	9.7	9.4	8.9	8.7	9.5	7.9	15.8	13.3	13.4	16.8
		25	10.5	11.6	9.7	9.8	8.3	7.1	12.7	12.1	16.0	12.7	15.2	16.2
N(μ,100I)	(0.25,0.25)	12	11.4	9.1	10.4	9.2	7.6	8.1	12.6	9.5	13.5	13.5	6.8	15.1
		25	10.5	8.7	10.3	11.1	8.3	7.7	9.9	8.3	11.7	10.0	8.9	13.7
N(μ,100I)	--	--	10.2	9.0	10.1	10.2	9.4	8.8	10.4	8.9	13.8	11.0	11.6	14.9
Cauchy	(0,0.10)	12	9.9	11.4	9.2	8.2	9.9	9.0	10.0	9.7	9.7	10.2	11.4	8.6
		25	12.9	13.9	12.3	13.6	14.7	14.3	15.1	12.6	9.0	10.1	7.9	9.4
Cauchy	(0,0.25)	12	10.1	12.2	9.4	10.1	7.8	8.1	11.6	10.2	12.8	15.4	11.7	8.9
		25	11.3	12.1	11.6	11.3	12.2	11.8	13.5	11.1	12.5	13.2	13.8	13.5
Cauchy	(0.10,0.10)	12	10.2	13.2	9.7	9.7	9.6	11.4	13.6	12.1	8.5	10.8	11.9	8.3
		25	10.0	10.8	9.4	10.6	11.4	11.5	11.2	11.1	10.0	12.8	12.8	12.0
Cauchy	(0.25,0.25)	12	10.2	7.6	11.4	13.8	12.5	12.5	7.8	9.2	11.4	4.6	12.0	10.6
		25	10.4	11.2	9.6	9.7	8.6	10.0	10.9	10.4	11.5	7.3	13.1	13.8
Cauchy	--	--	10.6	11.6	10.3	10.9	10.8	11.1	11.7	10.8	10.7	10.6	11.8	10.6

TABLE 5.4.3

AVERAGE MARGINAL RANK OF EACH DISCRIMINANT PROCEDURE
(m=4)

			Discriminant Procedure							
			I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
Overall		(b=640)	10.2	10.4	10.0	10.1	10.0	9.9	10.8	10.6
δ=1		(b=320)	10.3	10.4	10.1	10.3	9.9	9.8	11.0	10.7
δ=3		(b=320)	10.1	10.4	9.8	9.9	10.1	10.0	10.7	10.5
n=12	δ=1	(b=160)	10.3	10.1	10.1	10.0	9.8	9.7	11.0	10.2
n=25	δ=1	(b=160)	10.3	10.7	10.2	10.5	9.9	9.9	11.1	11.3
n=12	δ=3	(b=160)	10.4	10.6	10.1	10.1	10.0	10.1	10.7	10.2
n=25	δ=3	(b=160)	9.8	10.1	9.5	9.7	10.3	10.0	10.6	10.7
n=12		(b=320)	10.3	10.4	10.1	10.0	9.9	9.9	10.8	10.2
n=25		(b=320)	10.1	10.4	9.9	10.1	10.1	10.0	10.9	11.0
(α,β)=(0,0.10)	δ=1	(b=80)	10.0	11.1	9.9	10.2	9.9	10.4	11.5	11.0
(α,β)=(0,0.25)	δ=1	(b=80)	10.4	10.5	10.2	10.7	10.0	10.0	11.6	11.5
(α,β)=(0.10,0.10)	δ=1	(b=80)	10.5	10.4	9.8	9.7	9.2	9.3	10.4	9.9
(α,β)=(0.25,0.25)	δ=1	(b=80)	10.2	9.7	10.6	10.4	10.3	9.7	10.6	10.5
(α,β)=(0,0.10)	δ=3	(b=80)	9.4	9.8	9.4	9.3	9.8	9.6	9.9	9.4
(α,β)=(0,0.25)	δ=3	(b=80)	10.1	10.5	10.2	10.1	10.4	10.1	10.9	10.8
(α,β)=(0.10,0.10)	δ=3	(b=80)	10.2	11.6	9.4	9.3	9.2	9.4	11.8	10.9
(α,β)=(0.25,0.25)	δ=3	(b=80)	10.6	9.6	10.3	10.9	11.2	11.1	10.0	10.7
(α,β)=(0,0.10)		(b=160)	9.7	10.5	9.7	9.8	9.9	10.0	10.7	10.2
(α,β)=(0,0.25)		(b=160)	10.3	10.5	10.2	10.4	10.2	10.0	11.2	11.1
(α,β)=(0.10,0.10)		(b=160)	10.4	11.0	9.6	9.5	9.2	9.3	11.1	10.4
(α,β)=(0.25,0.25)		(b=160)	10.4	9.6	10.5	10.6	10.8	10.4	10.3	10.6
G(x)=N(μ,9I)	δ=1	(b=80)	9.9	10.1	9.9	10.4	10.3	10.0	11.2	10.7
G(x)=N(μ,25I)	δ=1	(b=80)	10.3	10.8	10.2	10.3	10.1	10.5	11.2	11.2
G(x)=N(μ,100I)	δ=1	(b=80)	10.4	9.9	10.4	10.7	9.8	9.3	10.0	10.5
G(x)=Cauchy	δ=1	(b=80)	10.5	10.8	10.1	9.6	9.3	9.5	11.7	10.5
G(x)=N(μ,9I)	δ=3	(b=80)	9.8	10.4	9.7	9.6	10.1	10.2	10.6	10.7
G(x)=N(μ,25I)	δ=3	(b=80)	10.6	10.5	10.2	10.3	10.2	10.2	11.0	10.4
G(x)=N(μ,100I)	δ=3	(b=80)	10.5	10.1	10.1	10.0	10.6	9.9	10.5	10.7
G(x)=Cauchy	δ=3	(b=80)	9.5	10.4	9.3	9.7	9.7	9.9	10.6	10.0
G(x)=N(μ,9I)		(b=160)	9.9	10.2	9.8	10.0	10.2	10.1	10.9	10.7
G(x)=N(μ,25I)		(b=160)	10.4	10.7	10.2	10.3	10.1	10.3	11.1	10.8
G(x)=N(μ,100I)		(b=160)	10.4	10.0	10.3	10.3	10.2	9.6	10.3	10.6
G(x)=Cauchy		(b=160)	10.0	10.6	9.7	9.7	9.5	9.7	11.1	10.3

TABLE 5.4.3 (Continued)

			Discriminant Procedure											
			I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
Overall		(b=640)	10.0	10.2	9.8	10.0	9.5	9.5	10.6	10.1	12.6	11.6	12.0	12.7
δ=1		(b=320)	10.0	10.3	9.8	10.0	9.2	9.1	10.6	10.3	12.8	11.6	11.6	12.6
δ=3		(b=320)	9.9	10.2	9.7	9.9	9.8	9.9	10.6	10.0	12.4	11.5	12.3	12.8
n=12 δ=1		(b=160)	10.6	10.6	10.2	10.0	9.6	9.7	10.8	10.4	12.4	11.7	11.6	11.7
n=25 δ=1		(b=160)	9.5	9.9	9.4	10.0	8.9	8.4	10.3	10.2	13.3	11.5	11.7	13.5
n=12 δ=3		(b=160)	10.0	10.4	10.0	10.3	9.9	10.2	10.6	10.0	11.9	11.8	11.9	11.4
n=25 δ=3		(b=160)	9.7	10.0	9.4	9.6	9.8	9.6	10.7	10.0	12.8	11.2	12.8	14.3
n=12		(b=320)	10.3	10.5	10.1	10.2	9.7	10.0	10.7	10.2	12.1	11.8	11.7	11.5
n=25		(b=320)	9.6	10.0	9.4	9.8	9.3	9.0	10.5	10.1	13.1	11.4	12.2	13.9
(α,β)=(0,0.10)	δ=1	(b=80)	9.7	10.9	9.5	9.7	9.4	9.9	11.2	10.9	11.7	11.2	11.6	10.5
(α,β)=(0,0.25)	δ=1	(b=80)	9.8	10.0	9.6	9.8	9.1	9.4	10.8	10.4	12.0	11.4	11.3	12.3
(α,β)=(0.10,0.10)	δ=1	(b=80)	10.3	10.5	9.5	9.8	8.7	8.5	10.5	9.8	14.2	12.3	13.4	13.9
(α,β)=(0.25,0.25)	δ=1	(b=80)	10.3	9.6	10.7	10.7	9.7	8.6	9.8	10.0	13.4	11.6	10.2	13.8
(α,β)=(0,0.10)	δ=3	(b=80)	10.2	10.5	10.3	10.2	10.7	10.4	10.9	10.3	12.8	11.9	12.8	12.7
(α,β)=(0,0.25)	δ=3	(b=80)	9.8	9.8	9.9	10.1	10.3	9.9	10.3	9.6	12.2	11.4	12.1	12.5
(α,β)=(0.10,0.10)	δ=3	(b=80)	9.5	11.1	8.9	9.1	8.5	9.0	11.4	10.5	12.3	12.3	13.2	12.7
(α,β)=(0.25,0.25)	δ=3	(b=80)	10.1	9.5	9.8	10.3	9.7	10.3	9.9	9.6	12.2	10.4	11.2	13.5
(α,β)=(0,0.10)		(b=160)	9.9	10.7	9.9	10.0	10.0	10.2	11.1	10.6	12.2	11.6	12.2	11.6
(α,β)=(0,0.25)		(b=160)	9.8	9.9	9.7	10.0	9.7	9.7	10.5	10.0	12.1	11.4	11.7	12.4
(α,β)=(0.10,0.10)		(b=160)	9.9	10.8	9.2	9.5	8.6	8.8	11.0	10.1	13.3	12.3	13.3	13.3
(α,β)=(0.25,0.25)		(b=160)	10.2	9.5	10.2	10.5	9.7	9.4	9.9	9.8	12.8	11.0	10.7	13.6
$G(\underline{x})=N(\underline{\mu},9I)$	δ=1	(b=80)	10.0	10.4	9.7	10.1	10.1	9.8	10.6	11.0	11.8	11.9	11.2	11.4
$G(\underline{x})=N(\underline{\mu},25I)$	δ=1	(b=80)	9.3	10.1	9.2	9.7	9.0	9.1	10.4	9.3	13.1	11.7	12.1	12.9
$G(\underline{x})=N(\underline{\mu},100I)$	δ=1	(b=80)	10.4	9.2	10.3	10.6	8.6	8.2	9.6	10.4	14.2	11.6	12.3	14.2
$G(\underline{x})$=Cauchy	δ=1	(b=80)	10.5	11.3	10.1	9.7	9.2	9.2	11.7	10.4	12.2	11.5	11.0	11.8
$G(\underline{x})=N(\underline{\mu},9I)$	δ=3	(b=80)	9.8	10.3	9.8	9.7	9.9	10.6	10.0	10.5	12.1	12.0	12.4	12.4
$G(\underline{x})=N(\underline{\mu},25I)$	δ=3	(b=80)	8.9	10.0	8.7	9.1	9.0	9.2	10.5	9.8	12.7	12.4	13.5	13.5
$G(\underline{x})=N(\underline{\mu},100I)$	δ=3	(b=80)	10.2	9.0	10.1	10.2	9.4	8.8	10.4	8.9	13.8	11.0	11.6	14.9
$G(\underline{x})$=Cauchy	δ=3	(b=80)	10.6	11.6	10.3	10.9	10.8	11.1	11.7	10.8	10.7	10.6	11.8	10.6
$G(\underline{x})=N(\underline{\mu},9I)$		(b=160)	9.9	10.4	9.7	9.9	10.0	10.2	10.3	10.8	12.0	11.9	11.8	11.9
$G(\underline{x})=N(\underline{\mu},25I)$		(b=160)	9.1	10.0	9.0	9.4	9.0	9.1	10.4	9.5	12.9	12.0	12.8	13.2
$G(\underline{x})=N(\underline{\mu},100I)$		(b=160)	10.3	9.1	10.2	10.4	9.0	8.5	10.0	9.6	14.0	11.3	11.9	14.6
$G(\underline{x})$=Cauchy		(b=160)	10.6	11.4	10.2	10.3	10.0	10.1	11.7	10.6	11.4	11.0	11.4	11.2

best (8.5 for $\delta = 1$, 6.6 for $\delta = 3$). The 10% Gastwirth and 15% Trimmed LDF rank poorest for $\delta = 1$ (13.1 and 11.7 respectively); the 5% TRIM15 LDF ranks poorest for $\delta = 3$ (rank 12.3) followed by the 5% Sin LDF and the 10% TRIM25 LDF (rank 11.9).

Table 5.4.4 gives the actual probabilities of misclassification for each of the ten pairs of samples drawn within each uncontaminated parameter combination.

5.5. Summary

In this chapter, we have examined the results of the small sample simulation experiments for the four variable case. In general, the probabilities of misclassification were higher for $m = 4$ than for $m = 1$.

When there is no contamination and $n = 25$, the twenty procedures give very similar results. For $n = 12$, the Type I and Type III discriminant functions give a probability of misclassification 0.01 to 0.02 higher than the Robust Regression LDF and Fisher's LDF.

For mild single sample contamination, Fisher's LDF performs as well as the robust procedures. In general, however, as in the univariate study, Fisher's linear discriminant function does not perform as well as the robust procedures when the initial samples are scale contaminated. The Type II Robust Regression discriminant function performed very much as it did for $m = 1$, i.e., when Fisher's discriminant function performed poorly, the Type II LDF also performed poorly. The Type III Trimmed LDFs did not perform as well in the four variable case as in the univariate study. The Type I discriminant functions performed better than the Type III LDFs. Overall, the Type I LDFs with 10%

TABLE 5.4.4

PROBABILITIES OF MISCLASSIFICATION FOR INDIVIDUAL PAIRS OF SAMPLES
WHEN THERE IS NO CONTAMINATION
(m=4)

		Discriminant Procedure							
	Sample Number	I TRIM15 -05	I TRIM25 -05	I HUB -05	I HAMHUB -05	I HAMP -05	I SIN -05	I GAST -05	I TRI -05
n=12 δ=1	1	0.471	0.471	0.474	0.477	0.477	0.476	0.473	0.469
	2	0.347	0.345	0.349	0.350	0.350	0.349	0.346	0.348
	3	0.327	0.333	0.329	0.330	0.330	0.336	0.332	0.325
	4	0.330	0.339	0.329	0.329	0.329	0.333	0.337	0.336
	5	0.354	0.355	0.358	0.368	0.373	0.396	0.354	0.356
	6	0.342	0.350	0.347	0.346	0.343	0.348	0.350	0.349
	7	0.332	0.345	0.335	0.333	0.332	0.343	0.337	0.336
	8	0.397	0.391	0.401	0.414	0.417	0.412	0.389	0.396
	9	0.319	0.320	0.320	0.321	0.326	0.329	0.319	0.320
	10	0.414	0.405	0.413	0.413	0.418	0.422	0.411	0.412
n=25 δ=1	1	.0.321	0.320	0.323	0.325	0.326	0.324	0.321	0.322
	2	0.368	0.383	0.364	0.359	0.355	0.358	0.396	0.367
	3	0.311	0.311	0.311	0.311	0.311	0.310	0.312	0.310
	4	0.357	0.357	0.357	0.356	0.356	0.356	0.359	0.362
	5	0.325	0.324	0.325	0.325	0.325	0.325	0.326	0.327
	6	0.344	0.342	0.344	0.343	0.344	0.342	0.346	0.345
	7	0.312	0.312	0.311	0.309	0.309	0.310	0.316	0.317
	8	0.337	0.331	0.335	0.334	0.333	0.335	0.336	0.334
	9	0.361	0.358	0.364	0.363	0.365	0.361	0.353	0.348
	10	0.337	0.339	0.336	0.334	0.334	0.334	0.334	0.332
n=12 δ=3	1	0.090	0.097	0.090	0.084	0.082	0.085	0.098	0.092
	2	0.154	0.140	0.156	0.152	0.154	0.148	0.136	0.146
	3	0.099	0.099	0.101	0.099	0.098	0.100	0.098	0.097
	4	0.081	0.085	0.081	0.096	0.099	0.098	0.086	0.084
	5	0.091	0.089	0.090	0.090	0.089	0.088	0.090	0.091
	6	0.071	0.072	0.071	0.068	0.067	0.071	0.072	0.071
	7	0.075	0.070	0.070	0.075	0.070	0.074	0.070	0.075
	8	0.106	0.117	0.101	0.098	0.097	0.102	0.105	0.106
	9	0.076	0.074	0.076	0.075	0.075	0.075	0.074	0.075
	10	0.120	0.120	0.122	0.125	0.126	0.124	0.122	0.117
n=25 δ=3	1	0.074	0.074	0.074	0.075	0.076	0.075	0.075	0.075
	2	0.095	0.096	0.095	0.094	0.093	0.094	0.096	0.095
	3	0.084	0.086	0.084	0.085	0.086	0.085	0.086	0.083
	4	0.077	0.068	0.077	0.076	0.076	0.077	0.068	0.069
	5	0.094	0.093	0.093	0.091	0.091	0.092	0.092	0.094
	6	0.076	0.078	0.076	0.077	0.082	0.081	0.079	0.079
	7	0.073	0.072	0.073	0.073	0.073	0.072	0.071	0.072
	8	0.081	0.075	0.082	0.082	0.082	0.081	0.073	0.076
	9	0.090	0.087	0.089	0.089	0.090	0.090	0.088	0.091
	10	0.088	0.089	0.088	0.087	0.087	0.087	0.088	0.085

TABLE 5.4.4 (Continued)

	Sample Number	I TRIM15 -10	I TRIM25 -10	I HUB -10	I HAMHUB -10	I HAMP -10	I SIN -10	I GAST -10	I TRI -10	II ROBREG	III TLDF15	III TLDF25	FISHER
		Discriminant Procedure											
n=12 δ=1	1	0.511	0.510	0.513	0.515	0.514	0.513	0.511	0.509	0.437	0.475	0.435	0.435
	2	0.348	0.335	0.353	0.357	0.355	0.350	0.336	0.339	0.365	0.369	0.388	0.381
	3	0.341	0.337	0.342	0.343	0.343	0.341	0.337	0.338	0.327	0.331	0.363	0.331
	4	0.327	0.339	0.325	0.375	0.325	0.331	0.337	0.335	0.332	0.335	0.330	0.332
	5	0.332	0.332	0.335	0.340	0.342	0.348	0.330	0.333	0.341	0.382	0.372	0.358
	6	0.357	0.366	0.362	0.362	0.358	0.364	0.366	0.366	0.335	0.356	0.345	0.335
	7	0.319	0.328	0.321	0.320	0.319	0.327	0.324	0.321	0.314	0.311	0.311	0.312
	8	0.409	0.408	0.411	0.427	0.433	0.430	0.409	0.416	0.401	0.377	0.388	0.398
	9	0.320	0.321	0.321	0.322	0.323	0.326	0.321	0.320	0.314	0.315	0.311	0.319
	10	0.378	0.370	0.378	0.378	0.383	0.388	0.407	0.407	0.369	0.402	0.374	0.368
n=25 δ=1	1	0.326	0.323	0.329	0.331	0.331	0.330	0.325	0.327	0.334	0.339	0.340	0.329
	2	0.375	0.391	0.371	0.367	0.364	0.366	0.410	0.376	0.367	0.375	0.377	0.362
	3	0.310	0.311	0.310	0.310	0.310	0.310	0.312	0.310	0.309	0.311	0.312	0.311
	4	0.376	0.376	0.375	0.366	0.366	0.366	0.378	0.380	0.363	0.371	0.367	0.366
	5	0.328	0.326	0.328	0.327	0.328	0.328	0.336	0.328	0.335	0.346	0.347	0.332
	6	0.348	0.346	0.348	0.346	0.347	0.346	0.349	0.334	0.345	0.346	0.341	0.352
	7	0.310	0.310	0.310	0.309	0.310	0.309	0.312	0.314	0.311	0.323	0.324	0.311
	8	0.332	0.335	0.330	0.329	0.329	0.331	0.333	0.330	0.328	0.329	0.329	0.330
	9	0.353	0.342	0.356	0.355	0.357	0.354	0.339	0.351	0.363	0.350	0.356	0.364
	10	0.332	0.336	0.331	0.332	0.331	0.330	0.338	0.331	0.340	0.345	0.337	0.339
n=12 δ=3	1	0.092	0.095	0.091	0.105	0.103	0.107	0.095	0.093	0.093	0.087	0.087	0.085
	2	0.145	0.131	0.147	0.143	0.145	0.139	0.127	0.137	0.159	0.174	0.140	0.150
	3	0.096	0.098	0.097	0.095	0.094	0.096	0.094	0.094	0.077	0.077	0.077	0.085
	4	0.089	0.079	0.091	0.095	0.081	0.102	0.079	0.092	0.083	0.091	0.098	0.090
	5	0.085	0.090	0.084	0.092	0.091	0.090	0.091	0.077	0.083	0.083	0.078	0.092
	6	0.072	0.074	0.071	0.072	0.071	0.072	0.074	0.072	0.068	0.071	0.080	0.068
	7	0.069	0.077	0.070	0.070	0.070	0.069	0.076	0.078	0.073	0.076	0.083	0.074
	8	0.113	0.125	0.105	0.098	0.096	0.102	0.109	0.115	0.090	0.087	0.092	0.076
	9	0.075	0.079	0.075	0.075	0.075	0.075	0.079	0.075	0.071	0.071	0.078	0.074
	10	0.108	0.107	0.110	0.113	0.114	0.112	0.110	0.105	0.186	0.172	0.225	0.153
n=12 δ=3	1	0.075	0.077	0.075	0.077	0.078	0.078	0.078	0.076	0.075	0.071	0.074	0.077
	2	0.092	0.089	0.092	0.091	0.090	0.091	0.089	0.088	0.082	0.085	0.092	0.080
	3	0.088	0.090	0.088	0.089	0.090	0.089	0.090	0.084	0.076	0.073	0.070	0.070
	4	0.071	0.072	0.071	0.071	0.071	0.072	0.072	0.072	0.070	0.070	0.078	0.073
	5	0.087	0.100	0.087	0.085	0.085	0.086	0.099	0.102	0.076	0.078	0.096	0.086
	6	0.081	0.083	0.083	0.083	0.094	0.092	0.083	0.085	0.075	0.094	0.089	0.076
	7	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.070	0.077	0.080	0.077	0.077
	8	0.086	0.082	0.087	0.087	0.087	0.086	0.079	0.082	0.076	0.076	0.076	0.075
	9	0.086	0.086	0.085	0.085	0.085	0.086	0.087	0.090	0.082	0.085	0.081	0.082
	10	0.085	0.084	0.085	0.084	0.084	0.085	0.083	0.082	0.084	0.093	0.094	0.084

Gnanadesikan-Kettenring variance estimates performed better than the Type I LDFs with 5% variance estimates. This varied from one parameter combination to another, however, with no clear pattern. The 5% Type I LDFs gave remarkably similar results; similarly for the 10% Type I LDFs.

As in the univariate study, differences between methods decreased as the sample size increased. This is consistent with our asymptotic results, i.e., as n increases, Fisher's LDF becomes optimal. For samples of size 25, however, there are still many cases where Fisher's LDF performs poorly when the initial samples are scale contaminated.

CHAPTER VI

SUMMARY AND RECOMMENDATIONS FOR FURTHER RESEARCH

6.1. Summary

In this study we have investigated one aspect of the robustness of Fisher's linear discriminant function--robustness against contamination in the initial samples. In Chapter II, we examined the asymptotic effect of normal contamination, both location contamination and scale contamination. We found that with scale contamination and equal prior probabilities for π_1 and π_2, Fisher's LDF is optimal, i.e., when the prior probabilities of the two underlying populations are equal, scale contamination has no effect on our ability to make correct classifications. When the prior probabilities are unequal, however, scale contamination adversely affects ability to discriminate between the two populations. With mild contamination, only small increases in probability of misclassification are observed. With heavier contamination, viz., the contaminating covariance matrix twenty-five to one hundred times the uncontaminated matrix, contamination has a definite harmful effect. Location contamination may be even more harmful than scale contamination, depending upon the direction of the contamination. In single sample location contamination, for example, the worst effects occur when the mean of the contaminating distribution is on the opposite side of the uncontaminated population.

In Chapter II we found that for large samples scale contamination of the initial samples has no effect on our ability to discriminate when the prior probabilities of the two underlying populations are equal. We described in Chapter III the methodology to be used in a small sample study of such contamination. The small sample study consisted of a series of computer sampling experiments. Finding that with small samples Fisher's LDF may perform poorly when the initial samples are scale contaminated, we suggested three types of alternatives to the usual linear discriminant function to be used when scale contamination is suspected. The Type I LDF is based on robust estimates of location and scale; the Type II LDF is based on robust regression coefficients; the Type III LDF is a trimmed sample discriminant function.

In Chapters IV and V we presented the results of the small sample experiments. Chapter IV gave the results for the univariate study (m=1); Chapter V gave the results for the four variable study (m=4). The results were quite similar in the two studies. In both cases we found that for mild single sample contamination, Fisher's linear discriminant function performs as well as the robust procedures. In general, however, Fisher's LDF does not perform as well as the robust procedures when the initial samples are scale contaminated. Among the robust procedures, the Type II Robust Regression discriminant function was not protective against contamination. When Fisher's LDF performed poorly, the Type II LDF also performed poorly.

For m = 1, all the Type I and Type III discriminant functions were remarkably similar to one another. For m = 4, the Type III functions tended not to perform as well as the Type I LDFs.

When the initial samples are scale contaminated, there is little difference among the Type I discriminant functions. Overall, the Sin-10 LDF gave the smallest probability of misclassification. However, the difference in probability of misclassification between the Sin-10 LDF and the worst of the Type I procedures was only 0.004 for $m = 1$, 0.005 for $m = 4$.

It should be pointed out that all the discrimination procedures became more similar to one another as n increased. Fisher's LDF gave probabilities of misclassification closer to the robust procedures for $n = 25$ than for $n = 12$. This is to be expected on the basis of our asymptotic results. However, for $n = 25$ there were still many situations where Fisher's LDF performed quite poorly.

If we are uncertain as to whether or not our initial samples are scale contaminated, we would like the robust procedure we use also to perform well when there is no contamination. For one variable, all the robust procedures performed as well as Fisher's LDF when there was no contamination. For four variables there was a little more variability among the robust procedures in the uncontaminated situation. The Sin LDF gave the largest probabilities of misclassification (0.375, 0.372 as compared to 0.357 for Fisher's LDF for $n = 12$, $\delta = 1$; 0.084, 0.083 as compared to 0.077 for Fisher's LDF for $n = 25$, $\delta = 3$).

In summary, if only mild scale contamination of the initial samples is suspected, Fisher's discriminant function will perform as well as the robust procedures. For heavier contamination, however, we may be able to discriminate better if we try a Type I robust discriminant function. Since the differences in probability of misclassification among the Type I functions are so small, any one of the Type I

LDFs would be a feasible alternative to Fisher's LDF when the initial samples are scale contaminated. The 25% Trimmed Mean LDF might be a good selection since it is quite easy to compute, easily understood by a layman and performs quite well for both contaminated and uncontaminated parameter combinations.

6.2. Recommendations for Further Research

The results of this study suggest several areas where further research might lead to useful results.

In the area of large sample scale contamination, we have examined the effect of normal contaminating distributions where the contaminating covariance matrix is a multiple of the uncontaminated matrix. These results might be extended to other types of contaminating covariance matrices and to other types of contaminating distributions. Similarly for large sample location contamination, the results of this study might be extended to other types of location contaminating distributions.

In the area of small sample contamination, we have dealt in this study only with the case of scale contamination with equal prior probabilities. In addition, we dealt with only a limited number of contaminating distributions, normal and Cauchy. It might be useful to look at other types of long or heavy tailed contaminating distributions such as Normal(0,1)/Uniform(0,1) and Normal(0,1)/Uniform($0,\frac{1}{3}$). We might also wish to reproduce the results of this study for a larger number of variables such as $m = 8$ or 10. A consideration of larger sample sizes such as $n = 50$ or 100 would be useful to determine when the asymptotic results begin to apply with moderate to heavy contamination. A natural extension of the results presented here is the consideration of unequal prior probabilities with scale contaminated initial samples. In this case, the problem mentioned in Chapter III concerning the bias correction factor for the Gnanadesikan-Kettenring covariance matrix must receive further attention. We might also investigate other robust scale estimates to be used in the Type I LDF.

Perhaps the most interesting and productive area for further research is in the area of small sample location contamination. The small sample effect of location contamination on Fisher's LDF should be investigated. The three types of robust discriminant functions studied in our small sample investigation of scale contamination can be studied in the context of location contamination. Modifications of the three types of robust LDFs (such as a one-sided Trimmed LDF) might also be considered.

LIST OF REFERENCES

[1] Anderson, J. A. "Separate Sample Logistic Discrimination." Biometrika, Volume 59, 1972, pp. 19-36.

[2] Anderson, T. W. An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc., 1958.

[3] Andrews, D. F.; Bickel, P. J.; Hampel, F. R.; Huber, P. J.; Rogers, W. H.; and Tukey, J. W. Robust Estimates of Location. Princeton, New Jersey: Princeton University Press, 1972.

[4] Barnard, M. M. "The Secular Variations of Skull Characters in Four Series of Egyptian Skulls." Annals of Eugenics, Volume 6, 1935, pp. 352-371.

[5] Cochran, W. G. "On the Performance of the Linear Discriminant Function." Technometrics, Volume 6, 1964, pp. 179-190.

[6] Cochran, W. G. and Hopkins, C. E. "Some Classification Problems with Multivariate Qualitative Data." Biometrics, Volume 17, 1961, pp. 10-32.

[7] Cooper, P. W. "Statistical Classification with Quadratic Forms." Biometrika, Voluem 50, 1963, pp. 439-448.

[8] Cox, D. R. and Brandwood, L. "On a Discriminatory Problem Connected with the Works of Plato." JRSS (B), 1959, Volume 21, pp. 195-200.

[9] Day, N. E. and Kerridge, D. F. "A General Maximum Likelihood Discriminant." Biometrics, 1967, Volume 23, pp. 313-323.

[10] Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems." Annals of Eugenics, Volume 7, 1936, pp. 179-188.

[11] Forsythe, A. B. "Robust Estimation of Straight Line Regression Coefficients by Minimizing pth Power Deviations." Technometrics, Volume 14, 1972, pp. 159-166.

[12] Gentleman, W. M. "Robust Estimation of Multivariate Location by Minimizing pth Power Deviations." Unpublished Ph.D. Thesis, Princeton University.

[13] Gessaman, M. P. and Gessaman, P. H. "Comparison of Some Multivariate Discrimination Procedures." JASA, Volume 67, 1972, pp. 468-472.

[14] Gilbert, E. S. "On Discrimination Using Qualitative Variables." JASA, Volume 63, 1968, pp. 1399-1412.

[15] Gilbert, E. S. "The Effect of Unequal Variance Covariance Matrices on Fisher's Linear Discriminant Function." Biometrics, Volume 25, 1969, pp. 505-516.

[16] Gnanadesikan, R. and Kettenring, J. R. "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data." Biometrics, Volume 28, 1972, pp. 81-124.

[17] Hampel, F. R. "The Influence Curve and Its Role in Robust Estimation." JASA, Volume 69, 1974, pp. 383-394.

[18] Hoel, P. G. and Peterson, R. P. "A Solution to the Problem of Optimum Classification." Ann. Math. Stat., Volume 20, 1949, pp. 433-438.

[19] Huber, P. J. "The 1972 Wald Lecture--Robust Statistics: A Review." Ann. Math. Stat., Volume 43, 1972, pp. 1041-1067.

[20] Lachenbruch, P. A. "Discriminant Analysis When the Initial Samples are Misclassified." Technometrics, Volume 8, 1966, pp. 657-662.

[21] Lachenbruch, P. A. and Mickey, M. R. "Estimation of Error Rates in Discriminant Analysis." Technometrics, Volume 10, 1968, pp. 1-11.

[22] Lachenbruch, P. A. "Some Results on the Multiple Group Discriminant Problem." University of North Carolina Mimeo Series, No. 829, 1972.

[23] Lachenbruch, P. A., Sneeringer, C., and Revo, L. T. "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality." Communications in Statistics, Volume 1, 1973, pp. 39-56.

[24] Lachenbruch, P. A. "Discriminant Analysis When the Initial Samples are Misclassified II: Non-Random Misclassification Models." Technometrics, Volume 16, 1974, pp. 419-424.

[25] Linhart, H. "Techniques for Discriminant Analysis with Discrete Variables." Metrika, Volume 2, 1959, pp. 138-149.

[26] Lubischev, A. A. "On the Use of Discriminant Functions in Taxonomy." Biometrics, Volume 18, 1962, pp. 455-477.

[27] Mosteller, F. and Wallace, D. L. "Inference in an Authorship Problem." JASA, Volume 58, 1963, pp. 275-309.

[28] Neave, H. R. "On Using the Box-Müller Transformation with Multiplicative Congruential Pseudo-random Number Generators." Applied Statistics, Volume 22, 1973, pp. 92-97.

[29] Parzen, E. "On Estimation of a Probability Density Function and Its Mode." Ann. Math. Stat., Volume 33, 1962, pp. 1065-1076.

[30] Revo, L. T. "On Classifying With Certain Types of Ordered Qualitative Variates: An Evaluation of Several Procedures." University of North Carolina Mimeo Series, No. 708, 1970.

[31] Specht, D. F. "Series Estimation of a Probability Density Function." Technometrics, Volume 13, 1971, pp. 409-424.

[32] Truett, J.; Cornfield, J.; and Kannel, W. "A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham." Journal of Chronic Diseases, Volume 20, 1967, pp. 511-524.

[33] Tukey, J. W. "A Survey of Sampling from Contaminated Distributions." Contributions to Probability and Statistics. Edited by I. Oklin. Stanford, California: Stanford University Press, 1960, pp. 448-485.

[34] Weiner, J. and Dunn, O. J. "Elimination of Variates in Linear Discrimination Problems." Biometrics, Volume 22, 1966, pp. 268-275.

[35] Welch, B. L. "Note on Discriminant Functions." Biometrika, Volume 31, 1939, pp. 218-220.