DISCRIMINANT ANALYSIS UNDER TRUNCATION\textsuperscript{1}

Judith R. O'Fallon\textsuperscript{2}

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 885

August, 1973

\textsuperscript{1} Ph.D. dissertation, August, 1973.

\textsuperscript{2} This work was partially supported by the U.S. Army Research Office, Durham, under contract DAHC04-71-C-0042.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td><strong>I</strong> INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation for the Research</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Content of the Chapters</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Notation and Abbreviations</td>
<td>4</td>
</tr>
<tr>
<td>1.4 References to Equations, Theorems, etc</td>
<td>5</td>
</tr>
<tr>
<td><strong>II</strong> LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS WITH EQUAL COVARIANCE MATRICES BASED ON A TRUNCATED SAMPLING PROCEDURE</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Two Linear Discriminant Functions</td>
<td>7</td>
</tr>
<tr>
<td>2.2 The Joint Distribution of the ODF and the BLDF</td>
<td>14</td>
</tr>
<tr>
<td>2.3 The Decision Rules Based on the ODF and the BLDF</td>
<td>15</td>
</tr>
<tr>
<td>2.3.1 Statement of Basic Assumptions</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Analysis of the Functions $\xi(\cdot)$, $\omega(\cdot)$, and $\psi(\cdot)$</td>
<td>17</td>
</tr>
<tr>
<td>2.5 Computation of the Critical Value $k_2$</td>
<td>23</td>
</tr>
<tr>
<td>2.5.1 The Equation Which $k_2$ Must Satisfy</td>
<td>23</td>
</tr>
<tr>
<td>2.5.2 The Relationship Between $k_2$ and the Critical Value of the Optimal Decision Rule Based on $Z$</td>
<td>26</td>
</tr>
<tr>
<td>2.5.3 Existence and Uniqueness of $k_2$</td>
<td>28</td>
</tr>
<tr>
<td>2.5.4 Computation of $k_2$</td>
<td>32</td>
</tr>
<tr>
<td>2.6 Location of the Point $(k_1, k_2)$</td>
<td>33</td>
</tr>
<tr>
<td>2.7 Formulas for Computing the Probabilities of Agreement and Disagreement</td>
<td>36</td>
</tr>
<tr>
<td><strong>III</strong> LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS WITH UNEQUAL COVARIANCE MATRICES (NO TRUNCATION)</td>
<td>41</td>
</tr>
<tr>
<td>3.1 Basic Assumptions, Definitions, and General Procedure</td>
<td>41</td>
</tr>
<tr>
<td>3.2 The Optimal Decision Rule Based on a Specified Linear Function</td>
<td>43</td>
</tr>
<tr>
<td>3.3 The Corresponding Probability of Misclassification</td>
<td>45</td>
</tr>
<tr>
<td>3.4 Some Properties of the Functions $F$ and $F^o$</td>
<td>49</td>
</tr>
<tr>
<td>3.4.1 Some Properties of $D$ and $h$</td>
<td>49</td>
</tr>
<tr>
<td>3.4.2 Properties of $F$ and $F^o$</td>
<td>52</td>
</tr>
<tr>
<td>3.5 Comparison of the OLDF with Fisher's BLDF</td>
<td>61</td>
</tr>
<tr>
<td>3.5.1 The BLDF and the MBLDF</td>
<td>61</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>3.5.2 Numerical Examples Comparing the OLDF, the BLDF, and the MBLDF.</td>
<td>63</td>
</tr>
<tr>
<td>3.5.3 Conclusions Drawn from the Examples</td>
<td>69</td>
</tr>
<tr>
<td>3.6 Linear Discrimination When $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$.</td>
<td>70</td>
</tr>
<tr>
<td>3.6.1 Some Properties of $F$ When $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$.</td>
<td>72</td>
</tr>
<tr>
<td>3.6.2 Numerical Investigation into the Properties of $F$.</td>
<td>82</td>
</tr>
<tr>
<td>3.6.3 Rules for Predicting the OLDF When $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$.</td>
<td>92</td>
</tr>
</tbody>
</table>

**IV** LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS WITH UNEQUAL COVARIANCE MATRICES BASED ON A TRUNCATED SAMPLING PROCEDURE. | 93 |
| 4.0 Basic Assumptions and Definitions. | 93 |
| 4.1 The Optimal Decision Rule Based on a Specified Linear Function. | 94 |
| 4.2 When is it Optimal to Ignore the Data Based on $Y=X_1$? | 101 |
| 4.3 Behavior of the Function $Q(y)$. | 112 |
| 4.4 Behavior of the Function $G(y)$. | 115 |
| 4.4.1 Behavior of $g(y)$. | 116 |
| 4.4.2 Behavior of $b_1$ and $b_2$. | 119 |
| 4.4.3 Behavior of $a_1$ and $a_2$ When $k=0$. | 127 |
| 4.5 Behavior of the Decision Function $U(y)$. | 129 |
| 4.5.1 Behavior of $U(y)$ When $\pi'=(1,0,1,1,-2/3,1/3,1/2,0)$. | 133 |
| 4.5.2 The Limits of $U(y)$ as $y \to \infty$ When $k=0$, $\sigma_1=\sigma_2=1$, and $\rho_1\rho_2$. | 149 |
| 4.5.3 Geometrical Implications of the Limiting Values of $U(y)$. | 159 |
| 4.5.4 A Numerical Study of $U(y)$. | 166 |
| 4.6 The Probability of Misclassification. | 180 |
| 4.6.1 A General Procedure for Finding the Optimal Linear Partition. | 180 |
| 4.6.2 Numerical Studies of the "Misclassification Probability" Function. | 182 |
| 4.6.3 Results of the General Parameter Study. | 183 |
| 4.6.4 Results of the Truncation Study. | 191 |
| 4.6.5 Conclusions Drawn from the Numerical Studies. | 200 |

**V** PLANS FOR FURTHER STUDY. | 202 |
<p>| 5.1 Unfinished Business. | 202 |</p>
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Generalizations</td>
<td>203</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>204</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>205</td>
</tr>
</tbody>
</table>
Acknowledgements

When a person has studied in a department off and on for thirteen years, she becomes indebted to dozens of people for benefits large and small. There's no way to thank individually all the professors, secretaries, and fellow students throughout the years who have lent a helping hand, but I remember these favors with gratitude.

There are several people, though, without whose help I never could have completed the requirements for a Ph.D. degree. In chronological order they are:

The late George E. Nicholson, Jr., who, as departmental chairman in 1967, admitted me to the Ph.D. program on very special terms to accommodate my wish to combine the roles of student and mother in approximately equal parts, and who offered financial assistance and cheering up when each was needed.

N. L. Johnson, who agreed to direct my dissertation way back in 1967 and who has guided it with tolerant amusement and amazing patience through six long years, even obtaining an Army grant to help support it. I will be forever grateful to him for proposing a research topic that wasn't polished off by someone else in the interim!

James O. Kitchen, who for several years has cheerfully dashed off elegant programs to perform all sorts of sophisticated computations on electronic calculators and huge computers whenever my own severely limited prowess on the computers failed and who has always been willing to
share his knowledge and judgment.

Rolland and Toby Wrenn, who provided a loving home for our two sons during the hours I was in school and who, with moral support and the benefit of their experience and wisdom, helped me to manage the dual mother-student role.

Most of all, I am grateful to and for my family: My parents, Budd and Pat Rich, who loved and nurtured me and encouraged me always to do my best at whatever I chose to do. My in-laws, John and Luella O'Fallon, who supplied much encouragement and many prayers in this endeavor. Our sons, Sean and Liam, now seven and three, respectively, who cooperated so often and in so many ways to "help Mommy." And, above all, my husband Mike, who cheerfully paid the bills, supplied love and moral support, took responsibility for getting graphs and tables duplicated to meet Graduate School standards, painstakingly checked and collated the final copies, and otherwise tended to so many of the mundane tasks involved in producing a dissertation.

Bless them all!
CHAPTER I
INTRODUCTION

1.1 Motivation for the Research.

Consider the following hypothetical situation. There are two diseases which have, as their initial symptom, sharply elevated blood pressure. One of them occurs more rarely but is much more serious than the other. Unfortunately, they require different treatment procedures. To help distinguish between the two diseases, the doctor must administer a difficult and painful test; consequently, he postpones doing so until the patient's blood pressure registers above a certain critical level.

From published medical research the doctor knows the (bivariate normal) distributions of the values of the two indicators, i.e. blood pressure and test score, in the two populations. He also knows fairly accurately the probability of occurrence of each disease.

Being a busy man without easy access to a computer, he would like to find a simple but good method of making a diagnosis from his bivariate data---a linear function of the two scores, say, which minimizes his chances of wrong diagnosis. He also might want to use a cost function to adjust the weights on the probabilities of misclassification for each disease.

This dissertation attempts to help the doctor's statistician to find an appropriate solution to this problem for a large variety of bivariate normal distributions. Chapter II deals with the problem when
both disease populations have the same covariance matrix, and Chapter IV deals with it when they do not. At various stages in these two chapters it proved necessary to assume that (in terms of the hypothetical example) the rarer but more serious disease has a higher mean blood pressure than does the other disease. It was also convenient to assume that the critical blood pressure level is chosen to be no higher than the mean of the more serious disease. The result of these two assumptions is that the doctor will be able to observe elements from at least half of the distribution of the more serious disease but may be unable to observe elements from the major share of the less serious disease. Figure 1 is a diagram of two populations and a critical value satisfying these assumptions.

Now, obviously, the problem is the same in the case when the mean of the first variate is smaller in the rarer population than it is in the other one, provided the region of truncation is the set of points with values of the first variate exceeding a specified critical value. The theory in the following chapters has been developed in terms of this second version of the problem.

1.2 Content of the Chapters.

Chapter II solves the above problem when both populations have the same covariance matrix, for the optimal decision rule is indeed based on a linear function of the variates. This solution does not depend on any
restrictions concerning the means or the truncation point. It was decided to investigate another well-known linear discriminant function, namely Fisher's "best linear" discriminant function, for this problem; and when it proved to be different from the optimal linear discriminant function, the two were compared analytically. It was in the course of this analytical comparison that it became necessary to make the additional assumptions concerning the means and truncation point.

Chapter IV develops a procedure for finding the optimal linear function of the variates and the decision regions based on it for the case when the two populations have unequal covariance matrices. This development does not depend on the assumptions about the means and truncation point, but the analysis of the behavior of the procedure does.

Chapter III develops a procedure for finding the optimal linear function of the variates and the decision regions based on it for the case when the two populations have unequal covariance matrices and there is no truncation involved in the sampling procedure. This was done because, apparently, it has not been done previously. There are four papers dealing with the subject of linear discrimination between two multivariate normal populations having unequal covariance matrices: Kullback [9], Clunies-Ross and Riffenburgh [6], Anderson and Bahadur [2], and Bechtel, Gavin, and Bachand [3]. However, in each of these papers the stated purpose is to find a linear boundary which partitions the $X$-space into exactly two cells and which is in some sense "optimal"; that is, to find a linear function $a'X$ and a critical value $c$ such that the decision rule, "Choose Population 1 if and only if $a'X < c$", minimizes one of the following quantities:
(a.) one misclassification probability subject to a fixed value for the other; or

(b.) the maximum misclassification probability; or

(c.) the total probability of misclassification given prior probabilities.

Thus, the goals of these four papers differ from the goal of Chapter III, which is to find the partition with parallel linear boundaries which minimizes quantity (c), with no restrictions imposed concerning the number of cells this partition may have.

In fact, it is shown in Section 3.2 that, except for certain very special cases, the optimal decision rule based on a specified linear function $d'X$ partitions the $X$-plane into three cells.

As in Chapter II, Fisher's "best linear" discriminant function is compared, both numerically and analytically, to the optimal linear discriminant function for many of the cases under consideration in Chapter III. The identity of the optimal linear discriminant function in the special case when both variates have the same mean and same variance in both populations is studied extensively.

1.3 Notation and Abbreviations.

The following is a small list of the more unusual or non-standard abbreviations and notation used in this dissertation.

1.) Matrix notation.

   a.) A vector is denoted by an underscored lower-case letter; e.g. $\mathbf{d}$ or $\mathbf{w}$.

   b.) A matrix is denoted by an underscored capital letter; e.g. $\mathbf{D}$ or $\mathbf{R}$.

   c.) $\mathbf{u}_i$, $i=1,2$, denotes the unit vectors: $\mathbf{u}_1=(1,0)$, $\mathbf{u}_2=(0,1)$.

   d.) $\mathbf{j}$ denotes the vector containing all 1's.
2.) Other notation.
   a.) \([a,b]\), a pair of numbers inside square brackets, denotes a closed interval.
   b.) \((a,b)\), a pair of numbers inside parentheses, denotes either
       i.) an open interval, or
       ii.) an ordered pair of numbers.
       The appropriate meaning should be clear from context.
   c.) □ indicates the end of a proof.

3.) Abbreviations.
   a.) "iff" means "if and only if".
   b.) "\((x,y) = a \pm b\)" means "\(x = a+b\) and \(y = a-b\)".
   c.) "\(d+d^*\)" means "\(d\) increases to \(d^*\)" or "\(d\) approaches \(d^*\) from the left".
   d.) "\(d+d^*\)" means "\(d\) decreases to \(d^*\)" or "\(d\) approaches \(d^*\) from the right".
   d.) "L.H.S.\) and "R.H.S.\) refer to the expressions on the left-hand-side and the right-hand-side, respectively, of an equal sign or inequality sign.

1.4 References to Equations, Theorems, etc.

Within a given chapter each theorem is designated with a Roman numeral identifying the chapter followed by an Arabic numeral indicating its order of appearance in the chapter. The lemmas and corollaries are designated in the same way, with their numbers being independent of the numbers assigned to the theorems.

To simplify and often clarify the discussion, the Roman numeral is usually omitted in references to theorems or lemmas stated in the same chapter. The Roman numeral is never dropped, however, in references to results from another chapter.

Major equations and definitions are designated by Arabic numerals within parentheses, assigned in order of appearance within each chapter. Minor results are generally designated by letters or lower-case Roman
numerals inside parentheses.

A proper name followed by a number in square brackets, e.g. Tallis [13], refers to a specific book or paper listed in the Bibliography.
CHAPTER II

LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS WITH EQUAL COVARIANCE MATRICES BASED ON A TRUNCATED SAMPLING PROCEDURE.

Suppose the population of interest is a mixture of two bivariate normal populations having the same covariance matrix

$$
\Sigma = \begin{bmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}
$$

and means \( \mu_1 \) and \( \mu_2 \), where all these parameters are known. Moreover, the proportion of each population in the mixture is known. Finally, because of limitations in the measuring apparatus, any element \( X'=(X_1,X_2) \) in the population cannot be detected if \( X_1 > k \), where \( k \) is known. This is a form of what Tallis [13] defines as rectangular truncation. Results obtained for rectangular truncation are applicable in cases of plane truncation (e.g. \( X \) cannot be observed if \( c_1 X_1 + c_2 X_2 \geq K \)), as Tallis shows in the paper just cited, for there exist transformations which reduce plane truncation to rectangular truncation.

Under these conditions, what is the "best" way, in the sense that it minimizes the total probability of misclassification, to discriminate between the populations? That is, given an element \( x \) from the mixture, what rule should be used to decide which population it came from?

2.1 Two Linear Discriminant Functions.

When the distributions in the two populations are known and their proportions in the mixture are known, then the decision rule which minimizes the total probability of misclassification is
Choose Population 1 iff $af_1(x) > (1-a)f_2(x)$,
where $f_i(x)$ denotes the probability density function (p.d.f.) corresponding to the $i$th population, $i=1,2$, and $a$ is the proportion of Population 1 in the mixture (cf. Theorem 6.3.1, p.131 of Anderson [1]).

For the case of rectangular truncation under consideration here, let $f_i(x|k)$ denote the p.d.f. of the $i$th population when it is truncated to the set $\{x: x_1 \leq k\}$. Then

$$f_i(x|k) = \begin{cases} \frac{\exp\left\{(-1/2)((x-\mu_i)'\Sigma^{-1}(x-\mu_i))\right\}}{\phi(k_i)(2\pi)^{1/2}} & \text{if } x \in \{x: x_1 \leq k\}, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\phi(t) = \int_{-\infty}^{t} \phi(u)du,$$

$$\phi(t) = (1/\sqrt{2\pi})\exp\left\{-t^2/2\right\},$$

$$k_i = (x-\mu_{i1})/\sigma_i, \ i=1,2,$$

$$\mu_i' = (\mu_{i1}, \mu_{i2}), \ i=1,2,$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$ 

Consequently, the optimal decision rule is to choose Population 1 iff

$$\exp\left\{-(x-\mu_1)'\Sigma^{-1}(x-\mu_1) - (x-\mu_2)'\Sigma^{-1}(x-\mu_2)/2\right\} > \frac{1-a}{a} \frac{\phi(k_1)}{\phi(k_2)},$$

or

$$\delta'\Sigma^{-1}x > k_1$$

with

$$k_1 = \ln\left(\frac{1-a}{a} \frac{\phi(k_1)}{\phi(k_2)}\right) + [\delta'\Sigma^{-1}(\mu_1+\mu_2)]/2.$$ 

The optimal decision rule, then, depends on a linear function of $X$,

$$Y = \delta'\Sigma^{-1}X = \lambda'X,$$

which we will call the optimal discriminant function (ODF). The ODF is
a linear function in this case because both populations in the mixture
have the same variance. Observe, too, that this is the same function as
the discriminant function which is optimal for the case when there is no
truncation, but the optimal decision rule produces a different partition
of the \( \mathbf{x} \)-plane when there is truncation than when there is not. Thus
the boundary for the optimal partition in the case of truncation is par-
allel to the boundary for the optimal partition when there is no trunca-
tion.

There is another well-known linear discriminant function, one pro-
posed by Fisher [7] for use in cases when the optimal discriminant func-
tion is nonlinear. This "best linear" discriminant function (BLDF) is
declared to be the linear function

\[
Z = \mathbf{n}' \mathbf{x}
\]

which has the property that the coefficient vector \( \mathbf{n} \) maximizes

\[
R^2 = \frac{[E_1(\mathbf{n}' \mathbf{x}) - E_2(\mathbf{n}' \mathbf{x})]^2}{\text{Var}_1(\mathbf{n}' \mathbf{x}) + \text{Var}_2(\mathbf{n}' \mathbf{x})}
\]

where \( E_i(\mathbf{x}) \) and \( \text{Var}_i(\mathbf{x}) \) are the mean and variance of \( \mathbf{x} \) in the \( i \)^{th} popula-
tion in the mixture, \( i=1,2 \).

**LEMMA II.1:** The BLDF has coefficient vector \( \mathbf{n}' = c\mathbf{\Theta}^{-1} \mathbf{y} \),

where \( c \) is any nonzero real number and

\[
\mathbf{\Theta} = \text{Var}_1(\mathbf{x}) + \text{Var}_2(\mathbf{x}),
\]

\[
\mathbf{y} = E_1(\mathbf{x}) - E_2(\mathbf{x}),
\]

provided \( \mathbf{\Theta} \) is a positive-definite matrix and \( \mathbf{y} \neq \mathbf{0} \).

**proof:** From definitions (11), (13), and (14), we have

\[
R^2 = \frac{(\mathbf{n}' \mathbf{y})^2}{\mathbf{n}' \mathbf{\Theta} \mathbf{n}}.
\]
Rao [12] states, in result (1f.1.1), p. 48, that such a function of \( n \) attains its supremum when \( n = 0^{-1} \). It is clear that the value of \( R^2 \) is unchanged when

\[
    n = c0^{-1} \gamma
\]

for any real \( c \neq 0 \).

Now, when the mixture consists of two normal distributions with the same covariance matrix and there is no truncation involved, the ODF and the BLDF are the same, i.e. \( Y = Z \). But when there is rectangular truncation on the observations, \( Y \neq Z \). To see this, we must obtain \( E_i(X|k) \) and \( \text{Var}_i(X|k) \), the mean and variance of the \( i \)th truncated distribution of \( X \), \( i = 1, \ldots, \), in order to get \( Z \).

**Lemma II.2**: If \( X_1 \sim N(\mu_{i1}, \sigma_{11}^2) \) truncated to the space \( \{X_1: X_1 \leq k\} \), then the mean and variance of the truncated distribution of \( X_1 \) are

\[
E_i(X_1|k) = \mu_{i1} - \sigma_{11} \xi(k_i), \quad (15)
\]

\[
\text{Var}_i(X_1|k) = \sigma_{11}^2(1 - \omega_i), \quad (16)
\]

with \( k_i \) defined in (3) and

\[
\xi(t) = \frac{\phi(t)}{\phi(t)}, \quad (17)
\]

\[
\omega(t) = \xi(t)[t + \xi(t)], \quad (18)
\]

\[
\omega_i = \omega(k_i). \quad (19)
\]

**Proof**: \( E_i(X_1|k) = E_i(X_1|X_1 \leq k) \)

\[
= \frac{\int_{-\infty}^{k_i} (\sigma_{11}t + \mu_{i1}) \exp(-t^2/2) \, dt}{\sqrt{2\pi} \phi(k_i)}
\]

\[
= \mu_{i1} - \sigma_{11} \xi(k_i).
\]

\[
E_i(X_1^2|k) = \frac{1}{\phi(k_i)} \int_{-\infty}^{k_i} (\sigma_{11}t + \mu_{i1})^2 \phi(t) \, dt
\]

\[
= \mu_{i1}^2 + \sigma_{11}^2 - \xi(k_i)[k_i \sigma_{11}^2 + 2\mu_{i1} \sigma_{11}].
\]
\[ \text{Var}_1(X_1|k) = E_1(X_1^2|k) - [E_1(X_1|k)]^2 = \sigma_1^2(1-\omega_1). \]  

**Lemma II.3:** If \( \mathbf{X} \sim N(\mu_1, \Sigma) \) truncated to the space \( \mathcal{A} \equiv \{ X : X_1 \leq k \} \), then the mean and variance of the truncated distribution of \( X_2 \) are:

\[ E_1(X_2|k) = \mu_{12} - \rho \sigma_2 \xi(k_1), \]  
\[ \text{Var}_1(X_2|k) = \sigma_2^2(1 - \rho^2 \omega_1), \]  

and the covariance between \( X_1 \) and \( X_2 \) in the truncated distribution is:

\[ \text{Cov}_1(X_1, X_2|k) = \rho \sigma_1 \sigma_2 (1-\omega_1), \]

where \( k_1, \xi(k_1), \) and \( \omega_1 \) are defined as in Lemma 2.

**Proof:** If \( \mathbf{X} \sim N(\mu_1, \Sigma) \), then \( (X_2|X_1) \sim N(\mu_{12} + \rho \sigma_2 \sigma_1^{-1}(X_1 - \mu_{11}), \sigma_2^2(1-\rho^2)) \).

To simplify notation somewhat, let\n
\[ E_{k,i}(g(X_1)) \text{ denote } E_i(g(X_1)|X_1 \leq k) \]

and\n
\[ E_i(g(X_2)|X_1) \text{ denote expectation with respect to the untruncated conditional distribution of } X_2 \text{ given } X_1. \]

Then\n
\[ E_1(X_2|k) = E_{k,i}[E_1(X_2|X_1)] \]
\[ = E_{k,i}[\mu_{12} + \rho \sigma_2 \sigma_1^{-1}(X_1 - \mu_{11})] \text{ from (23)} \]
\[ = (\mu_{12} - \rho \sigma_2 \sigma_1^{-1}\mu_{11}) + \rho \sigma_2 \sigma_1^{-1}(\mu_{11} - \sigma_1 \xi(k_1)) \text{ from (15)} \]
\[ = \mu_{12} - \rho \sigma_2 \xi(k_1). \]

\[ \text{Var}_1(X_2|k) = E_{k,i}[\text{Var}_1(X_2|X_1)] + \text{Var}_{k,i}[E_1(X_2|X_1)] \]
\[ = \sigma_2^2(1-\rho^2) + \rho^2 \sigma_2^2 \sigma_1^{-2} \text{Var}_{k,i}(X_1) \text{ from (23)} \]
\[ = \sigma_2^2(1-\rho^2 \omega_1) \text{ from (16).} \]

\[ E_1(X_1X_2|k) = E_{k,i}[X_1 \cdot E_1(X_2|X_1)] \]
\[ = (\mu_{12} - \rho \sigma_2 \sigma_1^{-1}\mu_{11})E_{k,i}(X_1) + \rho \sigma_2 \sigma_1^{-1}E_{k,i}(X_1^2) \text{ from (23)}. \]

\[ \text{Cov}_1(X_1, X_2|k) = E_1(X_1X_2|k) - E_1(X_1|k) \cdot E_1(X_2|k) \]
\[ = \rho \sigma_1 \sigma_2 (1-\omega_1). \]
LEMMA II.4: If \( \mathbf{X} \sim N(\mathbf{u}_1, \Sigma) \) truncated to the space \( A(\mathbf{X}) = \{ \mathbf{X}: x_1 \leq k \} \), then the mean vector and covariance matrix for the truncated distribution of \( \mathbf{X} \) are:

\[
E_1(\mathbf{X}|k) = \mathbf{u}_1 - \sigma_1^{-1}\xi(k_1)\mathbf{u}_1
\]

(24)

\[
\text{Var}_1(\mathbf{X}|k) = (1-\omega_1)\Sigma + \sigma_2^2(1-\rho^2)\omega_1 \mathbf{U}_{22}
\]

(25)

where \( k_1, \xi(k_1) \), and \( \omega_1 \) are defined by (3), (17), and (19), and \( \mathbf{u}_1 \) is the first unit vector,

(26)

and \( \mathbf{U}_{22} \) is the matrix with a 1 in the (2,2) position and zeroes elsewhere.

(27)

proof: Follows directly from Lemmas 2 and 3. \( \square \)

Having obtained the means and covariance matrices for the truncated distributions, we can now obtain Fisher's "best linear" discriminant.

LEMMA II.5: The BLDF for the truncated distributions has coefficient vector \( \eta = \Sigma^{-1}\xi - \psi \mathbf{u}_1 \), where

\[
\psi = \frac{2\Delta \sigma_1 \omega_1}{(2-\omega)\sigma_1^2},
\]

\( \omega = \omega_1 + \omega_2 \), and \( \Delta = \xi(k_1) - \xi(k_2) \).

proof: From Lemma 4 we can find \( \gamma \) and \( \Theta \):

Substituting (24) into (14) we get

\[
\gamma = \Delta - \Delta \sigma_1^{-1}\mathbf{u}_1,
\]

(1)

where \( \Delta = \Delta(k_1, k_2) = \xi(k_1) - \xi(k_2) \).

(28)

Substituting (25) into (13) we get

\[
\Theta = (2-\omega)\Sigma + \sigma_2^2(1-\rho^2)\omega \mathbf{U}_{22}
\]

where \( \omega = \omega_1 + \omega_2 \).

(29)

Hence

\[
\Theta^{-1} = (1/2)\left\{ \Sigma^{-1} + \sigma_1^{-2} \frac{\omega}{2-\omega} \mathbf{U}_{11} \right\}.
\]

(30)

By Lemma 1 we substitute (1)-(11) into (12) in order to obtain the coefficient vector for the BLDF:
\[ n = 2\theta^{-1}y = \Sigma^{-1}a - \frac{2\delta \sigma_1 - \omega I}{(2-\omega)\sigma_1^2} u_1. \]

The results obtained in this section can be summarized as follows:

**THEOREM II.1:** If the given population is a mixture of two bivariate normal populations, 100a\% having the \( N(\mu_1, \Sigma) \) distribution, the remainder having the \( N(\mu_2, \Sigma) \) distribution; and if the observations are truncated to the space \( A(\lambda) = \{ X: X_1 \leq k \} \), where all parameters are known, then the "optimal" discriminant function (ODF), i.e. the function of \( X \) in terms of which the Optimal Decision Rule (which minimizes the total probability of misclassification) is defined, is

\[ Y = \lambda'X = \delta'\Sigma^{-1}X, \]

and the "best linear" discriminant function (BLDF), i.e. the linear function which maximizes

\[ R^2 = \frac{[E_i(Z|k) - E_i(Z|k)]^2}{Var_i(Z|k) + Var_i(Z|k)}, \]

is

\[ Z = n'X = (\lambda - \psi u_1)'X, \]

where for \( i = 1, 2, \)

\[ \delta = \mu_1 - \mu_2, \]

\( E_i(Z|k) \) and \( Var_i(Z|k) \) are the mean and variance of \( Z \) in the \( i \)th truncated distribution of \( X, \)

\[ u_1' = (1, 0), \]

\[ \psi = \frac{2\delta \sigma_1 - \omega I}{\sigma_1^2(2-\omega_1 + \omega_2)}, \]

\[ \Delta = \xi(k_1) - \xi(k_2), \]

\[ k_i = (k - u_i'\Sigma_1^{-1}), \]

\[ \xi(t) = \phi(t)/\psi(t), \]

\[ \phi(t) = (1/\sqrt{2\pi}) \exp(-t^2/2), \]

\[ \omega_1 = \xi(k_1)[k_1 + \xi(k_1)], \quad \text{and} \quad \psi(t) = \int_0^t \phi(u) \, du. \]
2.2 The Joint Distribution of the ODF and the BLDF.

In order to compare the ODF and the BLDF, it is necessary to know their joint distribution.

THEOREM II.2: Under the given conditions of Theorem 1, if \( X \sim N(\mu_1, \Sigma) \) truncated to the region \( A(X) = \{ X : x_1 \leq k \} \), then the joint distribution of \( Y \), the ODF, and \( Z \), the BLDF, is that of a \( N(C\mu_1, C\Sigma C') \) random variable truncated to the region \( \{ (Y, Z) : (1/\psi)(Y-Z) \leq k \} \), \( i = 1, 2 \), where \( C' = (\lambda_1, \psi) \) and \( \psi, \lambda_1 \), and \( \psi \) are defined as in Theorem 1.

**proof:** Define \( Y' = (Y, Z) = X' (\lambda_1, \psi) = X' C' \).

Then \( C^{-1} = \frac{1}{\psi} \begin{pmatrix} \lambda_1 & \psi \\ \psi & \lambda_2 \end{pmatrix} \) \( Y = X \),

since \( C = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \lambda_1 & \lambda_2 \end{pmatrix} \).

Therefore \( X \in A(X) \) iff \( Y \in \{ Y : (1/\psi)(Y-Z) \leq k \} \in A(Y) \).

Applying these relationships to the p.d.f. for \( X \) given in Section 2.1, we obtain the joint density for \( Y' = (Y, Z) \):

\[
g_i(\psi|k) = \begin{cases} \frac{\exp[-(X-C\mu_1)'(C\Sigma C')^{-1}(X-C\mu_1)/2]}{2\pi\sqrt{|C\Sigma C'|} \psi(k_1)} & \text{if } Y \in A(Y); \\ 0 & \text{otherwise.} \end{cases}
\]

Now, the boundary of the region of positive probability, \( A(Y) \), is the straight line, \( Y-Z = k\psi \), which has a slope of \( 45^\circ \) in the \((Y, Z)\)-plane. (This is an instance of what Tallis [13] calls plane truncation.) The \( Z \)-intercept of this line is below or above the origin according as \( k\psi \) is positive or negative, and \( A(Y) \) is on the left or right side of this line according as \( \psi \) is positive or negative.

THEOREM II.3: Under the given conditions of Theorem 1, if \( X \sim N(\mu_1, \Sigma) \), then the means, variances, and covariance of \( Y \) and \( Z \) in the truncated
truncated distribution are
\[
E_i(Y|k) = \delta_i \Sigma^{-1} \mu_i - (\delta_i/\sigma_1) \xi(k_i),
\]
\[
E_i(Z|k) = E_i(Y|k) - \psi[\mu_i - \sigma_1 \xi(k_i)],
\]
\[
\text{Var}_i(Y|k) = \delta_i \Sigma^{-1} \delta_i - \omega_i (\delta_i/\sigma_1)^2,
\]
\[
\text{Var}_i(Z|k) = \text{Var}_i(Y|k) + \psi(1-\omega_i) (\psi \sigma_1^2 - 2 \delta_i),
\]
\[
\text{Cov}_i(Y,Z|k) = \text{Var}_i(Y|k) - \psi(1-\omega_i) \delta_i,
\]
where all quantities are defined as in Theorem 1.

**proof:** As in Theorem 2, define \( Y' = X'C = (Y,Z) \).

Then \( C' = (\Sigma^{-1} \delta, \Sigma^{-1} \delta - \psi \mu_i) \).

Since \( E_i(Y|k) = C E_i(X|k) \), we get \( E_i(Y|k) \) and \( E_i(Z|k) \) directly from (24) and the definition of \( C \).

Similarly, since \( \text{Var}_i(Y|k) = C \text{Var}_i(X|k) C' \), we get \( \text{Var}_i(Y|k) \), \( \text{Var}_i(Z|k) \), and \( \text{Cov}_i(Y,Z|k) \) directly from (25) and the definition of \( C \).

2.3 The Decision Rules Based on the ODF and the BLDF

Recall from Section 2.1 that the optimal decision rule is to choose Population 1 iff \( Y > k_1 \), where \( k_1 \) is defined in (8) and \( Y = \delta' \Sigma^{-1} X = \lambda' X \).

It can be shown, using the approach explained in Section 4.1, that the optimal decision rule based on the BLDF \( Z = n' \lambda' = (\lambda' - \psi \mu_i)' X \) is to choose Population 1 iff

\[
\frac{\phi(k_2) \exp[-(Z-v_1)^2/2\sigma_1^2] \phi(a_1-bZ)}{\phi(k_1) \exp[-(Z-v_2)^2/2\sigma_2^2] \phi(a_2-bZ)} > \frac{1-\alpha}{\alpha}, \tag{A}
\]

where \( v_1 = n' \mu_i \) and \( a_1 \) and \( b \) are complicated functions of \( \Sigma, \delta, n, \) and \( k \), \( i=1,2 \). Since it was not exactly obvious that this rule is equivalent to the one,

"Choose Population 1 iff \( Z > k_2 \)," \( \tag{B} \)

which is analogous in form to the Optimal Decision Rule (i.e. leads to a
certain type of two-cell partition of the \( \mathbb{X} \)-plane), it was decided to employ decision rule (B) and choose \( k_2 \) so as to minimize the probability of misclassification for this rule.

It will be shown in Section 2.5 that the value of \( k_2 \) which minimizes the probability of misclassification for decision rule (B) is the unique value satisfying equation (45), which is identical in form to the equation formed from (A).

### 2.3.1 Statement of Basic Assumptions

To simplify computations it is desirable to use the normalizing transformation, \( T_i = (X_i - \mu_{2i})/\sigma_i \), \( i=1,2 \), which, being a simple translation and change of scale, does not alter the values of the ratios, \( \mathbb{R}^2 \) and the ratio of the p.d.f.'s, and also does not change the rectangular truncation into a more complicated form of truncation. With no loss of generality, then, we make the following assumptions:

**BASIC ASSUMPTIONS:**

1.) \( \mathbb{X} \) has a \( N(\bar{\delta}, \mathbb{R}) \) distribution in Population 1 and a \( N(0, \mathbb{R}) \) distribution in Population 2, where

2.) \( \mathbb{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \) is positive definite.

3.) \( \alpha \) is the proportion of Population 1 in the mixture.

4.) An element \( x \) of the mixture cannot be observed unless, for a real number \( k, x \in A(x) \equiv \{ x : X_1 \leq k \} \).

5.) All parameters are known.

Under these assumptions, the optimal decision function is the linear function

\[
y = \bar{\delta}' R^{-1} X
\]

(29)

and the optimal decision rule is to choose Population 1 iff \( Y > k_1 \),

where

\[
k_1 = (1/2) \bar{\delta}' R^{-1} \bar{\delta} - \ln \left( \frac{\alpha \phi(k)}{1-\alpha \phi(k-\bar{\delta}_1)} \right).
\]

(30)
Similarly, the "best linear" decision function is

\[ Z = (\mathbf{H}^{-1} \hat{\delta} - \psi \omega_1)'X \]  

\[
\begin{align*}
\psi &= \frac{2\Delta - \delta_1(\omega_1+\omega_2)}{2 - (\omega_1+\omega_2)} \\
\Delta &= \xi(k-\delta_1) - \xi(k) \\
\xi(t) &= \frac{\phi(t)}{\phi(t)} \\
\phi(t) &= \frac{1}{\sqrt{2\pi}} \exp[-t^2/2] \\
\omega_i &= \xi(k_i)[k_i + \xi(k_i)] , i=1,2 \\
k_1 &= k - \delta_1 \\
k_2 &= k \\
\end{align*}
\]  

and the decision rule based on \( Z \) is to choose Population 1 iff \( Z > k_2 \), where \( k_2 \) is chosen to minimize the total probability of misclassification.

In order to compute \( k_2 \) and compare the two functions \( Y \) and \( Z \), it is necessary to know something about the properties of the functions \( \xi, \omega, \) and \( \psi \).

2.4 Analysis of the Functions \( \xi(\cdot), \omega(\cdot), \) and \( \psi(\cdot) \).

It follows from (29) and (31) that the difference between the two linear discriminant functions is \( Y-Z = \psi X_\delta \), where \( \psi \) is a rational function of \( \xi(\cdot) \) and \( \omega(\cdot) \), the arguments of which depend on the parameters \( k \) and \( \delta_1 \). Therefore, to compare \( Y \) and \( Z \) we must know some of the properties of the function \( \psi(\delta_1,k) \), which themselves depend on properties of the functions \( \xi(\cdot) \) and \( \omega(\cdot) \).

**Lemma II.6:** \( \psi(-\delta, k) = -\psi(\delta, k+\delta) \) for every \( \delta > 0 \).
proof: Define $\Delta(a, b) = \xi(a) - \xi(b)$.

Then from (32) we have $\psi(\delta, k) = \frac{2\Delta(k-\delta, k) - \delta[\omega(k-\delta) + \omega(k)]}{2 - [\omega(k-\delta) + \omega(k)]}$.

For any $\delta > 0$, then, $\psi(-\delta, k) = \frac{2\Delta(k+\delta, k) + \delta[\omega(k+\delta) + \omega(k)]}{2 - [\omega(k+\delta) + \omega(k)]}$.

Now, obviously, $\omega(k+\delta) + \omega(k) = \omega([k+\delta]-\delta) + \omega(k+\delta)$.

Also, $\Delta(k+\delta, k) = -\{\xi([k+\delta]-\delta) - \xi(k+\delta)\} = -\Delta([k+\delta]-\delta, k+\delta)$.

The last three statements combine to give the result. \qed

As a result of Lemma 6, we can, with no loss of generality, study $\psi(\delta, k)$ for positive values of $\delta$ only. Henceforth we shall make the following assumption:

ASSUMPTION II.6: $\delta_1 > 0$.

Lemmas 7-10, Theorem 4, and Corollary 2, stated below, have proved to be very useful, both in this chapter and in Chapter IV. Lemma 7 establishes the relationship between $\xi(\cdot)$ and Mills' ratio $R(\cdot)$, which is helpful inasmuch as Mills' ratio has been studied extensively in the literature.

LEMMA II.7: Let $R(t)$ denote Mills' ratio, i.e. $R(t) = \frac{1 - \Phi(t)}{\phi(t)}$ for $t > 0$. Then $R(t) = \frac{1}{\xi(-t)}$ for all $t > 0$.

proof: For $t > 0$, $\xi(-t) \equiv \frac{\Phi(-t)}{\phi(-t)} = \frac{\phi(t)}{1 - \phi(t)}$ by symmetry of the $N(0, 1)$ distribution. \qed

Lemma 8 gives the general properties of $\xi(\cdot)$.

LEMMA II.8: (i.) $\xi(t)$ is a strictly decreasing function of $t$.

(ii.) $\xi(t) \to +\infty$ as $t \to -\infty$. (iii.) $\xi(t) \to 0$ as $t \to +\infty$.

proof: (i.) For $t > 0$, $0 < \phi(t+\epsilon) < \phi(t) \leq \phi(0)$ for all $\epsilon > 0$, and $1/2 \leq \phi(t) < \phi(t+\epsilon) < 1$ for all $\epsilon > 0$.

Hence $\xi(t) > \frac{\Phi(t+\epsilon)}{\phi(t+\epsilon)} = \xi(t+\epsilon)$ for all $\epsilon > 0$.

For $t < 0$, we use equation 5.67 on p.137 of Kendall and Stuart [8]:
\[ R(|t|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp[-(tu)^2/2]}{1 + u^2} \, du, \]

which is clearly a decreasing function of \(|t|\). It follows from this and Lemma 7 that \(\xi(t)\) increases as \(t \to -\infty\).

(ii.) \(\lim_{t \to -\infty} \xi(t) = \lim_{t \to -\infty} \frac{\phi'(t)}{\phi(t)}\) by L'Hôpital's Rule

\[ = \lim_{t \to -\infty} \frac{-t\phi'(t)}{\phi(t)} = +\infty. \]

(iii.) \(\lim_{t \to +\infty} \xi(t) = \frac{\lim_{t \to +\infty} \phi(t)}{\lim_{t \to +\infty} \phi(t)} = \frac{0}{1} = 0. \] \(\Box\)

Lemma 9 gives an important relationship between \(\xi(\cdot)\) and \(\omega(\cdot)\).

**Lemma II.9:** \(\omega(t) = -\xi'(t)\), where \(\xi'(t) = \frac{d}{dt} \xi(t)\).

**proof:** \(\xi'(t) = \frac{d}{dt} \frac{\phi(t)}{\phi(t)}\), where \(\phi(t) = \int_{-\infty}^{t} \phi(u) \, du, \)

\[ = \frac{\phi' \phi - \phi \phi'}{\phi^2}, \]

where the arguments of the functions have been omitted for simplicity,

\[ = \frac{\phi(-t\phi) - \phi^2}{\phi^2} = -t\xi - \xi^2 = -\omega. \] \(\Box\)

Theorem 4 gives the general properties of \(\omega(\cdot)\) and also, in the course of the proof, establishes a limit that prove useful later on.

**Theorem II.4:** i.) \(\omega(t)\) is a strictly decreasing function of \(t\).

ii.) \(\omega(t) \to 1\) as \(t \to -\infty\). iii.) \(\omega(t) \to 0\) as \(t \to +\infty\).

iv.) \(\lim_{t \to +\infty} t\xi(t) = 0. \)

**proof:** Lemmas 8 and 9 together imply \(\omega(t) > 0\) for all real \(t\). (A)

(iii.) \(\lim_{t \to +\infty} \omega(t) = \lim_{t \to +\infty} t\xi(t) + \lim_{t \to +\infty} \xi^2(t)\) by definition of \(\omega(t)\).

a.) \(0 < \xi(t) \leq \xi(0) = 2\phi(0) < 0.8\) for all non-negative \(t\) implies \(0 \leq \lim_{t \to +\infty} \xi^2(t) \leq \lim_{t \to +\infty} \xi(t) = 0. \)

b.) \(\lim_{t \to +\infty} t\xi(t) = \frac{\lim_{t \to +\infty} t\phi'(t)}{\lim_{t \to +\infty} \phi(t)} = \lim_{t \to +\infty} \frac{t\phi(t)}{\exp[t^2/2]} = 0\) by L'Hôpital's Rule.
ii.) For $t < 0$, $\omega(t) = \frac{t}{R(|t|)} + \frac{1}{R^2(|t|)}$ by definition and Lemma 7.

Using the following lower bound for $R(\cdot)$, which was proposed by Birnbaum [4],

$$R(x) > \frac{2}{x + \sqrt{x^2+4}},$$

we get an upper bound of unity for $\omega(t)$ when $t < 0$; i.e.

$$\omega(t) < 1$$

for every $t < 0$.

Using the following upper bound for $R(x)$, which was proposed by Laplace [10],

$$R(x) \leq x^{-1} - x^{-3} + 3x^{-5},$$

we get the following lower bound for $\omega(t)$ when $t < 0$:

$$\omega(t) \geq 1 - \frac{t^6+7t^4-6t^2+9}{(t^4-t^2+3)^2},$$

which tends to 1 as $t \to -\infty$.

Therefore, $\omega(t) \to 1$ as $t \to -\infty$.

1.) To show that $\omega(t)$ is a strictly decreasing function of $t$, we prove that its derivative is negative for all real $t$.

Omitting arguments of the functions for simplicity, we have:

$$\omega' = \xi[1 + \xi'] + \xi'[t + \xi]$$

$$= \xi[1 - \xi(t + \xi)] - \xi(t + \xi)^2, \text{ since } \omega' = -\xi',$$

$$= \xi[1 - (t + \xi)(t + 2\xi)]. \quad (C)$$

Since $\xi(t) > 0$ for all real $t$, $\omega'(t)$ is positive or negative according as the second factor on the R.H.S. of ($C$) is positive or negative.

a.) For $t \geq 1$, $t+2\xi > t+\xi > t \geq 1$, so that $(t+\xi)(t+2\xi) > 1$. Therefore $\omega'(t) < 0$ for all $t \geq 1$.

b.) For $t < 0$, the inequality, $1 - [t+\xi(t)][t+2\xi(t)] < 0$, may be written equivalently as

$$2\xi^2(t) + 3t\xi(t) + (t^2-1) > 0. \quad (C')$$

Since $\xi(t)$ is a decreasing, non-negative function of $t$, the L.H.S. of ($C'$) is a quadratic in $\xi(t)$ which tends to $+\infty$ as $\xi(t)$ approaches zero or $+\infty$. Consequently, the zeros of the L.H.S. are

$$\xi(t) = \frac{-3t \pm \sqrt{t^2+8}}{4}.$$
Therefore (C') holds iff either of the following inequalities holds:

\[ \xi(t) > \frac{-3t + \sqrt{t^2+6}}{4}, \quad (D') \]

\[ \xi(t) < \frac{-3t - \sqrt{t^2+6}}{4}. \quad (E') \]

By Lemma 7, (D') holds iff (F') holds:

\[ R(|t|) < \frac{4}{3|t| + \sqrt{t^2+6}}. \quad (F') \]

Sampford [11] has proved that (F') holds. Therefore, \( \omega'(t) < 0 \) for all \( t < 0 \).

c.) For \( 0 \leq t < 1 \), we must resort to inelegant numerical inequalities; that is, we partition the interval and show that on each cell the product \([t+\xi(t)][t+2\xi(t)]\) exceeds unity.

For \( .75 \leq t < 1 \), \( t+2\xi(t) > t+\xi(t) > .75 + \xi(1) = .75 + .285 \).

For \( .50 \leq t < .75 \), \( t+\xi(t) > .50 + \xi(.75) = .50 + .385 = .885 \);
\[ t+2\xi(t) > .50 + 2(.385) = 1.27; \]
so their product exceeds 1.12.

For \( .375 \leq t < .5 \), \( t+\xi(t) > .375 + \xi(.5) = .375 + .5 = .875 \);
\[ t+2\xi(t) > .375 + 1.0 = 1.375; \]
so their product exceeds 1.20.

For \( .25 \leq t < .375 \), \( (t+\xi)(t+2\xi) > (.25+.57)(.25+1.14) = 1.13. \)
For \( .125 \leq t < .25 \), \( (t+\xi)(t+2\xi) > (.125+.64)(.125+1.28) = 1.07. \)
For \( 0 \leq t < .125 \), \( (t+\xi)(t+2\xi) > (0 + .715)(1.43) = 1.02. \)

COROLLARY II.1: \( \xi(t) > \frac{-3t + \sqrt{t^2+6}}{4} > 0 \) for all \( t < 1 \).

proof: In the proof of Part (b) of (i), it was shown (making no use of the condition that \( t < 0 \)) that \( \omega'(t) < 0 \) iff either (D') or (E') holds.

It was shown there that (D') holds for all negative \( t \).

From Theorem 4 we have \( \omega'(t) < 0 \) for all real \( t \).

Hence, for every value of \( t \) either (D') or (E') holds.

Now, (E') does not hold for \( t < 0 \), and it cannot hold for \( t \geq 0 \), as the R.H.S. is negative for all non-negative \( t \), while \( \xi(t) \) is strictly positive.

Consequently, (D') holds for all \( t \).

Unfortunately, this bound is negative for \( t > 1 \) and so is trivial as a lower bound for the positive function \( \xi(t) \).
As the following fact is used frequently, it is convenient to state it as a corollary.

**COROLLARY II.2**: \( 0 < \omega(t) < 1 \) for all real \( t \).

**proof**: Follows directly from the definitions of "strictly decreasing" and "limit". \( \square \)

Information about the comparative values of \( \xi \) at neighboring points is very useful.

**LEMMA II.10**: \( \xi(t-a) - \xi(t) < a \) for all \( t \) and for all \( a > 0 \). Furthermore, \( \lim_{t \to -\infty} [\xi(t-a) - \xi(t)] = a \) for all real \( t \) and all real \( a \).

**proof**: Let \( a \) be any positive number. For any real number \( t \), then,
\[
\xi(t-a) - \xi(t) = \int_{t-a}^{t} \xi'(u) \, du = \int_{t-a}^{t} \omega(u) \, du \quad \text{by Lemma 9}
\]
\[
< \int_{t-a}^{t} 1 \, du = a, \quad \text{by Corollary 2.}
\]

For any real \( t \) and \( a \), \( \xi(t-a) = \xi(t) - a\xi'(t^*) \), where \( t^* \) is between \( t \) and \( (t-a) \).

So \( \xi(t-a) - \xi(t) = a\omega(t^*) \), which tends to \( a \) as \( t \to -\infty \), by Theorem 4. \( \square \)

An important relationship between \( \delta_1 \) and \( \psi \) is established in the following theorem.

**THEOREM II.5**: \( \psi(\delta, k) < \delta \) for all \( k \) and for all positive \( \delta \).

**proof**: Suppose \( \delta \leq \psi(\delta, k) \) for some choice of \( \delta > 0 \) and \( k \). Then
\[
\delta \leq \frac{2[\xi(k-\delta) - \xi(k)] - \delta[\omega(k-\delta) + \omega(k)]}{2 - [\omega(k-\delta) + \omega(k)]} \quad \text{(i)}
\]

The following little algebraic lemma is easy to prove: If \( A > 0 \) and \( A + B > 0 \), then \( Y \leq X \iff Y \leq \frac{AX + BY}{A + B} \).

In (i) let \( A = 2 \) and \( B = -[\omega(k-\delta) + \omega(k)] \).
Then \( A + B > 2 - (1+1) = 0 \) by Corollary 2.

By the algebraic lemma, then, (i) holds iff (ii) holds:
\[
\delta \leq \xi(k-\delta) - \xi(k). \quad \text{(ii)}
\]
But (ii) contradicts Lemma 10, and so our supposition is wrong. □

2.5 Computation of the Critical Value \( k_2 \).

Recall the fact that \( k_2 \) is defined to be the value of \( k \) which minimizes the total probability of misclassification among all the decision rules of the form, "Choose Population 1 iff \( Z > k \)."

A procedure for calculating \( k_2 \) is obtained in the following subsections. The key steps in the development of the procedure are:

1.) to differentiate the total probability of misclassification, considered as a function of \( k \), in order to obtain an equation which \( k_2 \) must satisfy;

2.) to prove that this equation has exactly one root, i.e. \( k = k_2 \), and that \( k_2 \) does indeed minimize the total probability of misclassification.

It is also shown that the optimal decision rule based on the linear function \( Z \) necessarily partitions the real line into the same two half-lines as does the above decision rule; but it is not established that the two decision rules agree in their assignment of the decision regions.

2.5.1 The Equation Which \( k_2 \) Must Satisfy.

From Theorem 2 we have the joint distributions of \( Y \) and \( Z \) for \( i=1,2 \):

\[

f_{1|y}(y|k) = \begin{cases} 
\frac{\exp[-(y-y_1)'\Sigma^{-1}(y-y_1)/2]}{2\pi^{1/2}|\Sigma|^{1/2} \phi(k_1)} & \text{if } y \in A(y), \\
0 & \text{otherwise} 
\end{cases} \quad (33)

\]

where

\[

\begin{align*}
\varphi' &= (Y,Z) \\
A(y) &= \{y: (1/\psi)(Y-Z) \leq k\} \\
C' &= (R^{-1} \delta, R^{-1} \delta - \psi u_1) \\
\Theta &= CRC' \\
\varphi_1' &= \delta' C' \text{ and } \varphi_2 = 0
\end{align*}
\]

\]
Note that in $A(\nu)$ \[ \begin{cases} Y \leq Z + k\psi & \text{if } \psi > 0 \\ Y \geq Z + k\psi & \text{if } \psi < 0 \end{cases} \].

Observe, too, that there are several relationships among the elements of $\Theta$ and $\nu_1$:

$$\left\{ \begin{array}{l}
a = \frac{\delta_2}{\gamma} = (1-\rho^2)^{-1}(\delta_1^2-2\rho\delta_1\delta_2+\delta_2^2), \\
b = \psi \delta_1, \\
c = \psi^2, \\
\end{array} \right. \quad (35)$$

Then $\Theta = \mathcal{G}_c' = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{12} & \theta_{22} \end{bmatrix} = \begin{bmatrix} a & a-b \\ a-b & a-2b+c \end{bmatrix}$, \quad (36)

and $\nu_1' = \delta'G = (a, a-b) = (\theta_{11}, \theta_{12})$. \quad (37)

Finally, observe that for $\nu \in A(\nu)$, the function $\phi(k_1) F_1(z|k)$ is equal to the p.d.f. corresponding to the $N(\nu_1, \Theta)$ distribution, which can be factored into the product of the p.d.f.'s corresponding to the $N(\nu_{12}, \Theta_{22})$ distribution and the

$$N\left\{ \nu_{11} + \frac{\theta_{12}}{\theta_{22}} (z-\nu_{12}), \frac{\theta_{11}\theta_{22}-\theta_{12}^2}{\theta_{22}} \right\}$$

distribution. Therefore, letting

$$A = \frac{\theta_{12}}{\theta_{22}} \quad \text{and} \quad s^2 = \frac{\theta_{11}\theta_{22}-\theta_{12}^2}{\theta_{22}}$$ \quad (38)

we can compute the probability of misclassification for the decision rule "Choose Population 1 iff $Z > k$":

$$F(k) = \begin{cases} a\text{Pr}_1(YZ+\kappa, Z<\kappa|k) + (1-a)\text{Pr}_2(YZ+\kappa, Z>\kappa|k) & \text{if } \psi > 0, \\
a\text{Pr}_1(YZ+\kappa, Z<\kappa|k) + (1-a)\text{Pr}_2(YZ+\kappa, Z>\kappa|k) & \text{if } \psi < 0, \\
\end{cases}$$

$$= \begin{cases} \frac{\alpha}{\phi(k_1)} \int_{-\infty}^{(k-\theta_{12})/\sqrt{\theta_{22}}} \phi(t) \left( [1-A](\sqrt{\theta_{22}}t+\theta_{12}) + k\psi-\theta_{11}+\theta_{12} \right) / s \ dt + \\
+ \frac{1-\alpha}{\phi(k)} \int_{k/\sqrt{\theta_{22}}}^{\infty} \phi(t) \left( [1-A](\sqrt{\theta_{22}}t+k\psi) / s \right) dt & \text{if } \psi > 0, \\
\frac{\alpha}{\phi(k_1)} \int_{-\infty}^{(k-\theta_{12})/\sqrt{\theta_{22}}} \phi(t) \left( [1-A](\sqrt{\theta_{22}}t+\theta_{12}) + k\psi-\theta_{11}+\theta_{12} \right) / s \ dt + \\
+ \frac{1-\alpha}{\phi(k)} \int_{k/\sqrt{\theta_{22}}}^{\infty} \phi(t) \left( [1-A](\sqrt{\theta_{22}}t+k\psi) / s \right) dt & \text{if } \psi < 0. \\
\end{cases} \quad (39)$$
These expressions can be simplified considerably. For this purpose and for later computations the following lemma is helpful.

**LEMMA II.11:** These identities hold: i.) \( \theta_{22} - \theta_{12} = \psi(\psi - \delta_1) \).
ii.) \( \theta_{11} \theta_{22} - \theta_{12}^2 = \psi^2 (a - \delta_1^2) \).
iii.) \( a - \delta_1^2 = (1 - \rho^2)^{-1}(\rho \delta_1 - \delta_2)^2 \).
iv.) \( \theta_{22} = (a - \delta_1^2) + (\delta_1 - \psi)^2 \).

**proof:** Follows directly from equations (35) and (36).

Applying (38) and Lemma 11 to the arguments of the \( \Phi \)-functions in (39), we can write:

\[
\text{a.) } s^2 = \frac{\psi^2}{1 + A^* Z} \\
\text{where } A^* = \frac{\delta_1 - \psi}{\sqrt{a - \delta_1^2}} = \frac{(\delta_1 - \psi) \sqrt{1 - \rho^2}}{\rho \delta_1 - \delta_2},
\]

provided \( \delta_2 \neq \rho \delta_1 \).

\[
\text{b.) } \frac{1 - A}{s} = \frac{-A^*}{\sqrt{\theta_{22}}} \text{ sgn } \psi.
\]

\[
\text{c.) } \frac{k \psi}{s} = k \sqrt{1 + A^* Z} \text{ sgn } \psi.
\]

\[
\text{d.) } \frac{\theta_{12} - \theta_{11}}{s} = -\delta_1 \sqrt{1 + A^* Z} \text{ sgn } \psi.
\]

Observing that, in (39), when \( \psi < 0 \) the argument of \( \Phi(\cdot) \) is the negative of that for \( \psi > 0 \), we can write \( F(k) \) simply as:

\[
F(k) = \frac{\alpha}{\Phi(k - \delta_1)} \int_{-\infty}^{(k - \theta_{12})/\sqrt{\theta_{22}}} \phi(t) \phi(-A^* t + (k - \delta_1) \sqrt{1 + A^* Z}) \, dt + \]

\[
+ \frac{1 - \alpha}{\Phi(k)} \int_{k/\sqrt{\theta_{22}}}^{\infty} \phi(t) \phi(-A^* t + k \sqrt{1 + A^* Z}) \, dt,
\]

provided \( \delta_2 \neq \rho \delta_1 \). As each term on the R.H.S. of (44) can be differentiated with respect to \( k \) repeatedly, we have:

\[
F'(k) = \frac{1}{\sqrt{\theta_{22}}} \left\{ \frac{\alpha}{\Phi(k - \delta_1)} \phi \left( \frac{k - \theta_{12}}{\sqrt{\theta_{22}}} \right) \phi \left( -A^* \frac{k - \theta_{12}}{\sqrt{\theta_{22}}} + (k - \delta_1) \sqrt{1 + A^* Z} \right) + \right. \]

\[
- \frac{1 - \alpha}{\Phi(k)} \phi \left( \frac{k}{\sqrt{\theta_{22}}} \right) \phi \left( -A^* \frac{k}{\sqrt{\theta_{22}}} + k \sqrt{1 + A^* Z} \right) \}
\]

and therefore \( F'(k) = 0 \) iff \( k = k_2 \), where
\[
\phi(k) \frac{\phi\left(\frac{k_2 - \delta_{12}}{\sqrt{\theta_{22}\epsilon}}\right)}{\phi\left(\frac{-A^*}{\sqrt{\theta_{22}\epsilon}} k_2 + k\sqrt{1 + A^{*2}} - \frac{\psi}{\sqrt{1 + A^{*2}}}\right)} = \frac{1-a}{a}, \quad (45)
\]

since from (38), (40), (42), and (43) we have

\[
\left(\frac{\theta_{12}}{\sqrt{\theta_{22}\epsilon}} - \delta_{1}\sqrt{1 + A^{*2}}\right) \text{sgn} \psi = \frac{-|\psi|}{\sqrt{1 + A^{*2}}}.
\]

These results can be summarized as follows:

**THEOREM II.6:** If \( P(k) \) denotes the total probability of misclassification for the decision rule, "Choose Population 1 iff \( Z > k \)", and if \( \delta_2 = \rho \delta_1 \), then the derivative \( P'(k) \) vanishes iff \( k = k_2 \), where \( k_2 \) satisfies equation (45) above with

\[
\begin{align*}
A^* &= \frac{\delta_{1}-\psi}{\sqrt{A-\delta_1^2}} = \frac{(\delta_{1}-\psi)\sqrt{1-\rho^2}}{|\delta_2-\rho \delta_1|} > 0; \\
\theta_{22} &= (a-\delta_1^2) + (\delta_{1}-\psi)^2 = \frac{(\delta_2-\rho \delta_1)^2}{1-\rho^2} + (\delta_{1}-\psi)^2 > 0; \\
\theta_{12} &= (a-\delta_1^2) + \delta_1(\delta_{1}-\psi) > 0.
\end{align*}
\]

When \( \delta_2 = \rho \delta_1 \), there is a special relationship between \( \delta \) and \( R \):

\( \delta' = \delta(1, \rho) \) with \( \delta > 0 \). Then the ODF is \( Y = \delta X_1 \) and the BLDF is \( Z = (1 - \psi/\delta)Y \). By Theorem 5, \( \psi < \delta_1 \) and so \( Z \) is directly proportional to \( Y \). Consequently, the probability of misclassification for the rule based on \( Z \) is minimized when \( k_2 = (1 - \psi/\delta)k_1 \).

**LEMMA II.12:** For the special case when \( \delta_2 = \rho \delta_1 \), the decision rule based on the BLDF \( Z \) is equivalent to the optimal decision rule.

\[ \square \]

2.5.2 The Relationship Between \( k_2 \) and the Critical Value of the Optimal Decision Rule Based on \( Z \).

In Section 4.1 it is shown that the optimal decision rule based
on the linear function \( Z = d_1X_1 + d_2X_2 \equiv d'X \) is to choose Population 1

\[
\frac{\phi(k) \sqrt{\theta_1} \exp[-(Y-\nu)^2/2\theta_1] \phi(a_1-b_1Y)}{\phi[k-\delta_1] \sqrt{\theta_1} \exp[-Y^2/2\theta_2] \phi(a_2-b_2Y)} > \frac{1-\alpha}{\alpha},
\]

(1)

where \( Y = d^{-1}_2Z \) when \( d_2 \neq 0 \), [N.B.: in this subsection, \( Y \) does not denote the ODF] and where, for the special case when both populations have the same covariance matrix \( R \) and the linear function is \( Z = (R^{-1}\delta - \psi u_1)'X \), the other parameters in (i) are defined as follows:

\[
\begin{align*}
(d', \nu, \theta_1, \theta_2, a_1, a_2, d, b_1, b_2) &= (R^{-1}\delta - \psi u_1)' = \left( \frac{\delta_1-\rho \delta_2}{1-\rho^2}, -\frac{\delta_2-\rho \delta_1}{1-\rho^2} \right) ; \\
\nu &= d_2^{-1} d' \delta = d_2^{-1}(d' R^{-1} \delta - \psi \delta' u_1) ; \\
\theta_1 = \theta_2 &= d_2^{-2}(d' R d) = d_2^{-2}(d' R^{-1} \delta - 2\psi \delta' u_1 + \psi^2) \equiv \theta ; \\
\left( ii \right) a_1 &= \frac{k\theta - d_2^{-1} d' R(u_2, -u_1) \delta}{\sqrt{1-\rho^2} \sqrt{\theta}} = \frac{k\theta - d_2^{-1} \psi (\delta_2-\rho \delta_1)}{\sqrt{1-\rho^2} \sqrt{\theta}} = \frac{k\theta - \psi (1-\rho^2)}{\sqrt{1-\rho^2} \sqrt{\theta}} ; \\
a_2 &= \frac{k\sqrt{\theta}}{\sqrt{1-\rho^2}} ; \\
d &= \frac{d_1}{d_2} = \frac{(\delta_1-\rho \delta_2) - \psi (1-\rho^2)}{\delta_2-\rho \delta_1} ; \\
b_1 = b_2 &= b = \frac{d + \rho}{\sqrt{1-\rho^2} \sqrt{\theta}} .
\end{align*}
\]

Conceivably, this optimal decision rule based on \( Z \) could partition the real line into three or more cells; e.g. "Choose Population 1 iff \( z_1 < Z < z_2 \)", where \( z_1 \) and \( z_2 \) are finite numbers such that for \( i=1,2, \)

\( z_i = d_2 y_i \) and \( y_i \) is a root of the following equation:

\[
\frac{\phi(k) \exp[-(y-\nu)^2/2\theta] \phi(a_1-b_1y)}{\phi(k-\delta_1) \exp[-y^2/2\theta] \phi(a_2-b_2y)} = \frac{1-\alpha}{\alpha} .
\]

(iii)

No matter how many critical values there are for the optimal decision rule based on \( Z \), all of them must satisfy the conditions that \( z_i = d_2 y_i \) and \( y_i \) is a root of (iii).

Now, equation (45) stated in the last subsection is the equation which must be satisfied by the critical value for the decision rule,
"Choose Population 1 iff Z > k_2". Since equations (45) and (iii) are identical in form, let us see whether their arguments correspond. In the notation defined in (35) and (36) we can rewrite \( \nu \) and \( \theta \):

\[
\begin{align*}
\nu &= (a-b)/d_2 = \theta_{12}/d_2 \\
\theta &= (a-2b+c)/d_2^2 = \theta_{22}/d_2^2
\end{align*}
\]

Therefore, the ratio of exponential functions in (iii) is equal to the ratio of \( \phi \)-functions in (45). Moreover, from (41), (ii), (iv), and Lemma 11 we have:

\[
\begin{align*}
a_2 &= k\sqrt{1 + A^2Z} \\
a_1 &= k\sqrt{1 + A^2Z} - \frac{\psi}{\sqrt{1 + A^2Z}} \\
b &= d_2 \frac{A^2}{\sqrt{\theta_22}}
\end{align*}
\]

so that the ratio of \( \phi \)-functions in (iii) equals the ratio of \( \phi \)-functions in (45), since \( y = z/d_2 \).

Thus equations (iii) and (45) are equivalent. It is shown in the next subsection that (45) has exactly one root. Thus the optimal decision rule also has exactly one critical value and so cannot partition the real line into more than two cells. However, we cannot conclude, from the fact that both decision rules has the same critical value, that they are the same decision rule: it is possible that, for certain parametric combinations, the optimal decision rule is to choose Population 1 iff \( Z < k_2 \). What the optimal decision rule is depends on the behavior of the L.H.S. of (iii) considered as a function of \( Y = Z/d_2 \). An investigation of this function as such has not been made in this dissertation.

2.5.3 Existence and Uniqueness of \( k_2 \).

Theorem 6 gives a necessary and sufficient condition for the derivative \( F'(k) \) to vanish, but it doesn't prove that any root of (45) also
minimises the total probability of misclassification $F(k)$. This problem will now be considered.

**Lemma II.13:** Equation (45) can be rewritten as follows:

$$f(k_2) = k_1 - E^2/2,$$

(46)

where $k_1$ is defined as in equation (30),

$$f(t) = (1+B)t + \ell_n g(t^*) ,$$

(47)

$$g(t) = \frac{\phi(t+E)}{\phi(t)} ,$$

(48)

$$t^* = Ct + D ,$$

(49)

and

$$B = [\psi(\delta_1) - \psi]/\theta_{22} ,$$

$$C = -A^*/\sqrt{\theta_{22}} ,$$

$$D = k\sqrt{1+A^2},$$

$$E = -\psi(1)/\sqrt{1+A^2}$$

(50)

with $A^*$ and $\theta_{22}$ defined as in Theorem 6. Then $C < 0$ and $BE < 0$.

**Proof:** For simplicity, let $t$ denote $k_2$ in equation (45).

Then take the natural logarithm of both sides of (45) and apply definitions (50) to get the equivalent equation:

$$\frac{\theta_{12}}{\theta_{22}} t + \ell_n g(t^*) = k_1 + \frac{1}{2} \left( \frac{\theta_{12}^2}{\theta_{22}} - \delta' R^{-1} \delta \right).$$

(1)

From equations (35), (36), and (40) we get:

$$\begin{cases}
\frac{\theta_{12}}{\theta_{22}} = 1 + B , \\
\frac{\theta_{12}^2}{\theta_{22}} - \delta' R^{-1} \delta = -E^2
\end{cases}$$

(II)

Substituting (II) into (I) gives (46).

It follows immediately from Theorem 6 that $C < 0$.

Since $\delta_1 > \psi$, $B$ has the same sign as $\psi$ and $E$ has the opposite sign.

□

Lemmas 14-15 establish some useful properties of the function $g(\cdot)$.

**Lemma II.14:** For any $E > 0$, (i) $g(t)$ is a strictly decreasing function of $t$ which tends to $+\infty$ as $t \to -\infty$ and (ii) tends to 1 as
t tends to $+\infty$.

**proof:** (i) Suppose $g(t)$ is not decreasing at some point $t^0$.

Then the derivative of $g(t)$ is non-negative at that point:

$$
\frac{\phi(t^0) \phi(t^0+E) - \phi(t^0+E) \phi(t^0)}{\phi^2(t^0)} \geq 0.
$$

But this implies that $\xi(t^0+E) \geq \xi(t^0)$, where $E > 0$, and this contradicts Lemma 8.

(ii) As $t$ tends to $-\infty$,

$$
\lim g(t) = \lim \exp[-(2Et + E^2)/2] = +\infty
$$

by L'Hôpital's Rule and the fact that $\phi'(t) = \phi(t)$.

(iii) As $t$ tends to $+\infty$, $\lim g(t) = 1$ because $\lim \phi(t) = 1$.

□

**LEMMA II.15:** For any $E < 0$, (i) $g(t)$ is a strictly increasing function of $t$ which (ii) tends to zero as $t$ tends to $-\infty$, and (iii) tends to 1 as $t$ tends to $+\infty$.

**proof:** Define $t^0 = t + E$ and $E^0 = -E > 0$. Then

$$
g(t) = \frac{\phi(t^0)}{\phi(t^0+E^0)} = \frac{1}{g(t^0)}.
$$

Then all results follow directly from Lemma 14.

□

Lemmas 16-17 help to establish the fact that the coefficient of $t$ in $f(t)$ is positive.

**LEMMA II.16:** If $B$, $C$, and $E$ are defined as in Lemma 13, then $B = CE$.

**proof:** Follows directly from the definitions and the fact that

$$
\theta_{22}(1+A^2) = \frac{\theta_{22}^2}{a-\delta_1^2}.
$$

□

**LEMMA II.17:** If $\psi < 0$, then $-1 < B < 0$.

**proof:** Assumption 6 and $\psi < 0$ imply that $B \equiv [\psi(\delta_1-\psi)]/\theta_{22} < 0$.

Suppose $B \leq -1$. This implies that
\[ 0 \geq \frac{a-\delta_1^2}{\delta_1-\psi} + \delta_1, \]

which contradicts Assumption 6 or Lemma 11-(iii).

Lemma 18 shows that \( f(t) \) is a strictly increasing function of \( t \).

**LEMMA II.18:** Let \( f'(t) \) denote the derivative with respect to \( t \) of the function \( f(t) \) defined in (47). Then

(i) if \( E < 0 \), then \( f'(t) > 1 \).

(ii) If \( E > 0 \), then \( 0 < f'(t) < 1 \).

**proof:** By differentiating and collecting terms we get

\[ f'(t) = (1+B) + C[\xi(t^*+E) - \xi(t^*)]. \quad (52) \]

(i) When \( E < 0 \), then by Lemmas 8 and 10

\[ 0 < \xi(t^*+E) - \xi(t^*) < -E; \]

and since, by Lemma 13, \( C < 0 \), this gives

\[ 0 > C[\xi(t^*+E) - \xi(t^*)] > -CE = -B \]

by Lemma 16. Substitution into (52) gives the result.

(ii) When \( E > 0 \), then by the same line of reasoning we have \( f'(t)<1 \).

Now, suppose \( f'(t) \leq 0 \) for some value of \( t \). Then by (52) and the above results we have

\[ 1 + C[\xi(t^*+E) - \xi(t^*)] \leq |B|, \]

that is, \( |B| > 1 \). But this contradicts Lemma 17.

The next theorem establishes the existence and uniqueness of the critical value which minimizes the total probability of misclassification for all decision rules of the form, "Choose Population 1 iff \( Z > k \)."

**THEOREM II.7:** Given any set of parameters such that \( \delta_2 \neq \rho \delta_1 \) and \( \psi(\delta_1,k) \neq 0 \), there exists one and only one value \( k_2 \) which satisfies equation (45), that is, which causes \( F'(k) \) to vanish, and this value minimizes \( F(k) \).

**proof:** (1) By Lemma 18, \( f(t) \) is a strictly increasing function of \( t \) for all non-zero \( E \), that is, for all \( \psi \neq 0 \).

(2) If \( \psi < 0 \), then \( f(t) \to \pm \infty \) as \( t \to \pm \infty \). For \( \psi < 0 \) implies
E > 0 and, from Lemma 17, 1 + B > 0.
E > 0 implies, from Lemma 14, that g(t*) is a decreasing function of t* and hence an increasing function of t.
Therefore \( f(t) = (1 + B)t + \ln g(t^*) + \tau \) as \( t \to \pm \infty \).

(3) If \( \psi > 0 \), then \( f(t) \to \pm \infty \) as \( t \to \pm \infty \). For \( \psi < 0 \) implies \( E < 0 < B \).
Now, \( E < 0 \) implies \( \ln g(t^*) \to 0 \) as \( t \to -\infty \), by Lemma 15.
By Lemma 16, \( f'(t) > 1 \). So
\[
\int_0^t f'(u) \, du > t.
\]
Since \( f(0) = \ln g(D) \) is a fixed, finite number for all parameter combinations such that \( \delta_2 = \rho \delta_1 \), it follows that \( f(t) \to +\infty \) as \( t \to +\infty \).

(4) Since \( f(t) \) is a strictly increasing, continuous function whose range is the real numbers, there exists exactly one value of \( t \) for which
\[
f(t) = k_1 - E^2/2,
\]
which is a fixed real number for a given set of parameters.

(5) Since \( f(t) \) increases as \( t \) increases, \( f(t) \) is less than or greater than
\[
k_1 - E^2/2
\]
according as \( t \) is less than or greater than \( k_2 \), the unique number satisfying \((45)\).
From \((44-A)\) and the proof of Lemma 13 we see that \( F'(k) \) is positive or negative according as \( f(k) - (k_1 - E^2/2) \) is positive or negative.
Therefore, \( F(k) \) is a decreasing function for all \( k < k_2 \) and an increasing function for all \( k > k_2 \).
That is, \( F(k) \) is minimum when \( k = k_2 \).

\[\square\]

2.5.4 Computation of \( k_2 \).

For a given set of parameters \( \pi' = (\delta_1, \delta_2, \rho, \alpha, k) \), equation \((45)\) must be solved for \( k_2 \). In practice, it is easier to choose various values of \( k_2 \) and solve \((45)\) for the corresponding values of \( \alpha \). Thus \((45)\) can be written:
\[
\frac{C_1}{C_2} = \frac{1-\alpha}{\alpha},
\]
which has the solution:
\[ \alpha = \frac{C_2}{C_1+C_2}. \]

A program can be written to do these calculations in conversational mode (e.g. in APL). It would then take only a few minutes to find the value of \( k_2 \) which corresponds closely enough to the given \( \alpha \).

2.6 Location of the Point \((k_1, k_2)\).

In order to compare the performances of the two decision rules, especially to compute their probabilities of misclassification and the probabilities associated with events such as "the decision rule based on \( Z \) is correct but the ODR is incorrect", it is necessary to know whether the point \((k_1, k_2)\) lies in the region of positive probability. We assume, of course, that there are two distinct decision rules, i.e. that \( \psi = 0 \) and that \( \delta_2 = \rho \delta_1 \).

Recall, from Section 2.2, that \((k_1, k_2)\) is in the region of positive probability iff \[ (1/\psi)(k_1 - k_2) \leq k. \]

In this section sufficient conditions for this inequality to hold are obtained.

**Lemma II.19:** Considered as a function of \( k_2 \), the quantity \( k_1-k_2 \) is a strictly decreasing function of \( k_2 \) if \( \psi<0 \) and a strictly increasing function of \( k_2 \) if \( \psi>0 \).

**Proof:** By (46) \[ k_1-k_2 = f(k_2) - k_2 + E^2/2. \]

Denoting \( k_1-k_2 \) as \( h(k_2) \), we have \[ h'(k_2) = f'(k_2) - 1, \]

which Lemma 18 shows is positive or negative according as \( E \) is negative or positive, i.e. according as \( \psi \) is positive or negative.

**Lemma II.20:** \( k_2 \) is a strictly increasing function of \( k_1 \).

**Proof:** Theorem 7 says that for fixed \( k_1 \), \( k_2 \) is the unique value
such that \( f(k_2) = k_1 - E^2/2 \).

From Lemma 18, \( f'(k_2) > 0 \) for all \( k_2 \).

Therefore \( f(k_2) \) is a 1:1 mapping and consequently has an inverse mapping,

\[
k_2 = f^{-1}(k_1 - E^2/2),
\]

which is a function.

The two relations,

\[
k_1 = f(k_2) + E^2/2 \quad \text{and} \quad f'(k_2) > 0 \quad \text{for all} \quad k_2,
\]

imply that \( k_1 \) is a strictly increasing function of \( k_2 \).

Since \( f^{-1} \) is a 1:1 mapping, then, \( k_2 \) is a strictly increasing function of \( k_1 \).

\[
\square
\]

**Lemma II.21:** For any \( E < 0 \), there exists a number \( t(E) < -1 \) such that

\[
\frac{\phi(t+E)}{\phi(t)} < \exp(|E|t - E^2/2)
\]

for every \( t < t(E) \).

**Proof:** For \( E < 0 \) and \( t < 0 \) we have, from Lemma 7,

\[
\frac{\phi(t+E)}{\phi(t)} = \frac{\xi(t)}{\xi(t+E)} \frac{\phi(t+E)}{\phi(t)} = \frac{R(|t+E|)}{R(|t|)} \frac{\phi(t+E)}{\phi(t)}.
\]

For \( t < -1 \), we use bounds for Mills' ratio derived by Laplace [10] to get this inequality:

\[
g(t) = \frac{\phi(t+E)}{\phi(t)} < \frac{|t+E|^{-1}}{|t|^{-1} - |t|^{-3}} \frac{\phi(t+E)}{\phi(t)}
\]

\[
= \frac{|t|^3}{(t^2-1)|t+E|} \exp[-(Et + E^2/2)], \quad (i)
\]

Now, that coefficient of the exponential function in (i) is \( \leq 1 \)

iff \[ |t|^3 \leq (t^2-1)|t+E| \],

iff \[ -Et^2 + t + E \geq 0. \] \quad (ii)

Since \( E < 0 \), the quadratic on the L.H.S. of (ii) is concave; that is, it will be positive for all \( t \) less than some critical value \( t(E) \), which is the smaller zero of the L.H.S. of (ii).

Applying this result to (i), we see that for \( t < t(E) \)

\[
g(t) < \exp(|E|t - E^2/2).
\]

\[
\square
\]

**Lemma II.22:** For any \( E > 0 \), there exists a number \( t(E) < -1 \) such
that for all $t < t(E)$, $\frac{\phi(t+E)}{\phi(t)} > \exp(E|t| - E^2/2)$.

**Proof:** Let $t^0 = t+E$ and $E^0 = -E < 0$.

Then

$$\frac{\phi(t+E)}{\phi(t)} = \frac{\phi(t^0)}{\phi(t+E^0)}.$$  

From Lemma 21 we know that there exists a number $t(E^0) < -1$ such that for every $t < t(E^0)$,

$$\frac{\phi(t)}{\phi(t+E^n)} > \exp(-|E^0|t + E^0^2/2) > \exp(-|E^n|t - E^n^2/2) = \exp(E|t| - E^2/2).$$

Noting that $t^0 < t(E^0)$ implies $t < t(-E) - E < -1$, we have the desired result.

$\square$

**Theorem II.8:** The point $(k_1, k_2) \in A(\psi)$, the region of positive probability, if (i) $\psi > 0$ or (ii) $\psi < 0$ and $k \geq 0$.

**Proof:** By Lemma 13 we can write

$$k_1 - k_2 = E + \ln g(k^2) + E^2/2,$$  \hspace{1cm} (a)

with $k_2^* = Ck_2 + D$, and the other quantities defined in (50).

Note that $C < 0$ implies that $k_2^* \to -\infty$ as $k \to -\psi$.

1. If $\psi > 0$, then $B < C$ and $A(\psi) = \{(Y, Z): Y - Z \leq k\psi\}$.

By Lemma 21, $B < C$ implies that for $k_2^*$ sufficiently small (i.e., for $k_2$ sufficiently large)

$$g(k_2^*) < \exp(|E|k_2^* - E^2/2).$$  \hspace{1cm} (b)

Statements (a) and (b) together imply that for $k_2$ sufficiently large,

$$k_1 - k_2 = E + (|E|k_2^* - E^2/2) + E^2/2$$

$$= (B-CE)k_2 + ED \quad \text{by definition of $k_2^*$},$$

$$= k_2 \psi \quad \text{by Lemma 16 and (50)}.$$

Since, by Lemma 19, $k_1 - k_2$ is a strictly increasing function of $k_2$ when $\psi > 0$, we have

$$k_1 - k_2 < k_2 \psi \quad \text{for all $k_2$},$$

that is, $(k_1, k_2) \in A(\psi)$ for all values of $k_2$.

2. If $\psi < 0$, then $B < C$ and $A(\psi) = \{(Y, Z): Y - Z \geq k\psi\}$.

From Lemma 22 it follows that for $k_2^*$ sufficiently small (i.e., for $k_2$ sufficiently large)

$$g(k_2^*) > \exp(|E|k_2^* - E^2/2).$$  \hspace{1cm} (c)

Statements (a) and (c) together imply that for $k_2$ sufficiently large,

$$k_1 - k_2 > E + E|Ck_2 + D|.$$  \hspace{1cm} (c)
If we assume that \( k \geq 0 \), then \( D = k\sqrt{1 + A^2} \geq 0 \).

Since (c) holds subject to the condition that \( k_2^* < t(E) < -1 \), we have

\[
Ck_2 < t(E) - D,
\]

and so, for \( k_2 \) sufficiently large,

\[
|Ck_2 + D| = |Ck_2| - D = -Ck_2 - D.
\]

Substituting this into (d), we get, by the same reasoning as above,

\[
k_1 - k_2 > k_\psi.
\]

Because, by Lemma 19, \( k_1 - k_2 \) is a strictly decreasing function of \( k_2 \) when \( \psi < 0 \), this proves that \( (k_1, k_2) \in A(\omega) \) for all values of \( k_2 \).

\[\square\]

Because of its necessity in the preceding proof and because it is a reasonably nonrestrictive assumption to make (basically, it requires us to observe no less than half of the possible values in Population 2), we shall henceforth assume:

ASSUMPTION II.7: \( k \geq 0 \).

2.7 Formulas for Computing the Probabilities of Agreement and Disagreement.

It would be of interest to compare the performances of the decision rules based on the two linear discriminant functions, \( Y \) and \( Z \). One way to measure their comparative performances for a specified set of parameters is to compute the probabilities associated with the regions of the \( X \)-plane in which the decision rules agree and the regions in which they disagree. Consider the following partition of \( A(\omega) \), the region of positive probability:

\[
\begin{align*}
\{a\} &= \text{the set of points which } Y \text{ and } Z \text{ both assign to the } N(\delta, \Theta) \text{ population;} \\
\{b\} &= \text{the set of points which } Y \text{ and } Z \text{ both assign to the } N(\Omega, \Theta) \text{ population;} \\
\{c\} &= \text{the set of points which } Y \text{ assigns to } N(\delta, \Theta) \text{ but } Z \text{ assigns to } N(\bar{\Omega}, \Theta). \\
\{d\} &= \text{the set of points which } Y \text{ assigns to } N(\bar{\Omega}, \Theta) \text{ but } Z \text{ assigns to } N(\delta, \Theta).
\end{align*}
\]
THEOREM II.9: The probabilities, $Pr_i((Y,Z) \in \{j\}|k)$, $i=1,2; j=a,b,c,d$, can be calculated using only tables of the standard normal distribution function $\Phi(\cdot)$ and tables of $L(h,k,r)$ for the standardized bivariate normal distribution.

**proof:**

**CASE 1:** $\psi < 0$.

Then $\psi k < 0$, and

$$A(\psi) = \{(Y,Z): Y-Z \geq k\psi\},$$

{a} = \{(Y,Z): Y>k_1, Z>k_2, Y-Z \geq k\psi\},
{b} = \{(Y,Z): Y<k_1, Z<k_2, Y-Z \geq k\psi\},
{c} = \{(Y,Z): Y>k_1, Z<k_2\},
{d} = \{(Y,Z): Y<k_1, Z>k_2, Y-Z \geq k\psi\};

(see Figure I).

It is shown below that the three probabilities,

$$Pr_i((Y,Z) \in \{a\} + \{d\}|k),$$
$$Pr_i((Y,Z) \in \{b\} + \{d\}|k),$$
$$Pr_i((Y,Z) \in \{c\}|k),$$

can each be expressed in the form

$$\frac{L(h,k,r)}{\Phi(k)},$$

where $L(h,k,r) = Pr(U_1|h, U_2>k)$ when $(U_1, U_2)$ has a bivariate normal distribution with zero means, unit variances, and correlation coefficient $r$.

Since

$$\{a\}+\{b\}+\{c\}+\{d\} = A(\psi)$$

and $Pr_i((Y,Z) \in A(\psi)|k) = 1$, $i=1,2,$

the probabilities for regions {a},{b}, and {d} can be obtained from the following equations, where

$$Pr_i((j)|k) = Pr_i((Y,Z) \in \{j\}|k), j=a,b,c,d:$$

$$\begin{align*}
Pr_i(\{d\}|k) &= Pr_i(\{a\}+\{d\}|k) + Pr_i(\{b\}+\{d\}|k) + Pr_i(\{c\}|k) - 1 \\
Pr_i(\{a\}|k) &= Pr_i(\{a\}+\{d\}|k) - Pr_i(\{d\}|k) \\
Pr_i(\{b\}|k) &= Pr_i(\{b\}+\{d\}|k) - Pr_i(\{d\}|k)
\end{align*}$$

FIGURE II.1
(1) \[ \Pr(c|k) = \begin{cases} L \left\{ \frac{k_1 - \theta_{11}}{\sqrt{\theta_{11}}} , - \frac{k_2 - \theta_{12}}{\sqrt{\theta_{22}}} , - \frac{\theta_{12}}{\sqrt{\theta_{11}\theta_{22}}} \right\} \phi(k-\delta_1) \quad & \text{if } X \sim N(\delta, \Sigma); \\ \phi(k) \quad & \text{if } X \sim N(0, \Sigma). \end{cases} \]

**proof:** From the joint distribution of Y and Z, given in (33), we have:

\[ \Pr_1((c)|k) = \frac{\int_{k_1}^{\infty} \exp\left[-\frac{(w-v_1)'^T\Sigma^{-1}(w-v_1)}{2}\right] \, dw}{\phi(k_1) 2\pi \sqrt{\det \Sigma}}. \]

Define \( u = D(w-v_1) \) with \( D = \text{diag}(1/\sqrt{\theta_{11}}, -1/\sqrt{\theta_{22}}) \).

This transformation and the relationships, \( \nu_1' = (\theta_{11}, \theta_{12}), \nu_2 = 0, k_1 = k-\delta_1, k_2 = k, \)
given earlier in (32), (37), and the Basic Assumptions, give the result.

(2) \[ \Pr((a)+(a)|k) = \begin{cases} L \left\{ -(k-\delta_1), \frac{k_2 - \theta_{12}}{\sqrt{\theta_{22}}} , - \frac{\delta_{1}-\psi}{\sqrt{\theta_{22}}} \right\} \phi(k-\delta_1) \quad & \text{if } X \sim N(\delta, \Sigma); \\ \phi(k) \quad & \text{if } X \sim N(0, \Sigma). \end{cases} \]

**proof:** \( \Pr_1((a)+(a)|k) = \Pr_1(Z > k_2, Y > Z + k \psi|k) \).

The transformation \( u = DB(w-v_1) \) with

\[ B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma = \text{diag}(1/\sqrt{\theta_{11} + \theta_{22}}, 1/\sqrt{\theta_{22}}), \]

shows that
\[
\int_{k_2}^{\infty} \int_{k_2 \psi + Z}^{\infty} f_{\mathcal{W}}(w \mid k) \, dw = L(h, k, r) / \Phi(k_1),
\]

where the integrand is the p.d.f. of \( \mathcal{W}' = (Y, Z) \) given in (33) and
\[
\begin{align*}
h &= \frac{[k\psi - (\nu_{11} - \nu_{12})]}{\sqrt{\theta_{11} + \theta_{22} - 2\theta_{12}}}, \\
k &= \frac{(k_2 - \nu_{12})}{\sqrt{\theta_{22}}}, \\
r &= \frac{(\theta_{12} - \theta_{22})}{\sqrt{\theta_{22}(\theta_{11} + \theta_{22} - 2\theta_{12})}}.
\end{align*}
\]

These arguments can be simplified: From equations (35)-(37) we get
\[
\begin{align*}
\theta_{11} + \theta_{22} - 2\theta_{12} &= \psi^2, \\
\theta_{11} - \theta_{12} &= \psi \delta_1, \\
\theta_{12} - \theta_{22} &= \psi(\delta_1 - \psi),
\end{align*}
\]

so that \( h = (k - \delta_1) \text{ sgn} \psi = -(k - \delta_1) \) and \( r = [(\delta_1 - \psi) / \sqrt{\theta_{22}}] \text{ sgn} \psi = - (\delta_1 - \psi) / \sqrt{\theta_{22}} \).

(3)
\[
\Pr_i(b + d \mid k) = \begin{cases} 
\left[1 / \Phi(k - \delta_1) \right] L \left( - (k - \delta_1), - \frac{k_1 - \theta_{11}}{\sqrt{\theta_{11}}}, \frac{\delta_1}{\sqrt{\theta_{11}}} \right), & i = 1; \\
\left[1 / \Psi(k) \right] L \left( - k, - \frac{k_1}{\sqrt{\theta_{11}}}, \frac{\delta_1}{\sqrt{\theta_{11}}} \right), & i = 2.
\end{cases}
\]

\[
\Pr_i(b + d \mid k) = \Pr_i(Y < k_1, Z < Y - k \psi \mid k)
\]
\[
= \int_{-\infty}^{k_1} \int_{-\infty}^{Y-k\psi} f_{\mathcal{W}}(w \mid k) \, dw
\]
\[
= L(h, k, r) / \Phi(k_1) \text{ by the transformation } u = DB(w - \nu_1),
\]

with \( B = \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix} \) and
\[
D = \text{diag}(1 / \sqrt{\theta_{11} + \theta_{22} - 2\theta_{12}}, 1 / \sqrt{\theta_{11}}),
\]

and the arguments are the \( h \) given at the top of the page,
\[
k = -(k_1 - \nu_{11}) / \sqrt{\theta_{11}},
\]
\[
r = (\theta_{12} - \theta_{11}) / \sqrt{\theta_{11}(\theta_{11} + \theta_{22} - 2\theta_{12})}.
\]

Application of (53) simplifies these arguments, to give the desired result.
CASE 2: $\psi > 0$.

Now $k\psi > 0$ and

$A(w) = \{(Y,Z): Y-Z \leq k\psi\}$, and so

{a} = \{(Y,Z): Y>k_1, Z<k_2, Y-Z \leq k\psi\},
{b} = \{(Y,Z): Y<k_1, Z<k_2, Y-Z \leq k\psi\},
{c} = \{(Y,Z): Y>k_1, Z<k_2, Y-Z \leq k\psi\},
{d} = \{(Y,Z): Y<k_1, Z>k_2\};

(see Figure 2).

It can be shown, using proofs strictly analogous to those employed in CASE 1, that the probabilities of the four regions can be calculated as follows:

$$\begin{align*}
\Pr_i(\{c\}|k) &= \Pr_i(\{b\}+\{c\}|k) + \Pr_i(\{a\}+\{c\}|k) + \Pr_i(\{d\}|k) - 1; \\
\Pr_i(\{b\}|k) &= \Pr_i(\{b\}+\{c\}|k) - \Pr_i(\{c\}|k) \\
\Pr_i(\{a\}|k) &= \Pr_i(\{a\}+\{c\}|k) - \Pr_i(\{c\}|k)
\end{align*}$$

and

$$\begin{align*}
\Pr_i(\{d\}|k) &= \begin{cases}
L \left( \frac{k_1-\theta_{11}}{\sqrt{\theta_{11}}}, \frac{k_2-\theta_{12}}{\sqrt{\theta_{22}}}, \frac{-\theta_{12}}{\sqrt{\theta_{11}\theta_{22}}} \right) \cdot \left[ 1/\Phi(k-\delta_1) \right] & \text{if } i=1; \\
L \left( \frac{-k_1}{\sqrt{\theta_{11}}}, \frac{k_2}{\sqrt{\theta_{22}}}, \frac{-\theta_{12}}{\sqrt{\theta_{11}\theta_{22}}} \right) \cdot \left[ 1/\Phi(k) \right] & \text{if } i=2;
\end{cases}
\end{align*}$$

where

$$\begin{align*}
\Pr_i(\{a\}+\{c\}|k) &= \begin{cases}
L \left( -(k-\delta_1), \frac{k_1-\theta_{11}}{\sqrt{\theta_{11}}}, \frac{-\delta_1}{\sqrt{\theta_{11}}} \right) \cdot \left[ 1/\Phi(k-\delta_1) \right] & \text{if } i=1; \\
L \left( -k, \frac{k_1}{\sqrt{\theta_{11}}}, \frac{-\delta_1}{\sqrt{\theta_{11}}} \right) \cdot \left[ 1/\Phi(k) \right] & \text{if } i=2;
\end{cases}
\end{align*}$$

$$\begin{align*}
\Pr_i(\{b\}+\{c\}|k) &= \begin{cases}
L \left( -(k-\delta_1), -\frac{k_2-\theta_{12}}{\sqrt{\theta_{22}}}, \frac{\delta_1-\psi}{\sqrt{\theta_{22}}} \right) \cdot \left[ 1/\Phi(k-\delta_1) \right] & \text{if } i=1; \\
L \left( -k, -\frac{k_2}{\sqrt{\theta_{22}}}, \frac{\delta_1-\psi}{\sqrt{\theta_{22}}} \right) \cdot \left[ 1/\Phi(k) \right] & \text{if } i=2.
\end{cases}
\end{align*}$$
CHAPTER III
LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS
WITH UNEQUAL COVARIANCE MATRICES (NO TRUNCATION)

3.1 Basic Assumptions, Definitions, and General Procedure.

Given any mixture of two bivariate normal distributions with unequal
covariance matrices, let us designate the population comprising at least
half the mixture as the first population and then transform the variables
so that: (1) in Population 2 each variate has a standard normal distribu-
tion, and (2) in Population 1 each variate has a non-negative mean.
This type of transformation is chosen so that the results obtained in
this chapter will be comparable to those of the next chapter, where the
assumption of rectangular truncation prohibits all transformations ex-
cept translations and scale changes. [Note that, in the case of rec-
tangular truncation, the transformation which reduces the covariance ma-
trix of Population 2 to the identity matrix also "warps" the line \( x_1 = k \),
defining the truncation boundary, into something other than a line para-
allel to one of the coordinate axes—that is, rectangular truncation has
been destroyed.] Consequently, we assume:

**BASIC ASSUMPTIONS:**

1.) \( X \sim N(\delta, \Sigma) \) in Population 1, where
   a.) \( \delta_i \geq 0, i=1,2; \)
   b.) \( \Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 \\ \rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \) is positive definite.
2.) \( X \sim N(0, \Sigma) \) in Population 2,

\[
\text{where } \Sigma = \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix} \text{ is positive definite.}
\]

3.) \( \Sigma = \Sigma \).

4.) The proportion of Population 1 in the mixture is \( \alpha \geq 1/2 \).

5.) All parameters are known.

Under these assumptions, the distribution of a specified linear function \( \delta'X \) in each population is known. Therefore, we define:

**DEFINITION:** The optimal decision rule based on a specified linear function \( Y = \delta'X \) is to choose Population 1 iff

\[
af_1(y) > (1-\alpha)f_2(y),
\]

where \( f_i(y) \) denotes the p.d.f. of the random variable \( Y \) in the \( i \)-th population, \( i=1,2 \).

This decision rule is optimal in that it minimizes the total probability of misclassification among all decision rules based on \( Y \) (cf. Anderson [1], p. 131, Theorem 6.3.1).

**DEFINITION:** The optimal linear discriminant function (OLDF) is the linear function \( Y^* = \hat{\alpha}^*X \) which has the property that the optimal decision rule based on \( Y^* \) has the smallest total probability of misclassification in the class of decision rules based on linear functions of \( X \).

**PROCEDURE:** The procedure used to find \( Y^* \) is:

1.) To obtain the optimal decision rule based on an arbitrary linear function \( Y = \delta'X \), showing in the process that with no loss of generality we need consider only the linear functions \( Y = X_2 \) and \( Y = X_1 + dX_2 \) for all real \( d \).

2.) To obtain the total probability of misclassification corresponding to this optimal decision rule based on \( Y \).

3.) Considering this total probability of misclassification as a function of \( d \), to find \( \hat{\alpha}^* \) to minimize it. This has to be done numerically; i.e. the misclassification function must be evaluated numerically for specified values of \( d \) in order to find \( \hat{\alpha}^* \).
3.2 The Optimal Decision Rule Based on a Specified Linear Function.

Let \( \mathbf{d}' = (d_1, d_2) \) be any coefficient vector, and let \( Y = \mathbf{d}'X \). Then

\[
Y \sim \begin{cases} 
N(\mathbf{d}'\mathbf{\theta}, \mathbf{d}'\mathbf{\Sigma}_d) & \text{in Population 1;} \\
N(0, \mathbf{d}'\mathbf{Rd}) & \text{in Population 2;} 
\end{cases}
\]  

(1)

The optimal decision rule based on \( Y \) is to choose Population 1 iff

\[
\frac{\sqrt{\theta_2} \exp[-(y-\mathbf{\nu})^2/(2\theta_1)]}{\sqrt{\theta_1} \exp[-y^2/(2\theta_2)]} > \frac{1-\alpha}{\alpha}
\]

(2)

Let us regard the L.H.S. of (2) as a function of \( \mathbf{d} \) and denote it by \( L(\mathbf{d}'|\mathbf{\Sigma}, \mathbf{R}, \mathbf{\theta}) \). It follows directly from the definitions given in (1) that

\[
\left\{ \begin{array}{l}
L(0, d_2 | \mathbf{\Sigma}, \mathbf{R}, \mathbf{\theta}) = L(0, 1 | \mathbf{\Sigma}, \mathbf{R}, \mathbf{\theta}) \\
L(d_1, d_2 | \mathbf{\Sigma}, \mathbf{R}, \mathbf{\theta}) = L(1, d_2/d_1 | \mathbf{\Sigma}, \mathbf{R}, \mathbf{\theta}) \text{ when } d_1 \neq 0
\end{array} \right.
\]

(3)

Therefore, with no loss of generality we need consider only the linear functions \( Y = X_2 \) and \( Y = X_1 + dX_2 \) for all real \( \mathbf{d} \); that is, \( \mathbf{d}' = (0, 1) \) or \( (1, 1) \).

Now, inequality (2) can be written equivalently as

\[
U(Y) \equiv (\theta_2 - \theta_1)y^2 - 2\nu \theta_2 y + \theta_2 \left[ \nu^2 - \theta_1 \ln \left( \frac{\alpha}{1-\alpha} \right)^2 \frac{\theta_2}{\theta_1} \right] < 0.
\]

(4)

To simplify, let us define:

\[
D(\mathbf{d}) \equiv \theta_2(\mathbf{d}) - \theta_1(\mathbf{d});
\]

(5)

\[
C = \ln \left( \frac{\alpha}{1-\alpha} \right)^2 \geq 0 \text{ because } \alpha \geq 1/2.
\]

(6)

\[
G(\mathbf{d}) \equiv C + \ln \frac{\theta_2(\mathbf{d})}{\theta_1(\mathbf{d})} = \ln \left( \frac{\alpha}{1-\alpha} \right)^2 \frac{\theta_2(\mathbf{d})}{\theta_1(\mathbf{d})}.
\]

(7)

**CASE 1**: \( D(\mathbf{d}) > 0 \), i.e. \( \theta_1(\mathbf{d}) < \theta_2(\mathbf{d}) \).

\( U(y) \) is a U-shaped quadratic in \( y \) and so \( U(y) = 0 \) iff

\[
y = \frac{\nu \theta_2 \pm \sqrt{\theta_1 \theta_2} \nu \sqrt{\nu^2 + 4DEG}}{D} \equiv ((y'', y')), \quad \text{iff}
\]

(8)

where we define the notation \(((W, V)) = A \pm B\)

(9)
to mean that \( W=A+B \) and \( V=A-B \). Consequently, \( U(y)<0 \) iff \( y'<y<y'' \).

Now, \( y'' \) and \( y' \) are real and distinct because

1. \( \theta_1 < \theta_2 \) implies \( G = C + \ln(\theta_2/\theta_1) > C \geq 0 \).

2. Therefore, \( \nu^2 + DG > 0 \) for all \( \nu, \alpha \geq 1/2 \), and \( \theta_1 < \theta_2 \).

Thus the optimal decision rule based on \( Y \) is:

Choose Population 1 iff \( y' < y < y'' \).

\textit{CASE 2:} \( D(\bar{d}) < 0 \), i.e. \( \theta_1(\bar{d}) > \theta_2(\bar{d}) \)

In this case, \( U(y) \) is an arch-shaped quadratic in \( y \), and so \( U(y)<0 \) iff \( y \) is not in the closed interval \([y'',y']\), with \( y',y'' \) defined in (8). Unfortunately, \( y',y'' \) are not real for certain parameter combinations. Therefore, the optimal decision rule based on \( Y \) is:

Always choose Population 1 if \( \nu^2 < |D| \).

Choose Population 1 for all \( y \notin [y'',y'] \) otherwise.

\textit{CASE 3:} \( D(\bar{d}) = 0 \) and \( \nu(\bar{d}) \neq 0 \).

In this case \( \theta_1=\theta_2=\theta \) and so \( U(y) = -2\nu\theta y + \theta(\nu^2-\theta C) \), a linear function of \( y \). Thus the optimal decision rule based on \( Y \) is:

Choose Population 1 iff \( \begin{cases} y > y^0 \text{ when } \nu > 0; \\ y < y^0 \text{ when } \nu < 0; \end{cases} \)

where \( y^0 = \frac{\nu^2 - \theta C}{2\nu} \). \hspace{1cm} (10)

\textit{CASE 4:} \( D(\bar{d}) = 0 = \nu(\bar{d}) \).

\( Y \) has the same distribution in both populations, and so the optimal decision rule is to ignore \( y \) and simply choose Population 1. (If \( \alpha=1/2 \), the decision maker has the option of basing his choice on the flip of an unbiased coin.)

The decision rules just derived have interesting geometrical ramifications. For any choice of \( \Sigma \) and \( \mathcal{R} \) satisfying the Basic Assumptions
and for any linear function $X_1 + dX_2$, the function $D(d) = D(1,0)$ is at most a quadratic in $d$ and so has at most two zeros. Consequently, Cases 1 and 2 apply for all $d$-values except at most two. In geometric terms this means that the optimal partition of the $X$-plane based on a family of lines having almost any slope is a three-cell partition.

It may well be, however, that the probability that $X$ occurs in one of these three cells is negligible in both populations. If so, we shall call this a trivial three-cell partition.

### 3.3 The Corresponding Probability of Misclassification

Let $F(d)$ denote the probability of misclassification for the optimal decision rule based on the linear function $Y = d'X$, and let $Pr_j(E)$ denote the probability that event $E$ occurs with respect to the distribution in Population $j$, $j=1,2$. Then

\[ F(d) = \alpha Pr_1(U(y) \geq 0) + (1-\alpha)Pr_2(U(y) < 0) \]

where $U(y)$ is defined by (4).

**CASE 1:** $D(d) > 0$, i.e. $\theta_1 < \theta_2$

\[ F(d) = \alpha \left[ Pr_1(y \leq y') + Pr_1(y \geq y'') \right] + (1-\alpha)Pr_2(y' < y < y'') \]

\[ = \alpha \left[ \phi \left( \frac{y' - \nu}{\sqrt{\sigma_1}} \right) + \phi \left( -\frac{y'' - \nu}{\sqrt{\sigma_1}} \right) \right] + (1-\alpha) \left[ \phi \left( \frac{y''}{\sqrt{\sigma_2}} \right) - \phi \left( \frac{y'}{\sqrt{\sigma_2}} \right) \right] \]

with $y''$, $y'$ defined in (8).

**CASE 2:** $D(d) < 0$, i.e. $\theta_1 > \theta_2$

(a.) If $v^2 < G|D|$, then $F(d)$ is the probability of selecting an item from Population 2:

\[ F(d) = 1-\alpha \]

(b.) If $v^2 \geq G|D|$, then, with $y''$, $y'$ defined in (8),

\[ F(d) = \alpha Pr_1(y'' \leq y \leq y') + (1-\alpha)[Pr_2(y < y'') + Pr_2(y > y')] \]

\[ = \alpha \left[ \phi \left( \frac{y' - \nu}{\sqrt{\sigma_1}} \right) - \phi \left( \frac{y'' - \nu}{\sqrt{\sigma_1}} \right) \right] + (1-\alpha) \left[ \phi \left( \frac{y''}{\sqrt{\sigma_2}} \right) + \phi \left( -\frac{y'}{\sqrt{\sigma_2}} \right) \right] . \]
CASE 3: \( D(\bar{a}) = 0, \quad v(\bar{a}) = 0; \) i.e. \( \theta_1 = \theta_2 = \theta \).

a.) If \( v > 0 \), then

\[
F(\bar{a}) = \alpha \text{Pr}_1(y \leq y^0) + (1-\alpha) \text{Pr}_2(y > y^0)
\]

\[
= \alpha \Phi\left(\frac{y^0 - v}{\sqrt{\theta}}\right) + (1-\alpha) \Phi\left(-\frac{y^0}{\sqrt{\theta}}\right)
\]

\[
= \alpha \Phi\left(-\frac{v^2 + \theta c}{2\sqrt{\theta}}\right) + (1-\alpha) \Phi\left(-\frac{v^2 - \theta c}{2\sqrt{\theta}}\right)
\]

by definition of \( y^0 \) in (10).

b.) If \( v < 0 \), then

\[
F(\bar{a}) = \alpha \text{Pr}_1(y \geq y^0) + (1-\alpha) \text{Pr}_2(y < y^0)
\]

\[
= \alpha \Phi\left(\frac{y^2 + \theta c}{2\sqrt{\theta}}\right) + (1-\alpha) \Phi\left(\frac{y^2 - \theta c}{2\sqrt{\theta}}\right).
\]

c.) From (a) and (b), then, we have

\[
F(\bar{a}) = \alpha \Phi\left(-\frac{v^2 + \theta c}{2\sqrt{\theta}}\right) + (1-\alpha) \Phi\left(-\frac{v^2 - \theta c}{2\sqrt{\theta}}\right).
\]

CASE 4: \( D(\bar{a}) = 0 = v(\bar{a}) \).

\[
F(\bar{a}) = 1-\alpha.
\]

\[
\square
\]

To simplify calculations, define the vector of all the parameters used to specify the two distributions of \( X \):

\[
\pi' = (\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha), \quad (11)
\]

where \( \alpha \geq 1/2 \) is the proportion of elements from Population 1 in the mixture, \( \rho_i \) is the correlation between \( X_1 \) and \( X_2 \) in the \( i^{th} \) population, \( i=1, 2, \) and \( \delta_j \geq 0 \) and \( \sigma_j \) are the mean and standard deviation of \( X_j \) in Population 1, \( j=1, 2. \) (Recall that in Population 2 both variates have zero means and unit variances.) Let \( E_i(Y|\pi) \) and \( \text{Var}_i(Y|\pi) \) denote the mean and variance of the linear function \( Y = \bar{a}'X \) in the \( i^{th} \) population, \( i=1, 2, \) when the distributions of \( X \) are specified by the parameters of \( \pi \).

If \( \bar{a}' = (0, 1) \), i.e. if \( Y = X_2 \), then
\( v \equiv E_1(Y|\pi) = \delta_2 \),
\( \theta_1 \equiv \text{Var}_1(Y|\pi) = \sigma_2^2 \),
\( \theta_2 \equiv \text{Var}_2(Y|\pi) = 1 \),
\( D \equiv \theta_2 - \theta_1 = 1 - \sigma_2^2 \),

and
\( G \equiv C + \ln(\theta_2/\theta_1) = C - \ln \sigma_2^2 \),
so that
\[ v^2 + DG = \delta_2^2 + (1-\sigma_2^2)(C - \ln \sigma_2^2). \]

Because this function arises again, define
\[
H(\alpha, \delta, \sigma) \equiv \delta^2 + (1-\sigma^2)(C - \ln \sigma^2),
\]
where \( C \) is given by (6). Substitution of these values into (8) gives
\[
((y'', y')) = \frac{\delta_2 \pm \sigma_2 \sqrt{H(\alpha, \delta_2, \sigma_2)}}{1-\sigma_2^2}.
\]

Consequently, for the coefficient vector \( \mathbf{a}' = (0, 1) \),

\[
\begin{align*}
((\mathbf{y}' - \mathbf{y}')) &= \frac{\sigma_2 \delta_2 \pm \sqrt{H(\alpha, \delta_2, \sigma_2)}}{1-\sigma_2^2} \equiv ((A_1(\alpha, \delta_2, \sigma_2), A_2(\alpha, \delta_2, \sigma_2))) \\
((\mathbf{y}''', \mathbf{y}''')) &= \frac{\delta_2 \pm \sigma_2 \sqrt{H(\alpha, \delta_2, \sigma_2)}}{1-\sigma_2^2} \equiv ((A_3(\alpha, \delta_2, \sigma_2), A_4(\alpha, \delta_2, \sigma_2))) \\
\frac{v^2 + \theta C}{2 |v| \sqrt{\theta}} &= \frac{\delta_2^2 \pm C}{2 \delta_2} \equiv ((A_5(\alpha, \delta_2), A_6(\alpha, \delta_2)))
\end{align*}
\]

For the linear function \( Y = X_2 \), then, the above results can be summarized as follows:

**Theorem III.1**: Let \( P^0(\alpha, \delta_2, \sigma_2) \) denote the total probability of misclassification for the optimal decision rule based on \( Y = X_2 \). Then

\[
P^0(\alpha, \delta_2, \sigma_2) = \begin{cases} 
\alpha[\Phi(A_2) + \Phi(-A_1)] + (1-\alpha)[\Phi(A_3) - \Phi(A_4)] & \text{if } \sigma_2 < 1; \\
\alpha[\Phi(-A_5) + (1-\alpha)[\Phi(-A_6) & \text{if } \sigma_2 = 1, \delta_2 \neq 0; \\
1-\alpha & \text{if } \sigma_2 = 1, \delta_2 = 0; \\
\alpha[\Phi(A_2) - \Phi(A_1)] + (1-\alpha)[\Phi(A_3) + \Phi(-A_4)] & \text{if } \sigma_2 > 1, H > 0;
\end{cases}
\]

where \( H \) denotes the function \( H(\alpha, \delta_2, \sigma_2) \) defined in (12).

\[ \square \]

If \( \mathbf{a}' = (1, d) \), where \( d \) is any real number, i.e. \( Y = X_1 + dX_2 \), then
\[
\begin{align*}
\begin{cases}
\nu(d|\pi) &\equiv E_1(Y|\pi) = \delta_1 + d\delta_2, \\
\theta_1(d|\pi) &\equiv \text{Var}_1(Y|\pi) = \sigma_1^2 + 2\rho_1\sigma_1\sigma_2d + \sigma_2^2d^2, \\
\theta_2(d|\pi) &\equiv \text{Var}_2(Y|\pi) = 1 + 2\rho_2d + d^2,
\end{cases}
\end{align*}
\]  
(14)

and the functions defined in (5) and (7) can be written:
\[
\begin{align*}
\begin{cases}
D(d|\pi) &\equiv \theta_2(d|\pi) - \theta_1(d|\pi) = (1-\sigma_1^2) + 2(\rho_2-\rho_1\sigma_1\sigma_2)d + (1-\sigma_2^2)d^2, \\
G(d|\pi) &\equiv C + \ln \theta_2(d|\pi) - \ln \theta_1(d|\pi)
\end{cases}
\end{align*}
\]  
(15)

where \( C = 2\ln[a/(1-a)] \). Moreover, if we define
\[
h(d|\pi) \equiv \nu^2(d|\pi) + D(d|\pi) \cdot G(d|\pi)
\]  
(16)

and temporarily drop the argument \((d|\pi)\) from all the above functions, we can write \( y' \) and \( y'' \) more simply as
\[
((y'', y')) = \frac{\nu\theta_2 \pm \sqrt{\nu\theta_2 \cdot \sqrt{h}}}{D}.
\]

Then
\[
\begin{align*}
\begin{cases}
((\sqrt{\theta_1}, \sqrt{\theta_1}^2 - Y)) &= \frac{\nu\sqrt{\theta_1} \pm \sqrt{\theta_2} \cdot \sqrt{h}}{D} \equiv ((B_1, B_2)) \\
((\sqrt{\theta_2}, \sqrt{\theta_2}^2 - Y')) &= \frac{\nu\sqrt{\theta_2} \pm \sqrt{\theta_1} \cdot \sqrt{h}}{D} \equiv ((B_3, B_4)) \\
\text{When } \theta_1=\theta_2, \\
&\frac{\nu^2 \pm \theta_2 C}{2|\nu|\sqrt{\theta_2}} \equiv ((B_5, B_6))
\end{cases}
\end{align*}
\]  
(17)

where each of the functions \( B_i \), \( i=1-6 \), implicitly has the argument \((d|\pi)\). For the linear function \( Y=X_1+dX_2 \), then, the above results can be summarized as follows:

**THEOREM III.2:** Given a vector of parameters \( \pi \), let \( F(d|\pi) \) denote the total probability of misclassification for the optimal decision rule based on the linear function \( Y=X_1+dX_2 \), where \( d \) is any real number.

\[
F(d|\pi) = \begin{cases}
\alpha[\phi(B_2)+\phi(-B_1)] + (1-\alpha)[\phi(B_3)-\phi(B_4)] & \text{if } D>0; \\
\alpha\phi(-B_5) + (1-\alpha)\phi(-B_6) & \text{if } D=0, \nu=0; \\
1-\alpha & \text{if } D=0, h>0; \\
\alpha[\phi(B_2)-\phi(B_1)] + (1-\alpha)[\phi(B_3)+\phi(-B_4)] & \text{if } D<0, h>0;
\end{cases}
\]

where all quantities involved in the formulation are defined in (14)-(17).

\[\Box\]

The following lemma is a direct consequence of statement (3) in
Section 3.2:

**Lemma III.1:** Given a vector of parameters \( \pi \), let \( \bar{d}' = (d_1, d_2) \) be any coefficient vector, and let \( F(d_1, d_2 | \pi) \) denote the probability of misclassification for the optimal decision rule based on \( \bar{d}'X \). Then

\[
F(d_1, d_2 | \pi) = \begin{cases} 
F^0(\alpha, \delta_2, \sigma_2) & \text{if } d_1 = 0 \\
F(\bar{d} | \pi) \text{ with } \bar{d} = d_2d_1^{-1} & \text{if } d_1 \neq 0 
\end{cases}
\]

Both of the functions, \( F \) and \( F^0 \), can be easily evaluated with the aid of a computer.

### 3.4 Some Properties of the Functions \( F \) and \( F^0 \)

Since the mathematical form of \( F(d | \pi) \) depends upon the functions \( D(d | \pi) \) and \( h(\bar{d} | \pi) \), we must know something about their behavior, especially when considered to be functions of the parameters of \( \pi \). To simplify notation in this section, we will use a prime and a double-prime on a function to denote the first- and second-partial derivatives, respectively, with respect to \( \bar{d} \). There is no ambiguity in this notation, as the functions \( y' \) and \( y'' \) defined in (8) are not considered in this section. The argument \( \bar{d} | \pi \) is often dropped for simplicity.

#### 3.4.1 Some Properties of \( D \) and \( h \)

**Lemma III.2:** The condition \( D(d | \pi) = D'(d | \pi) = D''(d | \pi) = 0 \) violates Assumption 3. If \( D(d_0 | \pi) = D'(d_0 | \pi) = 0 \), then \( D(d | \pi) \) attains a maximum or a minimum at \( d_0 \) according as \( \sigma_2 > 1 \) or \( \sigma_2 < 1 \).

**Proof:** \( D'(d | \pi) = 2[(1-\sigma_2^2)d + (\rho_2-\rho_1\sigma_1\sigma_2)] \) and \( D''(d | \pi) = 2(1-\sigma_2^2) \).

If \( D = D' = D'' = 0 \), then \( D'' = 0 \) implies \( \sigma_2 = 1 \), which together with \( D' = 0 \) implies \( \rho_2 = \rho_1\sigma_1 \), which together with \( D = 0 \) implies \( \sigma_1 = 1 \) and so \( \rho_1 = \rho_2 \). Thus \( \Sigma = \mathbb{R} \), which violates Assumption 3.

Therefore, if \( D(d_0 | \pi) = D'(d_0 | \pi) = 0 \), then \( D''(d_0 | \pi) \neq 0 \), that is, \( \sigma_2 \neq 1 \); and the result follows from elementary calculus.

\( \square \)
**Lemma III.3:** (1) If $D(d|\pi) > 0$, then $h(d|\pi) > 0$. (2) If $\theta_1 > \left(\frac{a}{1-a}\right)^2 \theta_2$, then $h(d|\pi) > 0$. (3) If $\theta_2 < \theta_1 \leq \left(\frac{a}{1-a}\right)^2 \theta_2$ and $v=0$, then $h(d|\pi) \leq 0$. (4) If $D<0$ and $v \neq 0$, then $h(d|\pi)$ will be negative, if at all, only if $\theta_1 = \theta_2 + \frac{v^2}{2}(1+\sqrt{1+4\theta_2/v^2})/2$. (5) If $D>0$ and $v^2 > \theta_2 \left(\frac{2a-1}{a(1-a)}\right)^2$, then $h(d|\pi) > 0$.

**Proof:** The proof is given in six parts, which do not correspond to the five statements comprising the Lemma.

A.) $D>0$ iff $\theta_1 < \theta_2$, which implies $\ln(\theta_2/\theta_1) > 0$, so that $G = C + \ln \theta_2 > C$. Therefore $h = v^2 + DG > v^2 \geq 0$.

This establishes Statement (1) of the Lemma.

B.) $\theta_1 > \left(\frac{a}{1-a}\right)^2 \theta_2$ implies $D<0$ and $G = \ln \left(\left(\frac{a}{1-a}\right)^2 \frac{\theta_2}{\theta_1}\right) < 0$.

Therefore, $h > v^2 \geq 0$.

This establishes Statement (2).

C.) $\theta_1 = \theta_2$ iff $D=0$, which implies $h = v^2 \geq 0$, with equality iff $v=0$.

D.) $D<0$ and $v=0$ imply $h = DG \leq 0$ iff $\theta_1 \leq \left(\frac{a}{1-a}\right)^2 \theta_2$.

Parts (C) and (D) establish Statement (3).

E.) $D<0$ and $v \neq 0$ imply $h = v^2 - D[\ln(1-D/\theta_2)-C]$ by (5) and (7).

Therefore, $h>0$ iff $(1-D/\theta_2)\exp(-v^2/D) \leq a^2/(1-a)^2$. (i)

Fix $\theta_2$ (i.e. fix $\rho_2$), and consider the function

$$g(y) = (1-ay) \exp(-b/y)$$

with $y=D<0$, $a=(1/\theta_2)>0$, and $b=v^2>0$. Then

$$g'(y) = y^{-2}(b-aby-ay^2) \exp(-b/y) = 0 \iff ay^2 + aby - b = 0,$$

i.e. iff $y = -b(1+\sqrt{1+4/(ab)})/2$.

Since $1+\sqrt{1+4/(ab)} < 0$, the unique critical point in the domain of negative numbers is $y_0 = -b(1+\sqrt{1+4/(ab)})/2$, i.e.

$$D_0 = -v^2[1 + \sqrt{1+4\theta_2/v^2}]/2.$$ (ii)

Since $a>0$ and $b>0$, $g(y) \to +\infty$ as $y \to -\infty$ and as $y \to 0$.

Therefore, $y_0$ is a minimum.
It follows from (i), (ii), and (16) that \( h \leq 0 \) iff \( g(D) \leq (\frac{\alpha}{1-\alpha})^2 \) and that \( g(D) \) is minimized when \( D = D_0 \), i.e., when \( \theta_1 = \theta_2 = D_0 \).

This establishes Statement (4).

F.) Let \( \theta_2^0 = \theta_2 - D_0 \), where \( D_0 \) is defined in (ii).

Statements (2) and (4) imply that if \( v \neq 0 \) and \( \theta_2^0 > \left( \frac{a}{1-a} \right)^2 \), then \( h > 0 \) for all \( \theta_1 > \theta_2 \).

That is, if \( \alpha, \delta_1, \delta_2, \) and \( \rho_2 \) are fixed, and \( \Pi \) is the set of all parameter vectors \( \pi \) having these fixed values, and if the two conditions

\[
v \neq 0 \quad \text{and} \quad \theta_2^0 > \left( \frac{a}{1-a} \right)^2 \theta_2
\]

are satisfied, then \( h > 0 \) for every set \( \pi \in \Pi \) having the property that \( \theta_1 > \theta_2 \).

Now, letting \( B = \left( \frac{a}{1-a} \right)^2 \geq 1 \), we see that \( \theta_2^0 > B \theta_2 \) iff \( \sqrt{1 + 4 \theta_2/v^2} > 2(B-1)\theta_2/v^2 - 1 \),

iff \( v^2 > \theta_2 \left( \frac{2a-1}{a(1-a)} \right)^2 \).

By this and Part (C), Statement (5) is established.

\[ \square \]

COROLLARY III.1: Given \( \alpha \geq 1/2, \delta \geq 0, \sigma > 0 \).

1. \( H(\alpha, \delta, \sigma) = h(0|\pi) \) for any parameter vector \( \pi \) having \( \delta_1 = \delta, \psi_1 = \sigma \), and the given \( \alpha \).

2. If \( \sigma < 1 \), then \( H(\alpha, \delta, \sigma) > 0 \).

3. If \( \sigma > \frac{a}{1-a} \geq 1 \), then \( H(\alpha, \delta, \sigma) > 0 \).

4. If \( \delta > 0 \), then \( H(\alpha, \delta, \sigma) \leq 0 \) for all \( \sigma \) such that \( 1 \leq \sigma \leq \frac{a}{1-a} \).

5. If \( \sigma > 0 \) and \( \sigma > 1 \), then \( H(\alpha, \delta, \sigma) \) will be negative, if at all, only if

\[
\sigma \leq \sqrt{1 + \delta^2(1 + \sqrt{1 + 4\delta^2})} / 2 \] \(1/2 \).

6. If \( \sigma > \frac{2a-1}{a(1-a)} \), then \( H(\alpha, \delta, \sigma) > 0 \) for all \( \sigma \geq 1 \).

proof: Statement (1) follows directly from definitions (12)-(16).

The remaining statements follow from (1) and Lemma 3.

\[ \square \]

COROLLARY III.2: If \( \alpha = 1/2 \), then

1. \( h(d|\pi) > 0 \) for all \( d \) such that \( D(d|\pi) \neq 0 \); and

2. \( h(d|\pi) = 0 \) iff \( D(d|\pi) = v(d|\pi) = 0 \).

proof: Statement (1) follows from Statements (1) and (2) of Lemma 3. Statement (2) follows from Lemma 3-(3) and the definition of \( h \).
**Lemma III.4:** (1) \( h(d|\pi) \) is a continuous and differentiable function of \( d \) for all real \( d \). (2) If \( D(d_0|\pi) = 0 \), then as \( d \to d_0 \),

\[
\begin{align*}
\theta_i &+ \theta, \ i=1,2; \\
G &\to C; \\
G' &\to D'(d_0)/\theta; \\
h &\to v_0^2; \\
h' &\to 2\delta_2 v_0 + C D'(d_0); \\
h'' &\to 2[\delta_2^2 + (D'(d_0))^2/\theta + C(1-\sigma_2^2)];
\end{align*}
\]

where \( v_0 \equiv v(d_0|\pi) \), \( \theta \equiv \theta_2(d_0|\pi) \), and \( D'(d_0) \) denotes the first partial derivative of \( D \) with respect to \( d \), evaluated at \( d=d_0 \).

**Proof:** \( h \) is a continuous and differentiable function of \( d \) because it is the sum of a quadratic, \( v^2 \), and a function which is the product of a quadratic, \( D \), and the natural logarithm function having an argument which is finite and well-defined because \( \theta_i > 0, i=1,2 \), for all \( d \).

If \( D(d_0) = 0 \), then as \( d \to d_0 \)

\[\begin{align*}
\text{a.} & \quad D(d) &\equiv \theta_2(d) - \theta_2(d_0) &\to 0 \text{ implies } \theta_i \to \theta_2(d_0), \ i=1,2. \\
\text{b.} & \quad G(d) &\equiv C + D \ln(\theta_2/\theta_1) &\to C > 0. \\
\text{c.} & \quad G'(d) &\equiv \theta_2'/\theta_2 - \theta_1'/\theta_1 \to [\theta_2'(d_0) - \theta_1'(d_0)]/\theta = [D'(d_0)]/\theta . \\
\text{d.} & \quad h(d) &\equiv v^2 + DG + v_0^2. \\
\text{e.} & \quad h'(d) &\equiv 2\delta_2 v + DG' + D'G + 2\delta_2 v_0 + C[D'(d_0)]. \\
\text{f.} & \quad h''(d) &\equiv 2\delta_2^2 + DG'' + 2D'G' + D''G. 
\end{align*}\]

\[\Box\]

**3.4.2 Properties of \( F \) and \( F^0 \).**

Theorem 3 tells us that despite the fact that \( F(d|\pi) \) is defined by four different functions over the range of \( d \), \( F \) is continuous.

**Theorem III.3:** For a given vector of parameters \( \pi \), \( F(d|\pi) \) is a continuous function of \( d \).

**Proof:** Let \( d_0 \) be any real number.

To simplify and clarify, \( \pi \) is omitted at some points in the proof.
Case 1: \( D(d_0 | \pi) \neq 0 \) and \( h(d_0 | \pi) \neq 0 \).

There exists a neighborhood of \( d_0 \), say \( N(\epsilon) \), such that for all \( d \in N(\epsilon) \), \( D(d) \cdot D(d_0) > 0 \) and \( h(d) \cdot h(d_0) > 0 \).

Therefore, in \( N(\epsilon) \) \( F(d) \) is evaluated by exactly one of the four functions specified in Theorem 2.

If \( F(d) \equiv 1 - \alpha \), it is clearly continuous!

\( F(d) \) is not a function of \( B_5 \) and \( B_6 \) in \( N(\epsilon) \) because \( D(d) \neq 0 \) for all \( d \in N(\epsilon) \).

Therefore, if \( F(d) \neq 1 - \alpha \), it is a linear function of \( \phi \)-functions, which are continuous functions of \( d \) if the arguments are continuous functions of \( d \).

Now, \( B_1 - B_4 \) are functions of the form, \((\sqrt{\theta_i} \pm \sqrt{\theta_j} \phi_i)/D, i \neq j = 1, 2\), which are real and finite because \( D \neq 0 \) and \( h > 0 \) for all \( d \in N(\epsilon) \) [for \( h \leq 0 \) implies \( F = 1 - \alpha \)].

Thus the arguments \( B_1 - B_4 \) are continuous functions of \( d \).

Case 2: \( D(d_0 | \pi) = 0 \) but \( h(d_0 | \pi) = 0 \).

Lemma 3-(1) says that if \( D > 0 \), then \( h > 0 \). Since \( h(d_0) = 0 \) and \( D(d_0) \neq 0 \), this means that \( D(d_0) < 0 \).

There exists, then, a neighborhood \( N(\epsilon) \) of \( d_0 \) such that \( D(d) < 0 \) for all \( d \in N(\epsilon) \).

Now, \( D(d_0) < 0 \) and \( h(d_0) = 0 \) imply that \( F(d_0) = 1 - \alpha \). (Theorem 2)

Let \( N_1 \) denote a one-sided neighborhood (possibly empty) of \( d_0 \) such that \( h(d) > 0 \) for all \( d \) in \( N_1 \).

Then in \( N_1 \), \( h(d) + h(d_0) = 0 \) as \( d \rightarrow d_0 \), and so as \( d \rightarrow d_0 \)

\[ \lim B_1 = \lim B_2 = \frac{\sqrt{\theta_1 \sigma}}{D_0} \quad \text{and} \quad \lim B_3 = \lim B_4 = \frac{\sqrt{\theta_2 \sigma}}{D_0}, \]

where \( \theta_i = \theta_i (d_0), i = 1, 2 \), and \( D_0 \equiv D(d_0) \).

It follows from Lemma A (below) that \( F(d) \rightarrow 1 - \alpha \) as \( d \rightarrow d_0 \) in \( N_1 \).

**Lemma A:** Let \( I \) be an open interval having \( d_0 \) for one endpoint.

If for every \( d \in I \), \( (1) \ D(d) < 0 \), \( (2) \ h(d) > 0 \), \( (3) \ \lim B_1 = \lim B_2 \), \( d \rightarrow d_0 \)

and \( (4) \ \lim B_3 = \lim B_4 \), \( d \rightarrow d_0 \)

then for \( d \in I \), \( \lim F(d) = 1 - \alpha \), \( d \rightarrow d_0 \)

**Proof:** By Theorem 2, \( F(d) = a[\phi(B_2) - \phi(B_1)] + (1 - \alpha)[\phi(B_3) + \phi(-B_4)] \)

for all \( d \in I \). Therefore, for all \( d \in I \),

\[ F(d) \rightarrow (1 - \alpha)[\phi(A) + \phi(-A)] = 1 - \alpha \quad \text{as} \quad d \rightarrow d_0, \]
where \( A = \lim_{d \to d_0} B_3 \).

Let \( N_2 \) denote a one-sided neighborhood (possibly empty) of \( d_0 \) such that \( h(d) \leq 0 \) for all \( d \in N_2 \).

By Theorem 2, \( F(d) = l-a \) for all \( d \in N_2 \).

Since \( N(\varepsilon) \) is the union of two intervals like \( N_1 \) and/or \( N_2 \),
\[ F(d) \to l-a = F(d_0) \text{ as } d \to d_0. \]

**Case 3:** \( D(d_0|\pi) = 0 \) but \( h(d_0|\pi) = \nu_0^2 > 0 \).

By Theorem 2, \( F(d_0) = a \phi(-B_5^0) + (1-a) \phi(-B_6^0) \) where \( B_j^0 = B_j(d_0), \ j=5,6. \)

\( h(d_0) > 0 \) implies there exists a neighborhood \( N(\varepsilon) \) of \( d_0 \) such that
for every \( d \in N(\varepsilon) \), \( D(d) \neq 0 \), \( h(d) > 0 \), and \( \operatorname{sgn} \nu(d) = \operatorname{sgn} \nu_0 \),
where \( \nu_0 \equiv \nu(d_0) = 0 \).

When \( \nu(d) \neq 0 \), the functions \( B_1-B_4 \) can be written in the form:
\[ (v/D)[\sqrt{\theta_i} \pm (\operatorname{sgn} \nu)\sqrt{\theta_j}/\sqrt{1+Q}] \quad (20) \]

where \( Q \equiv DG/\nu^2 \quad (21) \)
and \( i\neq j=1,2. \)

When there is a negative sign between the two terms of (20), the Taylor expansion of \( \nu/1+Q \) can be substituted into (20) to give
\[ (v/D)[\sqrt{\theta_i} - \sqrt{\theta_j}/\sqrt{1+Q}] = \frac{(i-j)\nu}{\theta_i + \sqrt{\theta_2}} - \frac{\sqrt{\theta_1} C}{2\nu} \left( 1 - \frac{Q}{4} (1+Q^*)^{-3/2} \right), \quad (22) \]

where \( Q^* \) is between \( Q \) and 0.

Since \( D_0 = 0 \) and \( \nu_0 \neq 0 \), we have: as \( d \to d_0 \),
\[ Q \to 0, \quad (G/\nu) + (C/\nu_0), \quad \theta_i \to \theta_i, \ i=1,2. \]

Then
\[ \frac{\nu}{D} \left( \sqrt{\theta_i} - \sqrt{\theta_j}/\sqrt{1+Q} \right) \to \frac{(i-j)\nu_0}{2\nu_0} - \frac{\sqrt{\theta C}}{2\nu_0} \text{ as } d \to d_0, \ i\neq j=1,2. \quad (23) \]

When there is a plus sign between the two terms of (20), then, by Lemma 2, \( D(d) \) must be either a linear or a quadratic function of \( d \).

Consequently, \( N(\varepsilon) \) is the union of two open intervals, each of which
is one of the following types:
\( I_1 \) has the property that \( D(d) \) is positive for all \( d \in I_1 \).
\( I_2 \) has the property that \( D(d) \) is negative for all \( d \in I_2 \).

Thus,
\[ \lim_{d \to d_0} \frac{\nu}{D} \left( \sqrt{\theta_i} + \sqrt{\theta_j}/\sqrt{1+Q} \right) = \begin{cases} (+\infty) (\text{sgn } \nu_0) \text{ in } I_1 \\ (-\infty) (\text{sgn } \nu_0) \text{ in } I_2 \end{cases}. \quad (24) \]

If \( \nu_0 < 0 \), then
\[ ((B_2,B_1)) = (v/D)[\sqrt{\theta_1} \pm \sqrt{\theta_2}/\sqrt{1+Q}] \text{ and} \]
\[ ((B_4, B_3)) = (\nu/D)[\sqrt{\theta_2} \pm \sqrt{\theta_1}/(\nu+Q)]. \]

By (24), \[ \lim_{d \to d_0} B_2 = \lim_{d \to d_0} B_4 = \begin{cases} -\infty & \text{in } I_1, \\ +\infty & \text{in } I_2. \end{cases} \]

By (23), \[ B_1 \to -[\nu_0/(2\sqrt{\theta}) + C\sqrt{\theta}/(2\nu_0)] = B_5^0 \text{ as } d \to d_0; \]
and \[ B_3 \to \nu_0/(2\sqrt{\theta}) - C\sqrt{\theta}/(2\nu_0) = -B_6^0 \text{ as } d \to d_0. \]

Consequently, \[ F(d) = \begin{cases} \alpha[\phi(-\infty) + \phi(-B_5^0)] + (1-\alpha)[\phi(-B_6^0) + \phi(\infty)] & \text{in } I_1, \\ \alpha[\phi(\infty) - \phi(B_5^0)] + (1-\alpha)[\phi(-B_6^0) + \phi(-\infty)] & \text{in } I_2. \end{cases} = F(d_0). \]

If \( \nu_0 > 0 \), then \( F(d) \to F(d_0) \) by similar reasoning.

**Case 4:** \( D(d_0) = h(d_0) = 0. \)

By Theorem 2, \( F(d_0) = 1-\alpha \).

Also, \( h(d_0) = \nu_0^2 + D(d_0) \cdot G(d_0) = 0 \) and \( D(d_0) = 0 \) implies \( \nu_0 = 0. \) \( \text{(iii)} \)

It follows from (19) that \( h'(d_0) = C[D'(d_0)]. \) \( \text{(iv)} \)

**Case 4-A:** \( D'(d_0) = D_0 \neq 0. \)

Since \( D(d) \) is strictly increasing or strictly decreasing in a neighborhood of \( d_0 \), there exist two disjoint open intervals having common endpoint \( d_0 \) and such that

- \( D(d) \) is positive for all \( d \) in \( I_1 \).
- \( D(d) \) is negative for all \( d \) in \( I_2 \).

If \( h(d) \) is positive, then functions \( B_1-B_4 \) can be written as follows:

\[ \sqrt{\theta_1} (\nu/D) \pm (\text{sgn } D) \sqrt{\nu/D^2} \quad \text{(v)} \]

and \[ (\nu/D) \to (\delta_2/D_0) = L, \quad \text{(vi)} \]

which is finite.

If \( \alpha = 1/2, \) then \( C=0 \) and so, by (iv), \( h'(d_0)=0. \)

However, \( h''(d_0) = 2[\delta_2^2 + D_0^2/\theta] > 0 \) implies \( h(d)>0 \) for all \( d \) in \( I_1 \cup I_2 \).

So \( (h/D^2) \to [h''(d_0)]/(2D_0^2) = L^2 + (1/\theta) \) as \( d \to d_0. \)

This fact, together with (v) and (vi), implies

\[ \lim_{d \to d_0} B_1 = \lim_{d \to d_0} B_3 = \begin{cases} A & \text{in } I_1, \\ B & \text{in } I_2 \end{cases} \]
and

\[ \lim_{d \to d_0} B_2 = \lim_{d \to d_0} B_4 = \begin{cases} B & \text{in } I_1, \\ A & \text{in } I_2 \end{cases} \]

where \((A,B)) = L\sqrt{\theta} \pm \sqrt{1+\theta L^2} \).
It follows from Lemma B (below) that \( F(d) + 1/2 = 1 - \alpha \) as \( d \to d_0 \).

**Lemma B:** Let \( I \) be an open interval having \( d_0 \) for one endpoint. If

1. \( \alpha = 1/2 \),
2. \( h(d) \) is positive for all \( d \) in \( I \),
3. \( D(d) \) is nonzero for all \( d \) in \( I \),
4. \( \lim_{d \to d_0} B_1 = \lim_{d \to d_0} B_3 \), and
5. \( \lim_{d \to d_0} B_2 = \lim_{d \to d_0} B_4 \),

then \( F(d) + 1/2 = 1 - \alpha \) as \( d \to d_0 \).

**Proof:** Let \( A \) and \( B \) denote the limits of \( B_1 \) and \( B_2 \), respectively, as \( d \to d_0 \).

By (3), either \( D(d) \) is positive for all \( d \) in \( I \) or \( D(d) \) is negative for all \( d \) in \( I \).

\[
F(d) = \begin{cases} 
(1/2)[\phi(B) + (-A) + (A) - \psi(B)] & \text{if } D > 0, \\
(1/2)[\phi(B) - \phi(A) + (A) + \psi(-B)] & \text{if } D < 0, 
\end{cases} = 1/2, \\
as d \to d_0. 
\]

If \( \alpha > 1/2 \), then \( C \) is positive, and \( h'(d_0) = CD_0 \) implies there exists a neighborhood \( N(\epsilon) \) of \( d_0 \) such that

\[
h(d) \cdot D(d) > 0
\]

for all \( d \to d_0 \) in \( N(\epsilon) \).

Thus, \( D < 0 \) implies \( h < 0 \), and so \( F(d) = 1 - \alpha \). Consequently,

\[
F(d) \to 1 - \alpha \quad \text{as } d \to d_0 \quad \text{in } I_2.
\]

In \( I_1 \),

\[
(h/D^2) \to CD_0/(2\lim DD') = C/0 = +\infty \quad \text{as } d \to d_0.
\]

This, together with (v) and (vi), implies that as \( d \to d_0 \) in \( I_1 \)

\[
((\lim B_1, \lim B_2)) = +\infty = ((\lim B_3, \lim B_4)).
\]

It follows from Lemma C (below) that \( F(d) \to 1 - \alpha \) as \( d \to d_0 \).

**Lemma C:** Let \( I \) be an open interval having \( d_0 \) for one endpoint. If

1. \( D(d) > 0 \) for all \( d \in I \),
2. \( \lim_{d \to d_0} B_1 = \lim_{d \to d_0} B_3 = +\infty \) as \( d \to d_0 \) in \( I \), and
3. \( \lim_{d \to d_0} B_2 = \lim_{d \to d_0} B_4 = -\infty \) as \( d \to d_0 \) in \( I \),

then \( F(d) \to 1 - \alpha \) as \( d \to d_0 \) in \( I \).

**Proof:** As \( d \to d_0 \) in \( I \),

\[
F(d) \to a[\phi(-\infty) + \psi(-\infty)] + (1 - \alpha)[\phi(+\infty) - \psi(-\infty)] = 1 - \alpha. \]

**Case d-B:** \( D'(d_0) = 0 \) and \( \delta_2 > 0 \).

By (iii), \( v_0 = 0 \), and since \( \delta_2 > 0 \), it follows that
\( v(d) \neq 0 \) for all \( d \neq d_0 \) and \( \text{sgn} \ v(d) = \text{sgn} (d - d_0) \).

As in Case 3, we can write \( B_1 - B_4 \) in the form

\[
(v/D)[\sqrt{\theta_1} \pm (\text{sgn} \ v(\sqrt{\theta_1})/1+Q} \text{ with } Q = DG/v^2.
\]

Furthermore, it follows from Lemma 2 that, since \( D(d_0) = D'(d_0) = 0 \), \( D(d) \) is positive for all \( d \neq d_0 \) or negative for all \( d = d_0 \) according as \( \sigma_2 < 1 \) or \( \sigma_2 > 1 \).

Thus \( (v/D) + (\pm)[\text{sgn}(d - d_0)][\text{sgn}(1 - \sigma_2)] \) as \( d \rightarrow d_0 \). \( \text{(vii)} \)

If \( \alpha = 1/2 \), it follows from (19) that

\[
C = G(d_0) = G'(d_0) = h'(d_0) = 0 \text{ and } h''(d_0) = 2\delta_2^2 > 0.
\]

Thus \( h(d) \) is positive for all \( d \neq d_0 \) in some neighborhood \( N(\epsilon) \) of \( d_0 \). Then \( (v/D)[\sqrt{\theta_1} - \sqrt{\theta_j^j(1 + Q)}] \) can be written as (22) with

\[
\lim Q = [\lim(DG^\prime + 2DG^\prime + D^\prime G)]/(2\delta_2^2) = 0, \quad \text{(viii)}
\]

\[
\lim(G/v) = (\lim G^\prime)/\delta_2 = 0 \quad \text{(ix)}
\]

as \( d \rightarrow d_0 \), so that

\[
(v/D)[\sqrt{\theta_1} - \sqrt{\theta_j^j(1 + Q)} + 0 \text{ as } d \rightarrow d_0. \quad \text{(ix)}
\]

Moreover, from (vii) and the fact that

\[
(\sqrt{\theta_1} + \sqrt{\theta_j^j(1 + Q)} + 2\delta > 0 \text{ as } d \rightarrow d_0
\]

it follows that

\[
(v/D)[\sqrt{\theta_1} + \sqrt{\theta_j^j(1 + Q)} + (\pm)[\text{sgn}(d - d_0)][\text{sgn}(1 - \sigma_2)] \text{ as } d \rightarrow d_0 \).
\]

Therefore

\[
\lim B_1 = \lim B_3 = \lim B_2 = \lim B_4 = 0, \text{ and } d \rightarrow d_0 \quad d \rightarrow d_0 \quad d \rightarrow d_0 \quad d \rightarrow d_0
\]

\[
\lim B_1 = \lim B_3 = (\pm)[\text{sgn}(1 - \sigma_2)], \text{ and } \quad d \rightarrow d_0 \quad d \rightarrow d_0
\]

\[
\lim B_2 = \lim B_4 = (\mp)[\text{sgn}(1 - \sigma_2)], \quad d \rightarrow d_0 \quad d \rightarrow d_0
\]

By Lemma B (above), \( F(d) + 1/2 = l - \alpha \) as \( d \rightarrow d_0 \).

If \( \alpha > 1/2 \), then \( G(d_0) = C > 0 \) and \( Q + C(l - \sigma_2^2)/\delta_2^2 \neq 0 \).

If \( \sigma_2 < 1 \), then \( D(d) > 0 \) for all \( d \neq d_0 \) implies \( h(d) > 0 \) for all \( d \neq d_0 \), by Lemma 3-(1). Also, \( Q \) tends to a positive number implies

\[
(\sqrt{\theta_1} + \sqrt{\theta_j^j(1 + Q)} + 2\delta > 0 \text{ as } d \rightarrow d_0, \text{ and}
\]

\[
(\sqrt{\theta_1} - \sqrt{\theta_j^j(1 + Q}) \rightarrow \text{a negative number as } d \rightarrow d_0;
\]
and this result, together with (vii), means that as $d \to d_0$

$$\lim B_1 = \lim B_3 = +\infty,$$ and
$$\lim B_2 = \lim B_4 = -\infty.$$ 

By Lemma C (above), $F(d) \to 1-\alpha$ as $d \to d_0$.

If $\sigma_2 > 1$, then $D(d) < 0$ for all $d \neq d_0$.

If $\sigma_2^2 > 1 + \delta_2^2/C$, then $\lim Q < -1$, and it follows from (19) that
$$h(d_0) = h'(d_0) = 0$$ and $h''(d_0) = 2\delta_2^2(1 + \lim Q) < 0,$
and so $h(d) < 0$ for all $d \neq d_0$ in a neighborhood $N(\epsilon)$ of $d_0$.

Therefore, $F(d) = 1-\alpha$ for all $d$ in $N(\epsilon)$.

If $\sigma_2^2 < 1 + \delta_2^2/C$, then $h''(d_0) > 0$ and so $h(d) > 0$ for all $d \neq d_0$ in $N(\epsilon)$. Furthermore, $-1 < \lim Q < 0$ implies
$$\lim(\sqrt{\theta_1} \pm \sqrt{\theta_2} \sqrt{1+Q}) > 0$$ as $d \to d_0$.

It follows from this result, together with (vii), that
$$\lim B_1 = \lim B_2 = \lim B_3 = \lim B_4 = (-\infty)[\operatorname{sgn}(d-d_0)].$$

By Lemma A (above), then, $F(d) \to 1-\alpha$ as $d \to d_0$.

If $\sigma_2^2 = 1 + \delta_2^2/C$, then $d_0$ is an inflection point of $h(d)$.

For $d$ in $N(\epsilon)$ such that $h(d) \leq 0$, it follows from Theorem 2 that $F(d) \equiv 1-\alpha$. For $d$ in $N(\epsilon)$ such that $h(d) > 0$, we have $\lim Q = -1$ and so $\lim(\sqrt{\theta_1} \pm \sqrt{\theta_2} \sqrt{1+Q}) = 0 > 0$. By (vii) and Lemma A, then, $F(d) \to 1-\alpha$ as $d \to d_0$.

Case 4-C: $D'(d_0) = \delta_2 = 0$.

By (iv), $h'(d_0) = C[D'(d_0)] = 0 = h(d_0)$.

Since, by (iii), $v_0 = 0$, it follows that
$$v(d) \equiv \delta_1 + \delta_2 = 0$$ for all $d$.

$$((B_1, B_2)) = \pm[\operatorname{sgn} D \sqrt{\theta_2} \sqrt{h/D^2}],$$
$$((B_3, B_4)) = \pm[\operatorname{sgn} D \sqrt{\theta_1} \sqrt{h/D^2}].$$

If $\alpha = 1/2$, then by Corollary 2 $h(d) > 0$ for all $d$ in $I_1 \cup I_2$.

Also, $C = 0$ implies $G(d) \equiv C + \ln(\theta_2/\theta_1) = \ln(1 + D/\theta_1)$, so that $h(d) \equiv v^2 + DG = D \ln(1 + D/\theta_1)$
$$= D[\theta_1 - (1/2)(D/\theta_1)^2(1+x^*)^{-2}]$$
where $x^*$ is between 0 and $D/\theta_1$, which tends to 0 as $d \to d_0$.

Thus $h/D^2 = \theta_1^{-1} + O(D) \to \theta_1^{-1}$ as $d \to d_0$.

It follows from this and (x) that as $d \to d_0$,
\(((\lim B_1,\lim B_2)) = ((\lim B_3,\lim B_4)) = \left\{ \begin{array}{ll} \pm 1 & \text{in } I_1 \\ \pm (-1) & \text{in } I_2 \end{array} \right\}.\]

By Lemma B, then, \(F(d) \to 1/2 = 1-\alpha\) as \(d \to d_0\).

**If \(\alpha > 1/2\), then \(C>0\) implies \(h''(d_0) = 2C(1-\sigma_2^2) > 0\).**

**If \(\sigma_2 < 1\), then for all \(d\) in \(N(\epsilon)\) we have \(D(d) > 0\) and \(h(d) > 0\).**

Therefore \(\lim(h/d^2) = h''(d_0)/0 = +\infty\),

which, together with \((x)\), implies

\[((\lim B_1,\lim B_2)) = \pm \infty = ((\lim B_3,\lim B_4)).\]

By Lemma C, \(F(d) \to 1-\alpha\) as \(d \to d_0\).

**If \(\sigma_2 > 1\), then \(h''(d_0) < 0\) implies \(h(d) < 0\) for all \(d\) in \(N(\epsilon)\), and so \(F(d) \equiv 1-\alpha\) for all \(d\) in \(N(\epsilon)\).**

\[\square\]

Corollaries 3 and 4 give relationships between the two misclassification functions, \(F\) and \(F^0\):

**COROLLARY III.3:** For a given vector of parameters \(\pi' = (\delta_1,\delta_2,\sigma_1,\sigma_2,\rho_1,\rho_2,\alpha)\), \(F(d|\pi) \to F^0(\alpha,\delta_1,\sigma_1)\) as \(d \to 0\).

**proof:** By Theorem 3, \(F(d|\pi) \to F(0|\pi)\) as \(d \to 0\).

When \(d=0\), then \(v = \delta_1, \theta_1 = \sigma_1^2, \theta_2 = 1, G = \delta_1^2 / \sigma_1^2, D = 1 - \sigma_1^2\), and \(h(0) = h(\alpha,\delta_1,\sigma_1)\), as shown in Corollary 1.

Substituting these values into (17) and comparing with (13), we find that \(B_1(0|\pi) = A_1(\alpha,\delta_1,\sigma_1)\) for every \(i = 1, 2, \ldots, 6\).

By Theorems 1 and 2, then, \(F(0|\pi) = F^0(\alpha,\delta_1,\sigma_1)\).

\[\square\]

**THEOREM III.4:** For a given vector of parameters \(\pi' = (\delta_1,\delta_2,\sigma_1,\sigma_2,\rho_1,\rho_2,\alpha)\), define the corresponding vector \(\pi'^* = (\delta_2,\delta_1,\sigma_2,\sigma_1,\rho_1,\rho_2,\alpha)\).

Then for every \(d \neq 0\), \(F(d|\pi) = F(d^{-1}|\pi^*)\).

**proof:** By definition, \(F(d|\pi)\) is the probability of misclassification for the optimal decision rule based on the linear function 
\(Y = X_1 + dX_2\).

If \(X_1\) and \(X_2\) are interchanged, then \(\pi^*\) is the vector of parameters specifying the bivariate normal distributions of \((X_2,X_1)\) in
the two populations.

Consider the linear function \( Y^* = X_2 + d^{-1}X_1 \).

By definition, \( F(d^{-1} | \pi^*) \) is the probability of misclassification for the optimal decision rule based on \( Y^* \).

Now, \( Y = dY^* \).

By Lemma 1, the probability of misclassification for the optimal decision rule based on \( Y \) is \( F(d^{-1} | \pi^*) \).

Therefore, \( F(d | \pi) = F(d^{-1} | \pi^*) \).

\[ \square \]

**COROLLARY III.4:** For a given vector of parameters \( \pi' = (\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha) \), \( F(d | \pi) \rightarrow F^0(\alpha, \delta_2, \sigma_2) \) as \( d \rightarrow \pm \infty \).

**proof:** By Theorem 4, \( F(d | \pi) = F(d^* | \pi^*) \), where \( d^* = d^{-1} \) and \( \pi^* = (\delta_2, \delta_1, \sigma_2, \sigma_1, \rho_1, \rho_2, \alpha) \). Therefore,

\[
\begin{align*}
\lim_{d \rightarrow +\infty} F(d | \pi) &= \lim_{d^* \rightarrow 0} F(d^* | \pi^*) \\
\lim_{d \rightarrow -\infty} F(d | \pi) &= \lim_{d^* \rightarrow 0} F(d^* | \pi^*)
\end{align*}
\]

by Corollary 3. \( \square \)

Corollary 5 gives an important property of \( F \) when \( \delta_1 = \delta_2 \) and \( \sigma_1 = \sigma_2 \).

**COROLLARY III.5:** If \( \delta_1 = \delta_2 \) and \( \sigma_1 = \sigma_2 \), then \( F(d | \pi) = F(d^{-1} | \pi) \).

**proof:** Since \( \pi = \pi^* \) when \( \delta_1 = \delta_2 \) and \( \sigma_1 = \sigma_2 \), the result follows immediately from Theorem 4. \( \square \)

The following lemma is concerned with the absolute maximum value of \( F(d | \pi) \) and the conditions under which it is attained.

**LEMMA III.5:**

1. \( F(d | \pi) = 1 - \alpha \) iff the condition, \( D(d | \pi) \leq 0 \) and \( h(d | \pi) \leq 0 \), holds. (2) If this condition does not hold, then \( F(d | \pi) < 1 - \alpha \).

3. If \( \alpha = 1/2 \), this condition holds iff \( D(d | \pi) = v(d | \pi) = 0 \).

**proof:** The optimal decision rule based on \( Y = X_1 + dX_2 \) partitions the real line into three cells:

- \( C_1 \) has the property that \( \alpha g_1(y) > (1 - \alpha) g_2(y) \); i.e. choose Pop. 1.
- \( C_2 \) has the property that \( \alpha g_1(y) < (1 - \alpha) g_2(y) \); i.e. choose Pop. 2.
- \( C_3 \) has the property that \( \alpha g_1(y) = (1 - \alpha) g_2(y) \); i.e. choose Pop. 2.
The probability of misclassification is
\[
F(d) = a \int_{C_2} g_1(y) \, dy + (1-a) \int_{C_1} g_2(y) \, dy,
\]
since \( C_3 \) contains at most two points. Thus \( F(d) = 1-a \) if \( C_2 \) is empty, and
\[
F(d) < (1-a) \int_{C_2} g_2(y) \, dy + (1-a) \int_{C_1} g_2(y) \, dy
\]
by definition of \( C_2 \) if \( C_2 \) is nonempty. From Section 3.2 we see that \( C_2 \) is empty iff \( D \leq 0 \) and \( h \leq 0 \). This establishes Statements (1) and (2). Statement (3) follows directly from Corollary 2.

\[
\square
\]

3.5 Comparison of the OLDF with Fisher's BLDF.

Recall that in Section 2.1 we considered the linear function proposed by Fisher for use in cases when the optimal discriminant function is not linear. An interesting question, which, unfortunately, we cannot fully explore here, is: under what conditions is the BLDF identical with the optimal linear discriminant function? (We do not go into this matter in exhaustive detail because this dissertation is primarily concerned with discriminant functions to be used when the data have a truncated distribution—-and this chapter does not deal with truncated distributions.)

3.5.1 The BLDF and the MBLDF.

**Lemma III.6:** Under the Basic Assumptions listed in Section 3.1, Fisher's "best linear" discriminant function, denoted BLDF, is
\[
\begin{cases}
X_1 + \frac{\alpha \delta_2 - \sigma_1}{\beta_1 - \sigma_2} X_2 & \text{if } \beta_1 \neq \sigma_2 \\
X_2 & \text{otherwise}
\end{cases}
\]
where \( \alpha = \sigma_1^2 + 1, \beta = \sigma_2^2 + 1, \) and \( \gamma = \rho_2 + \rho_1 \sigma_1 \sigma_2. \)

**Proof:** Lemma II.1 states that the BLDF has coefficient vector \( \eta = \Theta^{-1} X, \)
where \( c \) is any nonzero real number, \( \Theta = \text{Var}_1 X + \text{Var}_2 X \), and 
\( y = \text{E}_1(X) - \text{E}_2(X) \). Under the Basic Assumptions, then,

\[
\Theta = \Sigma + R = \begin{pmatrix} \sigma_1^2 + 1 & \rho_2 \rho_1 \sigma_2 \\ \rho_2 \rho_1 \sigma_2 & \sigma_2^2 + 1 \end{pmatrix} = \begin{pmatrix} a & c \\ c & b \end{pmatrix},
\]

\( y = \delta - 0 \).

Since \( \Theta \) is positive definite, we have

\[
n' = \delta' \begin{pmatrix} b & -c \\ -c & a \end{pmatrix} = \begin{cases} (1, (a\delta_2 - c\delta_1) / (b\delta_1 - c\delta_2)) & \text{if } b\delta_1 \neq c\delta_2, \\
(0, 1) & \text{otherwise}. \end{cases}
\]

We can then calculate the total probability of misclassification for the optimal decision rule based on the BLDF from either Theorem 1 or Theorem 2.

After comparing the misclassification probabilities for the OLDF and the BLDF using a number of different parameter vectors, we wondered whether the BLDF could be improved if it were modified to take into account the fact that when \( \alpha > 1/2 \) the mixture does not contain equal parts of the two populations.

**Lemma III.7:** The linear function \( n^* X \) which maximizes the quantity,

\[
R^{*2} = \frac{[\text{E}_1(n'X) - \text{E}_2(n'X)]^2}{a\text{Var}_1(n'X) + (1-a)\text{Var}_2(n'X)} = \frac{(n'X)^2}{n'\Theta n},
\]

has coefficient vector \( n^{*'} \) satisfying

\[
\Theta^* = \begin{pmatrix} a^* & c^* \\ c^* & b^* \end{pmatrix} = \alpha \Sigma + (1-\alpha) R,
\]

\( y = \text{E}_1(X) - \text{E}_2(X) = \delta \).

Because \( \text{E}_2(X) = 0 \), the linear function \( n^{*'} X \) also maximizes the quantity
\[ R^{**2} = \frac{[aE_1(n'X) - (1-\alpha)E_2(n'X)]^2}{n'\varrho n}. \]

We shall call \( n''X \) the Modified BLDF, denoted MBLDF.

**proof:** Strictly analogous to the proof of Lemma 6.

\[ \square \]

### 3.5.2 Numerical Examples Comparing the OLDF, the BLDF, and the MBLDF.

In this subsection are given nine examples showing the three linear functions and their probabilities of misclassification for a given set of parameters \( n' = (\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha) \). The graph of the misclassification probability function \( F(d|\pi) \) is also given for each parameter set, for values of \( d \) between -12.5 and +12.5.

**Example 1:** \( n' = (1, 1, \sqrt{2}, 1/2, -3/4, -1/4, 0.65) \)

\[ \varrho = \Sigma + R = \begin{bmatrix} 3.0 & -0.78033 \\ -0.78033 & 1.25 \end{bmatrix} \approx \begin{bmatrix} a & c \\ c & b \end{bmatrix}. \]

\[ n' = \delta' \varrho^{-1} = (1, 1) \begin{bmatrix} b & -c \\ -c & a \end{bmatrix} = (1, (a-c)/(b-c)) \approx (1, 1.86). \]

Therefore, the BLDF is \( X_1 + 1.86X_2. \)

\[ \varrho^* = a\Sigma + (1-\alpha)R = 0.65\begin{bmatrix} 2.0 & -0.53033 \\ -0.53033 & 0.25 \end{bmatrix} + 0.35\begin{bmatrix} 1.0 & -0.25 \\ -0.25 & 1.0 \end{bmatrix} \]

\[ = \begin{bmatrix} 1.65 & -0.4322 \\ -0.4322 & 0.5125 \end{bmatrix}. \]

Thus \( n'' = (1, (1.65+.4322)/(.51+.4322)) \), and the MBLDF \( \approx X_1 + 2.20X_2. \)

From the graph of the function \( F(d|\pi_1) \) in Figure 1, we see that the minimum value of \( F(d) \) is attained when \( d = 3 \); that is, the OLDF \( \approx X_1 + 3X_2. \) From this graph we can also read the probabilities of misclassification corresponding to the three linear discriminant functions: \( F(1.86) \approx .1175, \) \( F(2.20) \approx .1080, \) and \( F(3) \approx .1015. \)

These results are presented in tabular form, using the more precise figures obtained from the computer evaluation of \( F(d|\pi_1): \)
<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLDF = $X_1 + 1.86 , X_2$</td>
<td>0.1171</td>
</tr>
<tr>
<td>MBLDF = $X_1 + 2.20 , X_2$</td>
<td>0.1081</td>
</tr>
<tr>
<td>OLDF = $X_1 + 2.96 , X_2$</td>
<td>0.1017</td>
</tr>
</tbody>
</table>

**Example 2:** $\pi_2' = (1, 1, \sqrt{2}, 1/2, 3/4, 1/4, 0.65)$

The BLDF and the MBLDF are calculated in the way demonstrated in Example 1, and the approximate OLDF and the misclassification probabilities for the optimal decision rules based on the three linear functions can be read from the graph of $F(d|\pi_2)$ in Figure 1.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLDF = $X_1 + 4.74 , X_2$</td>
<td>0.2066</td>
</tr>
<tr>
<td>MBLDF = $X_1 + 15.17 , X_2$</td>
<td>0.1898</td>
</tr>
<tr>
<td>OLDF = $X_1 - 4.71 , X_2$</td>
<td>0.1609</td>
</tr>
</tbody>
</table>

**Example 3:** $\pi_3' = (1, 1, \sqrt{2}, 1, 1/4, 1/4, 0.65)$

See the graph of $F(d|\pi_3)$ in Figure 1.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLDF = $X_1 + 1.72 , X_2$</td>
<td>0.2632</td>
</tr>
<tr>
<td>MBLDF = $X_1 + 1.95 , X_2$</td>
<td>0.2624</td>
</tr>
<tr>
<td>OLDF = $X_1 + 2.42 , X_2$</td>
<td>0.2620</td>
</tr>
</tbody>
</table>

**Example 4:** $\pi_4' = (1/2, 1/2, 1/2, 1, 3/4, 1/4, 0.65)$

See the graph of $F(d|\pi_4)$ in Figure 2.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLDF = $X_1 + 0.45 , X_2$</td>
<td>0.2862</td>
</tr>
<tr>
<td>MBLDF = $X_1 + 0.27 , X_2$</td>
<td>0.2733</td>
</tr>
<tr>
<td>OLDF = $X_1 - 0.37 , X_2$</td>
<td>0.1967</td>
</tr>
</tbody>
</table>
Example 5: $\pi_5' = (1, 1/2, 1/2, 1, 3/4, 1/4, 0.65)$

See the graph of $F(d|\pi_5)$ in Figure 2.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BLDF = X_1$</td>
<td>0.1815</td>
</tr>
<tr>
<td>$MLD_2 = X_1 - 0.09 X_2$</td>
<td>0.1709</td>
</tr>
<tr>
<td>$OLDF = X_1 - 0.32 X_2$</td>
<td>0.1539</td>
</tr>
</tbody>
</table>

Example 6: $\pi_6' = (1, 1, 1/2, 1, 1/4, 3/4, 0.65)$

See the graph of $F(d|\pi_6)$ in Figure 2.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BLDF = X_1 + 0.33 X_2$</td>
<td>0.1807</td>
</tr>
<tr>
<td>$MLD_2 = X_1 + 0.26 X_2$</td>
<td>0.1787</td>
</tr>
<tr>
<td>$OLDF = X_1 + 0.15 X_2$</td>
<td>0.1773</td>
</tr>
</tbody>
</table>

Example 7: $\pi_7' = (1, 1/2, 1/2, 1/2, 3/4, 1/4, 0.65)$

See the graph of $F(d|\pi_7)$ in Figure 3.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BLDF = X_1 + 0.18 X_2$</td>
<td>0.1868</td>
</tr>
<tr>
<td>$MLD_2 = X_1 + 0.12 X_2$</td>
<td>0.1848</td>
</tr>
<tr>
<td>$OLDF = X_1 - 0.74 X_2$</td>
<td>0.1654</td>
</tr>
</tbody>
</table>

Example 8: $\pi_8' = (1, 1, 1, 1, 1/2, 1/4, 0.65)$

See the graph of $F(d|\pi_8)$ in Figure 3.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BLDF = X_1 + X_2$</td>
<td>0.2545</td>
</tr>
<tr>
<td>$MLD_2 = X_1 + X_2$</td>
<td>&quot;</td>
</tr>
<tr>
<td>$OLDF = X_1 + X_2$</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
Example 9: \[ \pi_9' = (1, 1, 1/2, 1/2, -3/4, -1/4, 0.65) \]

See the graph of \( F(d|\pi) \) in Figure 3.

<table>
<thead>
<tr>
<th>Linear Discriminant Function</th>
<th>Corresponding Misclassification Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLDF = ( X_1 + X_2 )</td>
<td>0.0633</td>
</tr>
<tr>
<td>MBLDF = ( X_1 + X_2 )</td>
<td>&quot;</td>
</tr>
<tr>
<td>OLDF = ( X_1 + X_2 )</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

3.5.3 Conclusions Drawn from the Examples.

Examination of the first seven examples suggests the following conclusions:

1.) The coefficients of \( X_2 \) in the BLDF and the MBLDF are not necessarily very close in value to the coefficient of \( X_2 \) in the OLDF.

2.) How close in value the misclassification probabilities for the Fisher estimators are to that of the OLDF depends not on the relative closeness of the coefficients of \( X_2 \) but on the shape of the function \( F(d|\pi) \) for the given parameter vector \( \pi \).

Regarding conclusion (2), compare the tables for Examples 3 and 4: in Example 3, the coefficients of the BLDF and the OLDF differ by 0.70, while the corresponding misclassification probabilities differ by a mere 0.0012; whereas in Example 4, the coefficients of the MBLDF and the OLDF differ by only .64 but their misclassification probabilities differ by 0.0766. A brief glance at the graphs of \( F(d|\pi) \) for the two parameter sets shows that in a neighborhood of the minimum, \( F(d|\pi_3) \) changes slowly, whereas \( F(d|\pi_4) \) changes relatively fast.

Another conclusion suggested by the data is this: while the MBLDF seems generally to be somewhat "better" than the BLDF, it probably isn't enough better to justify the extra computation involved. Example 2 shows, too, that there is no guarantee that the coefficient of \( X_2 \) in the MBLDF will lie between those in the OLDF and the BLDF.
3.6 Linear Discrimination When $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$.

Having observed, in Examples 8 and 9, the three linear discriminant functions to be equal in specific cases when $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$, we wonder whether this will always be true. Lemma 8 shows that the BLDF and the MELDF are always $X_1+X_2$ in this case.

**Lemma III.8:** If $\delta_1=\delta_2>0$ and $\sigma_1=\sigma_2$, then BLDF = MELDF = $X_1 + X_2$.

**Proof:** If $\delta_1=\delta_2=\delta>0$, and $\sigma_1=\sigma_2=\sigma$, then (using the notation of Lemma 7)

$$b^\#\delta_1 = c^\#\delta_2 \iff a\sigma^2 + (1-a) = a\rho_1 \sigma^2 + (1-a)\rho_2$$

$$\iff \sigma^2 = -[(1-a)(1-\rho_2)]/[a(1-\rho_1)] < 0.$$

Therefore, $b^\#\delta_1 \neq c^\#\delta_2$, and so by Lemma 7 the MELDF is

$$X_1 + [(a^*-c^*)/(b^*-c^*)]X_2 = X_1 + X_2,$$

because $a^* = b^* = a\sigma^2 + (1-a)$.

If $a=1/2$, the $\Theta^* = a\alpha^* + (1-a)\alpha = \Theta$. Therefore, this proof also establishes that the BLDF is $X_1 + X_2$.

Unfortunately, the OLDF isn't always equal to $X_1 + X_2$ when $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$. Consider the examples presented in the following table and in Figure 4.

<table>
<thead>
<tr>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\alpha$</th>
<th>Linear Discriminant Function</th>
<th>Prob. of Misclass.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1.6</td>
<td>.25</td>
<td>.75</td>
<td>.65</td>
<td>BLDF = $X_1 + X_2$</td>
<td>0.3140</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLDF = $X_1 - X_2$</td>
<td>0.2941</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.0</td>
<td>2.0</td>
<td>.25</td>
<td>.75</td>
<td>.65</td>
<td>BLDF = $X_1 + X_2$</td>
<td>0.3260</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLDF = $X_1 - X_2$</td>
<td>0.2566</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2.0</td>
<td>2.0</td>
<td>-.25</td>
<td>.75</td>
<td>.65</td>
<td>BLDF = $X_1 + X_2$</td>
<td>0.3094</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLDF = $X_1 - X_2$</td>
<td>0.2162</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>.75</td>
<td>.25</td>
<td>.65</td>
<td>BLDF = $X_1 + X_2$</td>
<td>0.1526</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLDF = $X_1 - X_2$</td>
<td>0.1497</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>.75</td>
<td>.25</td>
<td>.65</td>
<td>BLDF = $X_1 + X_2$</td>
<td>0.2408</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLDF = $X_1 - X_2$</td>
<td>0.2318</td>
</tr>
</tbody>
</table>
These examples suggest two more questions:

1) Is a linear function other than \( X_1 \pm X_2 \) ever the OLDF when \( \sigma_1=\sigma_2 \) and \( \delta_1=\delta_2 \) ?

2) Under what conditions will the OLDF be \( X_1 - X_2 \) ?

Attempts to answer these questions have produced the partial answers stated below. For the remainder of this section it will be assumed that \( \pi \) is a vector of parameters having the property that \( \delta_1=\delta_2=\delta \) and \( \sigma_1=\sigma_2=\sigma \). [Lemmas 9-14 referred to below are stated and proved in Subsection 3.6.1.]

1.) Because \( F(d|\pi) = F(d^{-1}|\pi) \) for all \( d \neq 0 \) [Corr. 5] and because
\[
F'(1|\pi) = F'(-1|\pi) = 0 \quad \text{[Lemma 9]},
\]
\( F(d|\pi) \) has a maximum or a minimum at each of the points \( d = \pm 1 \). Evaluation of \( F(d|\pi) \) for scores of parameter sets has never produced an instance of a minimum between \( +1 \) and \( -1 \).

2.) Lemmas 10 and 11 give necessary and sufficient conditions on the parameters so that, respectively, \( F(-1|\pi) = 1-\alpha \) and \( F(+1|\pi) = 1-\alpha \), which is the absolute maximum attained by \( F(d|\pi) \) [Lemma 5].

Lemma 12 gives necessary and sufficient conditions so that
\[
F(+1|\pi) = F(-1|\pi) = 1-\alpha.
\]
a.) Using the further assumption that \( \rho_1=\rho_2 \), Lemma 14 shows that
\[
F(+1|\pi) = F(-1|\pi) = 1-\alpha \quad \text{iff} \quad F(d|\pi) = 1-\alpha \quad \text{for all} \quad d.
\]
That is, if \( F(d|\pi) \) attains its absolute maximum value at both of the points \( d = \pm 1 \), then it has no minimum value.

b.) This result could not be established analytically for the case when \( \rho_1 \neq \rho_2 \), but all of the numerical examples with parameters chosen to make \( F(d|\pi) = 1-\alpha \) at \( d = \pm 1 \) showed \( F(d|\pi) = 1-\alpha \).

3.6.1 Some Properties of \( F \) when \( \delta_1=\delta_2 \) and \( \sigma_1=\sigma_2 \).

In this subsection are presented the lemmas which deal with the behavior of \( F(d|\pi) \) when \( \pi \) has the property that \( \delta_1=\delta_2 \) and \( \sigma_1=\sigma_2 \).

**LEMMA III.9:** For all parameter sets with \( \delta_1=\delta_2 \) and \( \sigma_1=\sigma_2 \),
\[
F'(-1|\pi) = F'(1|\pi) = 0,
\]
where \( F'(d|\pi) \) denotes the first derivative of \( F(d|\pi) \) with respect to \( d \). Consequently, \( F(d|\pi) \) has a maximum or a minimum at each of the points \( d = \pm 1 \).
proof: By Corollary 5, $F(d|\pi) = F(d^{-1}|\pi)$ for all $d \neq 0$. In particular,
\[
F\left(\frac{n-1}{n} \mid \pi \right) = F\left(\frac{n}{n-1} \mid \pi \right) \text{ for all } n=2,3, \ldots
\]

Since, by Theorem 3, $F(d|\pi)$ is continuous, it follows from Rolle's Theorem that $F'(+1|\pi) = 0$.

Similarly, $F'(-1|\pi) = 0$.

Because $F(d|\pi) = F(d^{-1}|\pi)$, $d = \pm 1$ cannot be inflection points.

\[\square\]

**LEMMA III.10:** If $\pi$ is a parameter vector with $\delta_1=\delta_2=\delta>0$ and $\sigma_1=\sigma_2=\sigma$, then $F(-1|\pi) = 1-\alpha$ iff $\sqrt{(1-\rho_2)/(1-\rho_1)} \leq \sigma \leq [\alpha/(1-\alpha)]\sqrt{(1-\rho_2)/(1-\rho_1)}$.

Note that there are no restrictions on the size of $\delta$.

**proof:** When $\delta_1=\delta_2=\delta$, then $\nu(d) = \delta_1+\delta_2 = \delta(d+1)$.

Therefore, $\nu(-1) = 0$ for all real $\delta$.

By Lemma 5, $F(d|\pi) = 1-\alpha$ iff $D(d) \leq 0$ and $h(d) \leq 0$.

From Lemma 3, parts (1)-(3), it follows that when $\nu(d)=0$, then $h(d) \leq 0$ iff $\theta_2 \leq \theta_1 \leq \theta_1[\alpha/(1-\alpha)]^2$.

Note that $\theta_2 \leq \theta_1$ iff $D(d) \leq 0$.

So $h(-1) \leq 0$ iff $\theta_2(-1) \leq \theta_1(-1) \leq \theta_1(-1)[\alpha/(1-\alpha)]^2$,

which gives the result.

\[\square\]

**LEMMA III.11:** If $\pi$ is a parameter vector with $\sigma_1=\sigma_2=\sigma$ and $\delta_1=\delta_2=\delta>0$,

then $F(+1|\pi) = 1-\alpha$ iff both of the following conditions hold:

1.) $\sqrt{(1+\rho_2)/(1+\rho_1)} \leq \sigma \leq [\alpha/(1-\alpha)]\sqrt{(1+\rho_2)/(1+\rho_1)}$, and

2.) $0 \leq \delta \leq \delta(\pi) \equiv \left\{ \frac{a^2(1+\rho_1)-2(1+\rho_2)}{2} \leq \ln \left[ \frac{a^{1-\alpha}}{\sigma} \sqrt{\frac{1}{1+\rho_1}} \right] \right\}$. \quad (24)

**proof:** $\nu(d) = \delta(d+1)$ implies $\nu(+1) = 2\delta$.

$D(d) = (d^2 + 2\rho_2d + 1) - \sigma^2(d^2 + 2\rho_1d + 1)$ implies $D(+1) = 2[(1+\rho_2)-\sigma^2(1+\rho_1)] \leq 0$ iff $\sigma^2 \geq (1+\rho_2)/(1+\rho_1)$.

$G(+1) \equiv C + \ln \left( \frac{\theta_2(+1)}{\theta_1(+1)} \right) = \ln \left( \left[ \frac{a^{1-\alpha}}{\sigma} \sqrt{\frac{1+\rho_2}{1+\rho_1}} \right] \right)$ implies $G(+1) \leq 0$ iff $\sigma^2 \geq [a^2(1+\rho_2)]/(1+\rho_1)$.

By definition, $h(+1) = \nu^2(+1) + D(+1) \cdot G(+1) = 4\delta^2 + D(+1) \cdot G(+1)$.

a.) When $\sigma < \sqrt{(1+\rho_2)/(1+\rho_1)}$, then $D(+1) > 0$ implies $h(+1) > 0$.
by Lemma 3-(l).

b.) When \( \sqrt{(1+\rho_2)/(1+\rho_1)} \leq \sigma \leq [\alpha/(1-\alpha)]\sqrt{(1+\rho_2)/(1+\rho_1)} \),
then \( D(+1) \leq 0 \) and \( G(+1) \geq 0 \) implies \( D(+1) \cdot G(+1) \leq 0 \).
Thus, \( h(+1) \leq 0 \) iff \( 4\delta^2 \leq -D(+1) \cdot G(+1) \)
iff \( 0 \leq \delta \leq \delta(\pi) \), defined in (24).

c.) When \( \sigma > [\alpha/(1-\alpha)]\sqrt{(1+\rho_2)/(1+\rho_1)} \), then \( D(+1) < 0 \) and
\( G(+1) < 0 \) implies \( h(+1) > 0 \).

The result follows from (a)-(c) and Lemma 5.

\[ \]
\[
\frac{1+\rho_2}{1+\rho_1} \leq \frac{\sigma^2(1-\rho_2)}{[1-\sigma^2(1-\rho_1)]},
\]

i.e. \(\rho_1(1-K\rho_2) \geq \rho_2-K\), \(1\)

where \(K = \frac{\sigma^2-(1-\sigma^2)}{[\sigma^2+(1-\sigma^2)]}\) is between 0 and 1 since \(1/2 < \sigma < 1\).

Now, \(1-K\rho_2 \leq 0\) implies \(\rho_2 \geq (1/K) > 1\), which is impossible.

Therefore, \((1)\) holds iff \(\rho_1 \geq (\rho_2-K)/(1-K\rho_2)\).

This lower bound, denoted \(L\), say, is contained in the interval \((-1, \rho_2)\); for \(L \leq 1\) implies \(\rho_2 \leq 1\), and \(L > \rho_2\) implies \(\rho_2^2 \geq 1\).

**Case 3:** \(\rho_1 > \rho_2\)

By an argument strictly analogous to that of Case 2, we can prove that \(I_1 I_2\) is nonempty iff \(\rho_1 \leq (\rho_2+K)/(1+K\rho_2)\) and that this upper bound belongs to the interval \((\rho_2, 1)\).

\[\Box\]

**LEMMA III.13:** If \(\pi\) is a parameter vector such that \(\delta_1 = \delta_2 = \delta\) and \(\sigma_1 = \sigma_2 = \sigma\) and \(\alpha = 1/2\), then \((1)\) the condition, \(F(+1|\pi) = F(-1|\pi) = 1/2\) can never hold; \((2)\) \(F(-1|\pi) = 1/2\) iff \(\sigma^2 = (1-\rho_2)/(1-\rho_1) \neq 1\); \(\sigma^2 = (1+\rho_2)/(1+\rho_1) \neq 1\) and \(\delta = 0\).

**proof:** Lemma 5 states that when \(\alpha = 1/2, F(d) = 1/2\) iff \(D(d) = v(d) = 0\).

When \(\delta_1 = \delta_2 = \delta\), \(v(d) = \delta(d+1) = 0\) iff \(d = -1\) or \(\delta = 0\).

By Lemma 16 (below), \(D(+1) = 0\) iff \(\sigma^2 = (1+\rho_2)/(1+\rho_1) \neq 1\), and \(D(-1) = 0\) iff \(\sigma^2 = (1-\rho_2)/(1-\rho_1) \neq 1\).

\[\Box\]

Table 2, on the last three pages of this section, shows numerical examples of the bounds on \(\rho_1\), \(\sigma\), and \(\delta\), given in Lemma 12, calculated for selected values of \(\rho_2\), \(\rho_1\), and \(\sigma\).

**LEMMA III.14:** If \(\pi\) is a parameter vector such that \(\sigma_1 = \sigma_2 = \sigma\), \(\rho_1 = \rho_2 = \rho\), \(\delta_1 = \delta_2 = \delta \geq 0\), and \(\alpha > 1/2\), then \(F(d|\pi) \equiv 1-\alpha\) for all \(d\) iff the following conditions are satisfied: \((1)\) \(1 < \sigma \leq \alpha/(1-\alpha)\) and \((2)\) \(0 \leq \delta \leq \sqrt{\frac{1}{2}(1+\rho)(G^2-1)(C-Ln \sigma^2)}\), where \(C = 2Ln[\alpha/(1-\alpha)]\).

In other words, \(F(-1|\pi) = F(+1|\pi) = 1-\alpha\) iff \(F(d|\pi) \equiv 1-\alpha\) for all \(d\).

**proof:** Suppose the above conditions are satisfied.

Then, by Lemma 17 (below), \(\rho_1 = \rho_2\) implies \(G(d) \equiv C-Ln \sigma^2\) for all \(d\). Let \(G \equiv C - Ln \sigma^2\).
\( \rho_1 = \rho_2 \) also implies \( D(d) = (1-\sigma^2)(d^2 + 2\rho d + 1) = A(d^2 + 2\rho d + 1) \).

Therefore, \( h(d) \equiv [v(d)]^2 + D(d) \cdot G(d) \)
\[ = (\delta^2 - |A|G)d^2 + 2(\delta^2 - \rho |A|G)d + (\delta^2 - |A|G) \]
\[ = Ud^2 + 2Vd + U. \]

Condition (2) can be rewritten: \( 0 \leq \delta \leq \sqrt{(1+\rho)|A|G}/2 \),
and so \( U = \delta^2 - |A|G < 0 \).

Therefore \( h(d) \) is an arch-shaped quadratic in \( d \) and has
\[ \max h(d) = h(-V/U) = \frac{(1-\rho)|A|G[2\delta^2 - (1+\rho)|A|G]}{|A|G - \delta^2} \leq 0 \]
by Condition (2).

Thus, \( h(d) \leq 0 \) for all \( d \).

By Theorem 2, then, \( F(d) \equiv 1-\alpha \) for all \( d \).

Suppose, now, that \( F(d) \equiv 1-\alpha \) for all \( d \).

Then by Lemma 12, the facts that \( F(+1|\pi) = F(-1|\pi) = 1-\alpha \) and \( \rho_1 = \rho_2 \) imply that Conditions (1) and (2) are satisfied, since these conditions are identical with Conditions (2) and (3) of that lemma in the special case when \( \rho_1 = \rho_2 \).

\[ \square \]

The next three lemmas have been relegated to the end because, though necessary, they deal with the subordinate functions, not with \( F(d) \). As in Section 3.4, primes and double-primes on functions denote the first- and second-partial derivatives, respectively, with respect to \( d \).

**Lemma III.15:** For all parameter vectors \( \pi \) with \( \sigma_1 = \sigma_2 \), the following statements hold for \( i=1,2 \):

1. \( \theta_i'(+1|\pi) = \theta_i(+1|\pi) \);
2. \( \theta_i'(-1|\pi) = -\theta_i(-1|\pi) \);
3. \( \theta_1''(d|\pi) = 2\sigma^2 \) for all \( d \), and \( \theta_2''(d|\pi) = 2 \) for all \( d \).

**Proof:** When \( \sigma_1 = \sigma_2 \), then
\[
\begin{align*}
\{ \theta_1(d) &= \sigma^2(d^2 + 2\rho d + 1) \\
\theta_2(d) &= d^2 + 2\rho d + 1 \}
\end{align*}
\]

All results follow from differentiation of these quadratics and substitution.

\[ \square \]
**Lemma III.16:** For all parameter vectors \( \pi \) with \( \sigma_1 = \sigma_2 \), the following statements hold:

1. \( D(+1|\pi) = 0 \) iff \( \sigma^2 = (1+\rho_2)/(1+\rho_1) \) and \( \rho_1 \leq \rho_2 \).
2. \( D(-1|\pi) = 0 \) iff \( \sigma^2 = (1-\rho_2)/(1-\rho_1) \) and \( \rho_1 \geq \rho_2 \).
3. \( D'(+1|\pi) = D(+1|\pi) \) and \( D'(-1|\pi) = -D(-1|\pi) \).
4. If \( D(+1|\pi) = 0 \), then \( D(d|\pi) \) has a maximum or minimum at \( d = +1 \) according as \( \rho_1 < \rho_2 \) or \( \rho_1 > \rho_2 \).
5. If \( D(-1|\pi) = 0 \), then \( D(d|\pi) \) has a maximum or minimum at \( d = -1 \) according as \( \rho_1 > \rho_2 \) or \( \rho_1 < \rho_2 \).

**Proof:** By definition, \( D(d) = \theta_2(d) - \theta_1(d) \).

From (25), then,

\[
\begin{align*}
D(+1) &= 2[(1+\rho_2) - \sigma^2(1+\rho_1)] \\
D(-1) &= 2[(1-\rho_2) - \sigma^2(1-\rho_1)]
\end{align*}
\]

(26)

Statements (1) and (2) follow immediately from this and the assumption that \( \mathbb{E} \neq \mathbb{R} \).

Statement (3) follows directly from Lemma 15-(1), (2).

From Lemma 15-(3) we have \( D''(d) = 2(1-\sigma^2) \) for all \( d \).

From this and Statements (1)-(3) we have:

- \( D(+1) = 0 \) implies \( D'(+1) = 0 \) and \( \sigma^2 = (1+\rho_2)/(1+\rho_1) \), which implies \( D''(d) = 2(\rho_1-\rho_2)/(1+\rho_1) \).
- Similarly, \( D(-1) = 0 \) implies \( D''(d) = 2(\rho_2-\rho_1)/(1-\rho_1) \).

Statements (4) and (5) follow from these two statements.

\[\square\]

**Lemma III.17:** For all parameter vectors \( \pi \) with \( \sigma_1 = \sigma_2 \), the following statements hold:

1. If \( \rho_1 = \rho_2 \), then \( G(d|\pi) = C - \ln \sigma^2 \) for all \( d \).
2. If \( \rho_1 < \rho_2 \), then \( G(d|\pi) \) has a unique maximum at \( d = +1 \) and a unique minimum at \( d = -1 \).
3. If \( \rho_1 > \rho_2 \), then \( G(d|\pi) \) has a unique maximum at \( d = -1 \) and a unique minimum at \( d = +1 \).
4. \( G''(+1|\pi) = (\rho_1-\rho_2)/[(1+\rho_1)(1+\rho_2)] \), \( G''(-1|\pi) = (\rho_2-\rho_1)/[(1-\rho_1)(1-\rho_2)] \);

that is, both are independent of \( \sigma \).
proof: By (25) we have

\[ G(d) = C + \ln \theta_2 - \ln \theta_1 = (C - \ln \sigma^2) + \ln \theta_2 - \ln \theta_1, \]

where \( Q = d^2 + 2\rho_1 d + 1. \)

Hence \[ G'(d) = \left( \frac{\theta_2'}{\theta_2} - \frac{\theta_1'}{\theta_1} \right) = \left( \frac{\theta_2'}{\theta_2} - \frac{\theta_1'}{\theta_1} \right) \frac{Q'}{Q} = \frac{2(\rho_2 - \rho_1)(1-d^2)}{(Q\theta_2)}. \]

Therefore, \( G'(d) \equiv 0 \) iff \( \rho_1 = \rho_2. \)

When \( \rho_1 = \rho_2, \) then \( Q(d) = \theta_2(d) \) for all \( d \) and so \( G(d) \equiv C - \ln \sigma^2. \)

If \( \rho_1 = \rho_2, \) then \( G'(d) = 0 \) iff \( d = \pm 1. \)

\( \rho_1 < \rho_2 \) implies \( G'(d) > 0 \) iff \( |d| < 1 \) and \( G'(d) < 0 \) iff \( |d| > 1. \)

\( \rho_1 > \rho_2 \) implies \( G'(d) > 0 \) iff \( |d| > 1 \) and \( G'(d) < 0 \) iff \( |d| < 1. \)

These three sentences establish Statements (2) and (3).

Since \( G''(d) = (2/\theta_2) - (\theta_2'/\theta_2)^2 - (2\sigma^2/\theta_1) + (\theta_1'/\theta_1)^2, \) we get Statement (4) from Lemma 15.

\[ \Box \]
<table>
<thead>
<tr>
<th>Given Values $\rho_2$</th>
<th>Corresponding Bounds on $\rho_1$</th>
<th>Selected Values of $\rho_1$</th>
<th>Corresponding Bounds on $\sigma$</th>
<th>Selected Values of $\sigma$</th>
<th>Corresponding Upper Bound on $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-.75</td>
<td>.65</td>
<td>-.92</td>
<td>1.768 $\leq \sigma \leq$ 1.773</td>
<td>1.77</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.75 ($=\rho_2$)</td>
<td>1.000 $\leq \sigma \leq$ 1.857</td>
<td>1.10</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30</td>
<td>0.248</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.77</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.85</td>
<td>0.048</td>
</tr>
<tr>
<td>-.50</td>
<td>.65</td>
<td>-.50</td>
<td>1.080 $\leq \sigma \leq$ 1.313</td>
<td>1.10</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30</td>
<td>0.078</td>
</tr>
<tr>
<td>-.34</td>
<td></td>
<td>-.34</td>
<td>1.143 $\leq \sigma \leq$ 1.143</td>
<td>1.143</td>
<td>0.007</td>
</tr>
<tr>
<td>-.50</td>
<td>.65</td>
<td>-.80</td>
<td>1.581 $\leq \sigma \leq$ 1.695</td>
<td>1.60</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.69</td>
<td>0.198</td>
</tr>
<tr>
<td>-.70</td>
<td></td>
<td>-.70</td>
<td>1.291 $\leq \sigma \leq$ 1.744</td>
<td>1.30</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.70</td>
<td>0.355</td>
</tr>
<tr>
<td>-.60</td>
<td></td>
<td>-.60</td>
<td>1.118 $\leq \sigma \leq$ 1.798</td>
<td>1.20</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.361</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.75</td>
<td>0.352</td>
</tr>
<tr>
<td>-.50</td>
<td></td>
<td>-.50 ($=\rho_2$)</td>
<td>1.000 $\leq \sigma \leq$ 1.857</td>
<td>1.01</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.10</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.365</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.77</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.857</td>
<td>0.010</td>
</tr>
<tr>
<td>-.45</td>
<td></td>
<td>-.45</td>
<td>1.017 $\leq \sigma \leq$ 1.771</td>
<td>1.02</td>
<td>0.200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.10</td>
<td>0.281</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.77</td>
<td>0.022</td>
</tr>
<tr>
<td>-.25</td>
<td></td>
<td>-.25</td>
<td>1.095 $\leq \sigma \leq$ 1.516</td>
<td>1.10</td>
<td>0.362</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.516</td>
<td>0.006</td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td>0.00</td>
<td>1.225 $\leq \sigma \leq$ 1.313</td>
<td>1.25</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30</td>
<td>0.110</td>
</tr>
<tr>
<td>.065</td>
<td></td>
<td>.065</td>
<td>1.267 $\leq \sigma \leq$ 1.272</td>
<td>1.27</td>
<td>0.049</td>
</tr>
<tr>
<td>Given Values</td>
<td>Corresponding Bounds on $\rho_1$</td>
<td>Selected Values of $\rho_1$</td>
<td>Corresponding Bounds on $\sigma$</td>
<td>Selected Values of $\sigma$</td>
<td>Corresponding Upper Bound on $\delta$</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td>---------------------------</td>
<td>-------------------------------</td>
<td>---------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>.65</td>
<td>-.55</td>
<td>1.491 $\leq$ $\sigma$ $\leq$ 1.491</td>
<td>1.491</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.25</td>
<td>1.155 $\leq$ $\sigma$ $\leq$ 1.662</td>
<td>1.20</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.496</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.66</td>
<td>0.523</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00 ($=\rho_2$)</td>
<td>1.000 $\leq$ $\sigma$ $\leq$ 1.857</td>
<td>1.01</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.20</td>
<td>0.438</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.517</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.66</td>
<td>0.444</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.85</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>.75</td>
<td>.800 $\leq$ $\rho_1$ $\leq$ .800</td>
<td>-.79</td>
<td>2.182 $\leq$ $\sigma$ $\leq$ 2.242</td>
<td>2.19</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>0.134</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.24</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.55</td>
<td>1.491 $\leq$ $\sigma$ $\leq$ 2.410</td>
<td>1.50</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>0.914</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.40</td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.25</td>
<td>1.155 $\leq$ $\sigma$ $\leq$ 2.683</td>
<td>1.20</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.759</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>1.093</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.40</td>
<td>1.104</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.68</td>
<td>1.061</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00 ($=\rho_2$)</td>
<td>1.000 $\leq$ $\sigma$ $\leq$ 3.000</td>
<td>1.01</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.20</td>
<td>0.635</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>0.931</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>1.091</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.68</td>
<td>0.835</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.99</td>
<td>0.163</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.25</td>
<td>1.155 $\leq$ $\sigma$ $\leq$ 2.683</td>
<td>1.20</td>
<td>0.802</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.50</td>
<td>1.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.68</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.55</td>
<td>1.491 $\leq$ $\sigma$ $\leq$ 2.410</td>
<td>1.50</td>
<td>1.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>0.769</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.40</td>
<td>0.178</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.79</td>
<td>2.182 $\leq$ $\sigma$ $\leq$ 2.242</td>
<td>2.19</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.20</td>
<td>0.382</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.24</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>Given Values $\rho_2$</td>
<td>Corresponding Bounds on $\rho_1$</td>
<td>Selected Values of $\rho_1$</td>
<td>Corresponding Bounds on $\sigma$</td>
<td>Selected Values of $\sigma$</td>
<td>Corresponding Upper Bound on $\delta$</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------</td>
<td>----------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>$.50$</td>
<td>$.65</td>
<td>$.065</td>
<td>$1.267 \leq \sigma \leq 1.272$</td>
<td>$1.27$</td>
<td>$.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.00$</td>
<td>$1.225 \leq \sigma \leq 1.313$</td>
<td>$1.25$</td>
<td>$.193</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.25$</td>
<td>$1.095 \leq \sigma \leq 1.516$</td>
<td>$1.10$</td>
<td>$.088</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.25$</td>
<td>$.470</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.632</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.516$</td>
<td>$.635</td>
</tr>
<tr>
<td>$.45$</td>
<td></td>
<td></td>
<td>$1.017 \leq \sigma \leq 1.771$</td>
<td>$1.02$</td>
<td>$.073</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.10$</td>
<td>$.371</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.25$</td>
<td>$.562</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.637</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.77$</td>
<td>$.448</td>
</tr>
<tr>
<td>$.50$ ($= \rho_2$)</td>
<td></td>
<td></td>
<td>$1.000 \leq \sigma \leq 1.857$</td>
<td>$1.01$</td>
<td>$.136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.10$</td>
<td>$.406</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.25$</td>
<td>$.578</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.633</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.77$</td>
<td>$.392</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.857$</td>
<td>$.017</td>
</tr>
<tr>
<td>$.60$</td>
<td></td>
<td></td>
<td>$1.118 \leq \sigma \leq 1.798$</td>
<td>$1.20$</td>
<td>$.570</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.517</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.75$</td>
<td>$.304</td>
</tr>
<tr>
<td>$.70$</td>
<td></td>
<td></td>
<td>$1.291 \leq \sigma \leq 1.744$</td>
<td>$1.30$</td>
<td>$.635</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.593</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.70$</td>
<td>$.297</td>
</tr>
<tr>
<td>$.80$</td>
<td></td>
<td></td>
<td>$1.581 \leq \sigma \leq 1.695$</td>
<td>$1.60$</td>
<td>$.424</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.69$</td>
<td>$.107</td>
</tr>
<tr>
<td>$.75$</td>
<td>$.65</td>
<td>$.340 \leq \rho_1 \leq .920</td>
<td>$1.143 \leq \sigma \leq 1.143$</td>
<td>$1.143$</td>
<td>$.014</td>
</tr>
<tr>
<td>$.50$</td>
<td></td>
<td></td>
<td>$1.080 \leq \sigma \leq 1.313$</td>
<td>$1.10$</td>
<td>$.198</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.30$</td>
<td>$.584</td>
</tr>
<tr>
<td>$.75$ ($= \rho_2$)</td>
<td></td>
<td></td>
<td>$1.000 \leq \sigma \leq 1.857$</td>
<td>$1.10$</td>
<td>$.439</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.30$</td>
<td>$.656</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.50$</td>
<td>$.684</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$1.85$</td>
<td>$.128</td>
</tr>
<tr>
<td>$.92$</td>
<td></td>
<td></td>
<td>$1.788 \leq \sigma \leq 1.773$</td>
<td>$1.77$</td>
<td>$.085</td>
</tr>
</tbody>
</table>
3.6.2 Numerical Investigation into the Properties of F.

In view of the fact that \( F(d|\pi) \) is a function of six variables, namely \( d, \delta, \sigma, \rho_1, \rho_2, \) and \( \alpha \), and depends on them in complex fashion, we finally had to abandon our efforts to determine analytically under what conditions \( X_1 - X_2 \) is the OLDF and whether the OLDF can ever be a function other than \( X_1 + X_2 \). Instead, we turned to the study of several examples in hopes of deriving rules-of-thumb for predicting the OLDF.

To illustrate the behavior of \( F(d|\pi) \) for various parameter combinations, seven tables and two figures are presented. Figure 5 shows how \( F(d|\pi) \) changes shape as \( \sigma \) increases, with the other parameters held fixed. Figure 6 shows how \( F(d|\pi) \) changes shape as \( \delta \) increases, with the other parameters held fixed.

Tables 3-8 contain the values of \( F(+1) \) and \( F(-1) \) and give the OLDF, if there is one, for each of 183 parameter sets such that \( \rho_1 \neq \rho_2 \). Table 9 contains the values of \( F(+1) \) and \( F(-1) \) for 135 parameter sets such that \( \rho_1 = \rho_2 \), with the presence of an asterisk to indicate those cases having no OLDF; in all other cases the OLDF is \( X_1 + X_2 \).
### TABLE III-3: The optimal linear discriminant function and the total probabilities of misclassification for the optimal decision rules based on the functions $X_1^2X_2$, for selected parameter sets with $\rho_1^1=\rho_2^1=\rho$, $\sigma_1^2=\sigma_2^2=\sigma$, and $\rho_1^1\neq\rho_2^1$.

<table>
<thead>
<tr>
<th>$\rho_1^{1^0.25}, \rho_2^{1^0.75}, \alpha^{1^0.65}$</th>
<th>( F(-1) = 0.1220 )</th>
<th>( F(-1) = 0.1220 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>( X_1^1X_2 )</td>
<td>( X_1^1X_2 )</td>
</tr>
<tr>
<td>0.5</td>
<td>( F(-1) = 0.3474 )</td>
<td>( F(-1) = 0.3474 )</td>
</tr>
<tr>
<td>( F(+1) = 0.2396 )</td>
<td>( F(+1) = 0.2292 )</td>
<td>( F(+1) = 0.1532 )</td>
</tr>
<tr>
<td>0.9</td>
<td>( F(-1) = 0.35 )</td>
<td>( F(-1) = 0.35 )</td>
</tr>
<tr>
<td>( F(+1) = 0.3346 )</td>
<td>( F(+1) = 0.3099 )</td>
<td>( F(+1) = 0.2266 )</td>
</tr>
<tr>
<td>1.1</td>
<td>( F(-1) = 0.3488 )</td>
<td>( F(-1) = 0.3488 )</td>
</tr>
<tr>
<td>( F(+1) = 0.3499 )</td>
<td>( F(+1) = 0.3350 )</td>
<td>( F(+1) = 0.2563 )</td>
</tr>
<tr>
<td>1.5</td>
<td>( F(-1) = 0.3050 )</td>
<td>( F(-1) = 0.3050 )</td>
</tr>
<tr>
<td>( F(+1) = 0.35 )</td>
<td>( F(+1) = 0.35 )</td>
<td>( F(+1) = 0.3046 )</td>
</tr>
<tr>
<td>1.9</td>
<td>( F(-1) = 0.2651 )</td>
<td>( F(-1) = 0.2651 )</td>
</tr>
<tr>
<td>( F(+1) = 0.35 )</td>
<td>( F(+1) = 0.35 )</td>
<td>( F(+1) = 0.3262 )</td>
</tr>
<tr>
<td>2.2</td>
<td>( F(-1) = 0.2411 )</td>
<td>( F(-1) = 0.2411 )</td>
</tr>
<tr>
<td>( F(+1) = 0.3499 )</td>
<td>( F(+1) = 0.3477 )</td>
<td>( F(+1) = 0.3222 )</td>
</tr>
<tr>
<td>2.6</td>
<td>( F(-1) = 0.2151 )</td>
<td>( F(-1) = 0.2151 )</td>
</tr>
<tr>
<td>( F(+1) = 0.3317 )</td>
<td>( F(+1) = 0.3281 )</td>
<td>( F(+1) = 0.3082 )</td>
</tr>
<tr>
<td>3.0</td>
<td>( F(-1) = 0.1943 )</td>
<td>( F(-1) = 0.1943 )</td>
</tr>
<tr>
<td>( F(+1) = 0.3090 )</td>
<td>( F(+1) = 0.3062 )</td>
<td>( F(+1) = 0.2914 )</td>
</tr>
<tr>
<td>4.0</td>
<td>( F(-1) = 0.1570 )</td>
<td>( F(-1) = 0.1570 )</td>
</tr>
<tr>
<td>( F(+1) = 0.2592 )</td>
<td>( F(+1) = 0.2592 )</td>
<td>( F(+1) = 0.2516 )</td>
</tr>
<tr>
<td>5.0</td>
<td>( F(-1) = 0.1321 )</td>
<td>( F(-1) = 0.1321 )</td>
</tr>
<tr>
<td>( F(+1) = 0.2240 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1)^1 = 0.35$ is $0.5774^\alpha^1 \leq 0.722$. A necessary condition for $F(+1)^1 = 0.35$ is $2.1974^\alpha^1 \leq 4.0808$. 
TABLE III-4: The optimal linear discriminant function and the total probabilities of misclassification for the optimal decision rules based on the functions $X_1 \pm X_2$, for selected parameter sets with $\delta_1 = \delta_2 = \delta$, $\sigma_1 = \sigma_2 = \sigma$, and $\rho_1 \neq \rho_2$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.0001</th>
<th>Values of $\delta$</th>
<th>0.4</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1$</td>
<td>$F(-1)=0.0889$</td>
<td>$F(-1)=0.0889$</td>
<td>$F(-1)=0.0889$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.0478$</td>
<td>$F(+1)=0.0396$</td>
<td>$F(+1)=0.0145$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$0.5$</td>
<td>$F(-1)=0.2960$</td>
<td>$F(-1)=0.2960$</td>
<td>$F(-1)=0.2960$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.1823$</td>
<td>$F(+1)=0.1531$</td>
<td>$F(+1)=0.0633$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$0.9$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2740$</td>
<td>$F(+1)=0.2305$</td>
<td>$F(+1)=0.1081$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$1.3$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3329$</td>
<td>$F(+1)=0.2815$</td>
<td>$F(+1)=0.1503$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$1.7$</td>
<td>$F(-1)=0.3435$</td>
<td>$F(-1)=0.3435$</td>
<td>$F(-1)=0.3435$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.34^\prime$</td>
<td>$F(+1)=0.3195$</td>
<td>$F(+1)=0.1883$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$2.1$</td>
<td>$F(-1)=0.3124$</td>
<td>$F(-1)=0.3124$</td>
<td>$F(-1)=0.3124$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.3492$</td>
<td>$F(+1)=0.2211$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$2.6$</td>
<td>$F(-1)=0.2764$</td>
<td>$F(-1)=0.2764$</td>
<td>$F(-1)=0.2764$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.2540$</td>
<td>$F(+1)=0.2540$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$3.2$</td>
<td>$F(-1)=0.2421$</td>
<td>$F(-1)=0.2421$</td>
<td>$F(-1)=0.2421$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.2751$</td>
<td>$F(+1)=0.2751$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$3.6$</td>
<td>$F(-1)=0.2236$</td>
<td>$F(-1)=0.2236$</td>
<td>$F(-1)=0.2236$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3395$</td>
<td>$F(+1)=0.3304$</td>
<td>$F(+1)=0.2772$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td>$4.0$</td>
<td>$F(-1)=0.2077$</td>
<td>$F(-1)=0.2077$</td>
<td>$F(-1)=0.2077$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3241$</td>
<td>$F(+1)=0.3163$</td>
<td>$F(+1)=0.2741$</td>
<td>$X_1^1+X_2^2$</td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1) = .35$ is $0.8452 \leq \sigma \leq 1.5695$. A necessary condition for $F(+1) = .35$ is $1.7321 \leq \sigma \leq 3.2166$. 

TABLE III-5: The optimal linear discriminant function and the total
probabilities of misclassification for the optimal decision rules
based on the functions $X_1+X_2$, for selected parameter sets with
$\delta_1=\delta_2=\delta$, $\sigma_1=\sigma_2=\sigma$, and $\rho_1 \neq \rho_2$.

<table>
<thead>
<tr>
<th>$\rho_1=-.25$, $\rho_2=+.25$, $\alpha=.65$</th>
<th>Values of $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.1</td>
<td>$F(-1)=0.0956$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.0618$</td>
</tr>
<tr>
<td>0.5</td>
<td>$F(-1)=0.3103$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2257$</td>
</tr>
<tr>
<td>0.9</td>
<td>$F(-1)=0.35$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3224$</td>
</tr>
<tr>
<td>1.3</td>
<td>$F(-1)=0.35$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
</tr>
<tr>
<td>1.7</td>
<td>$F(-1)=0.3319$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
</tr>
<tr>
<td>2.3</td>
<td>$F(-1)=0.2824$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
</tr>
<tr>
<td>2.7</td>
<td>$F(-1)=0.2556$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3396$</td>
</tr>
<tr>
<td>3.1</td>
<td>$F(-1)=0.2333$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3179$</td>
</tr>
<tr>
<td>3.5</td>
<td>$F(-1)=0.2146$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2978$</td>
</tr>
<tr>
<td>4.0</td>
<td>$F(-1)=0.1952$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2752$</td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1)=.35$ is $0.7746 \leq \sigma \leq 1.4385$.
A necessary condition for $F(+1)=.35$ is $1.2910 \leq \sigma \leq 2.3975$. 
TABLE III-6: The optimal linear discriminant function and the total probabilities of misclassification for the optimal decision rules based on the functions $X_1 \pm X_2$, for selected parameter sets with $\delta_1 = \delta_2 = \delta$, $\sigma_1 = \sigma_2 = \sigma$, and $\rho_1 \neq \rho_2$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.0001</th>
<th>$\delta$</th>
<th>0.4</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$F(-1)=0.0478$</td>
<td>$F(-1)=0.0478$</td>
<td>$F(-1)=0.0478$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.0889$</td>
<td>$F(+1)=0.0796$</td>
<td>$F(+1)=0.0445$</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>$F(-1)=0.1823$</td>
<td>$F(-1)=0.1823$</td>
<td>$F(-1)=0.1823$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2960$</td>
<td>$F(+1)=0.2660$</td>
<td>$F(+1)=0.1667$</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>$F(-1)=0.2740$</td>
<td>$F(-1)=0.2740$</td>
<td>$F(-1)=0.2740$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.3452$</td>
<td>$F(+1)=0.2500$</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>$F(-1)=0.3329$</td>
<td>$F(-1)=0.3329$</td>
<td>$F(-1)=0.3329$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.35$</td>
<td>$F(+1)=0.3029$</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3435$</td>
<td>$F(+1)=0.3382$</td>
<td>$F(+1)=0.3047$</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.3124$</td>
<td>$F(+1)=0.3083$</td>
<td>$F(+1)=0.2866$</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2736$</td>
<td>$F(+1)=0.2739$</td>
<td>$F(+1)=0.2607$</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$F(-1)=0.35$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2421$</td>
<td>$F(+1)=0.2406$</td>
<td>$F(+1)=0.2326$</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>$F(-1)=0.3395$</td>
<td>$F(-1)=0.3395$</td>
<td>$F(-1)=0.3395$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2236$</td>
<td>$F(+1)=0.2224$</td>
<td>$F(+1)=0.2165$</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>$F(-1)=0.3241$</td>
<td>$F(-1)=0.3241$</td>
<td>$F(-1)=0.3241$</td>
<td>$X_1 + X_2$</td>
</tr>
<tr>
<td></td>
<td>$F(+1)=0.2077$</td>
<td>$F(+1)=0.2069$</td>
<td>$F(+1)=0.2023$</td>
<td></td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1) = .35$ is $1.732 \leq \sigma \leq 3.2166$. A necessary condition for $F(+1) = .35$ is $0.8452 \leq \sigma \leq 1.5695$. 
TABLE III-7: The optimal linear discriminant function and the total probabilities of misclassification for the optimal decision rules based on the functions $X_1 \pm X_2$, for selected parameter sets with $\delta_1=\delta_2=\delta$, $\sigma_1=\sigma_2=\sigma$, and $\rho_1 \neq \rho_2$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$F(-1)=0.0666$</th>
<th>$F(-1)=0.0666$</th>
<th>$F(-1)=0.0666$</th>
<th>$X_1-X_2$</th>
<th>$X_1-X_2$</th>
<th>$X_1+X_2$</th>
<th>$X_1+X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$F(0.1220)$</td>
<td>$F(0.0711)$</td>
<td>$F(0.0044)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>0.5</td>
<td>$F(-0.2396)$</td>
<td>$F(-0.2396)$</td>
<td>$F(-0.2396)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>0.9</td>
<td>$F(-0.3346)$</td>
<td>$F(-0.3346)$</td>
<td>$F(-0.3346)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>1.3</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>1.7</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>2.1</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>2.5</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>3.2</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>4.0</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>4.5</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
<tr>
<td>5.0</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$F(-0.35)$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
<td>$X_1+X_2$</td>
</tr>
</tbody>
</table>

Values of $\delta$:

<table>
<thead>
<tr>
<th>Values of $\delta$</th>
<th>0.4</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F(-1)=0.0666$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(0.0711)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(0.0044)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.2396)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.3346)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
<tr>
<td>$F(-0.35)$</td>
<td>$X_1-X_2$</td>
<td>$X_1-X_2$</td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1)=.35$ is $2.1974 < \sigma < 4.0808$. A necessary condition for $F(+1)=.35$ is $0.5774 < \sigma < 1.0722$. 
TABLE III-8: The optimal linear discriminant function and the total probabilities of misclassification for the optimal decision rules based on the functions $X_1 \pm X_2$, for selected parameter sets with $\delta_1 = \delta_2 = \delta$, $\sigma_1 = \sigma_2 = \sigma$, and $\rho_1 \neq \rho_2$.

<table>
<thead>
<tr>
<th>$\rho_1 = +.25$, $\rho_2 = -.25$, $\alpha = .65$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1.3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1.7</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2.3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2.7</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3.1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3.5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4.0</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

The necessary and sufficient condition for $F(-1) = .35$ is $1.2910 \leq \sigma \leq 2.3975$. A necessary condition for $F(+1) = .35$ is $0.7746 \leq \sigma \leq 1.4385$. 
TABLE III-9: The total probability of misclassification for the optimal decision rule based on each of the functions $X_1+X_2$ for selected parameter sets with $\delta_1=\delta_2=\delta$, $\sigma_1=\sigma_2=\sigma$, $\rho_1=\rho_2=\rho$, and $\alpha=.65$. An asterisk indicates that no OLDF exists for the parameter set. Otherwise the OLDF is $X_1+X_2$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>F(-1)</th>
<th>$\delta$</th>
<th>- .75</th>
<th>- .25</th>
<th>0.00</th>
<th>+ .25</th>
<th>+ .75</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>.0770</td>
<td>0.0001</td>
<td>.0770</td>
<td>.0770</td>
<td>.0770</td>
<td>.0770</td>
<td>.0770</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.0441</td>
<td>.0640</td>
<td>.0671</td>
<td>.0689</td>
<td>.0712</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.0023</td>
<td>.0240</td>
<td>.0322</td>
<td>.0384</td>
<td>.0469</td>
</tr>
<tr>
<td>0.5</td>
<td>.2675</td>
<td>0.0001</td>
<td>.2675</td>
<td>.2675</td>
<td>.2675</td>
<td>.2675</td>
<td>.2675</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.1644</td>
<td>.2251</td>
<td>.2347</td>
<td>.2408</td>
<td>.2481</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.0233</td>
<td>.1044</td>
<td>.1290</td>
<td>.1473</td>
<td>.1725</td>
</tr>
<tr>
<td>0.9</td>
<td>.3492</td>
<td>0.0001</td>
<td>.3492</td>
<td>.3492</td>
<td>.3492</td>
<td>.3492</td>
<td>.3492</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.2426</td>
<td>.3070</td>
<td>.3172</td>
<td>.3235</td>
<td>.3310</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.0628</td>
<td>.1755</td>
<td>.2038</td>
<td>.2241</td>
<td>.2513</td>
</tr>
<tr>
<td>1.3</td>
<td>.35</td>
<td>0.0001</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.2998</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.1071</td>
<td>.2320</td>
<td>.2609</td>
<td>.2812</td>
<td>.3084</td>
</tr>
<tr>
<td>1.7</td>
<td>.35</td>
<td>0.0001</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.3206</td>
<td>.3488</td>
<td>.35</td>
<td>.35</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.1472</td>
<td>.2692</td>
<td>.2930</td>
<td>.3082</td>
<td>.3259</td>
</tr>
<tr>
<td>2.1</td>
<td>.3381</td>
<td>0.0001</td>
<td>.3381</td>
<td>.3381</td>
<td>.3381</td>
<td>.3381</td>
<td>.3381</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.3095</td>
<td>.3292</td>
<td>.3315</td>
<td>.3328</td>
<td>.3344</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.1788</td>
<td>.2771</td>
<td>.2925</td>
<td>.3020</td>
<td>.3128</td>
</tr>
<tr>
<td>2.5</td>
<td>.3114</td>
<td>0.0001</td>
<td>.3114</td>
<td>.3114</td>
<td>.3114</td>
<td>.3114</td>
<td>.3114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.2910</td>
<td>.3046</td>
<td>.3063</td>
<td>.3073</td>
<td>.3085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.1968</td>
<td>.2693</td>
<td>.2797</td>
<td>.2860</td>
<td>.2932</td>
</tr>
<tr>
<td>3.0</td>
<td>.2807</td>
<td>0.0001</td>
<td>.2807</td>
<td>.2807</td>
<td>.2807</td>
<td>.2807</td>
<td>.2807</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.2673</td>
<td>.2762</td>
<td>.2773</td>
<td>.2780</td>
<td>.2787</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.2035</td>
<td>.2531</td>
<td>.2598</td>
<td>.2639</td>
<td>.2687</td>
</tr>
<tr>
<td>4.0</td>
<td>.2334</td>
<td>0.0001</td>
<td>.2334</td>
<td>.2334</td>
<td>.2334</td>
<td>.2334</td>
<td>.2334</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4</td>
<td>.2267</td>
<td>.2311</td>
<td>.2317</td>
<td>.2320</td>
<td>.2324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>.1940</td>
<td>.2196</td>
<td>.2230</td>
<td>.2251</td>
<td>.2274</td>
</tr>
</tbody>
</table>
3.6.3 Rules for Predicting the OLDF when $\delta_1=\delta_2$ and $\sigma_1=\sigma_2$.

The rules-of-thumb presented here are based on Lemmas 10-14 and on the assumption that the OLDF must be either $X_1+X_2$ or $X_1-X_2$.

1.) There is no OLDF (i.e. the best decision rule is to ignore $X$ and to choose Population 1) iff the following three conditions hold:
   a.) $\frac{\rho_2-K}{1-\rho_2 K} \leq \rho_1 \leq \frac{\rho_2+K}{1+\rho_2 K}$, where $K = \frac{\alpha^2-(1-\alpha)^2}{\alpha^2+(1-\alpha)^2}$;
   b.) $\max\left\{\left[\frac{1-\rho_2}{1-\rho_1}\right]^{1/2}, \left[\frac{1+\rho_2}{1+\rho_1}\right]^{1/2}\right\} \leq \sigma \leq \min\left\{\frac{\alpha}{1-\alpha}, \frac{1-\rho_2}{1-\rho_1}, \frac{\alpha}{1-\alpha}, \frac{1+\rho_2}{1+\rho_1}\right\}$;
   c.) $0 \leq \delta \leq \delta(\pi) = \left\{\frac{\alpha^2(1+\rho_1)-(1+\rho_2)}{2 \ln \left\{\frac{\alpha}{1-\alpha}, \frac{1-\rho_2}{1+\rho_1}, \sigma^2\right\}\right\}^{1/2}$.

2.) The OLDF is $X_1-X_2$ if one of the following sets of conditions holds:
   a.) $\rho_1 < \rho_2$ and
      i.) $\max\left\{\left[\frac{1+\rho_2}{1+\rho_1}\right]^{1/2}, \left[\frac{\alpha}{1-\alpha}, \frac{1-\rho_2}{1+\rho_1}\right]\right\} < \sigma \leq \frac{\alpha}{1-\alpha}$;
      ii.) $\delta \leq \delta(\pi)$ or "not too much bigger" than $\delta(\pi)$.
   b.) $\rho_1 > \rho_2$ and
      i.) $0 < \sigma < \min\left\{\left[\frac{1-\rho_2}{1-\rho_1}\right]^{1/2}, \left[\frac{\alpha}{1-\alpha}, \frac{1+\rho_2}{1+\rho_1}\right]\right\}$;
      ii.) $0 < \delta \leq \delta(\pi)$ or "not too much bigger" than $\delta(\pi)$.

3.) The OLDF could be either $X_1+X_2$ or $X_1-X_2$ and so must be evaluated if the following two conditions hold:
   a.) $-1 < \rho_1 < \frac{\rho_2-K}{1-\rho_2 K}$ or $\frac{\rho_2+K}{1+\rho_2 K} < \rho_1 < 1$; and
   b.) $\min\left\{\left[\frac{\alpha}{1-\alpha}, \frac{1-\rho_2}{1-\rho_1}\right]^{1/2}, \left[\frac{\alpha}{1-\alpha}, \frac{1+\rho_2}{1+\rho_1}\right]^{1/2}\right\} < \sigma < \max\left\{\left[\frac{1-\rho_2}{1-\rho_1}\right]^{1/2}, \left[\frac{1+\rho_2}{1+\rho_1}\right]^{1/2}\right\}$.

4.) In all other cases, especially when $\delta$ is "large", the OLDF is $X_1+X_2$. 
CHAPTER IV

LINEAR DISCRIMINATION BETWEEN TWO BIVARIATE NORMAL POPULATIONS WITH UNEQUAL COVARIANCE MATRICES BASED ON A TRUNCATED SAMPLING PROCEDURE

In this chapter the population of interest is a mixture of two bivariate normal populations having unequal covariance matrices, where all parameters are known. In addition, the proportion of each population in the mixture is known. Finally, because of limitations in the measuring apparatus, an element \( \mathbf{x} \) in the mixture cannot be detected if \( x_1 > k \) (or, alternatively, \( x_1 < k \)), where \( k \) is known.

4.0 Basic Assumptions and Definitions.

The Basic Assumptions stated below are the five assumptions stated in Section 3.1 plus two assumptions dealing with truncation. The terms, optimal decision rule based on a specified linear function and optimal linear discriminant function, are defined exactly as in Section 3.1.

Recall the fact that, in order to retain truncation of the rectangular type, we cannot, in general, use a transformation which reduces a covariance matrix to the identity matrix.

BASIC ASSUMPTIONS:

1.) In Population 1, \( \mathbf{X} \sim N(\delta, \Sigma) \), where
   a.) \( \delta' = (\delta_1, \delta_2) \) with \( \delta_i \geq 0, \ i=1,2; \)
   b.) \( \Sigma = \begin{pmatrix} \sigma_1^2 & \rho_1 \sigma_1 \sigma_2 \\ \rho_1 \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \) is positive definite.

2.) In Population 2, \( \mathbf{X} \sim N(0, \mathbf{R}) \), where \( \mathbf{R} = \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix} \) is positive definite.

3.) \( \Sigma \neq \mathbf{R} \).

4.) The proportion of elements from Population 1 is \( \alpha \geq 1/2 \).

5.) Elements \( \mathbf{x}' = (x_1, x_2) \) such that \( x_1 > k \) cannot be detected.

6.) \( k > 0 \).

7.) All parameters are known.
We wish to find the coefficient vector $\mathbf{a}' = (a_1, a_2)$ such that the optimal decision rule based on the linear function $Y = \mathbf{a}'X$ has the smallest total probability of misclassification among decision rules based on linear functions of $X$.

### 4.1 The Optimal Decision Rule Based on a Specified Linear Function.

For a specified linear function $Y = d_1X_1 + d_2X_2 \neq 0$, the optimal decision rule based on $Y$ is to choose Population 1 iff

$$af_1(y|k) > (1-a)f_2(y|k),$$

where $f_i(y|k)$ denotes the p.d.f. of $Y$ in the $i^{th}$ population when $X_i$ has been truncated at $k$, $i=1,2$. Now, $Y = \mathbf{c}Z$, where

$$\begin{cases} Z = X_1 \text{ and } c = d_1 \neq 0 & \text{if } d_2 = 0; \\
Z = dX_1 + X_2 \text{ with } d = d_1/d_2 \text{ and } c = d_2 \text{ if } d_2 \neq 0; \\
\end{cases}$$

and so the density function of $Z$ is $g_i(z|k) = |c|f_i(y|k)$, $i=1,2$.

Since $(g_1/g_2) = (f_1/f_2)$, the optimal decision rule based on $Z$ is equivalent to the optimal decision rule based on $Y$. Consequently, it will be necessary to consider only the linear functions $Y = X_1$ and $Y = dX_1 + X_2$, where $d$ is any real number.

**THEOREM IV.1:** Given a specified linear function $Y = X_1$ or $Y = dX_1 + X_2$, where $d$ is any real number, the optimal decision rule based on $Y$ is to choose Population 1 iff

$$A.) \quad \frac{\phi(k) \exp[-(x_1 - \delta_1)^2/(2\sigma_1^2)]}{\phi((k-\delta_1)/\sigma_1 \exp[-x_1^2/2] \sqrt{a}} \quad \frac{1-a}{a} \quad \text{if } y = x_1;$$

$$B.) \quad \frac{\phi(k) \sqrt{\theta_2} \exp[-(y - \mu_2)^2/(2\theta_2)] \phi(a_1 - b_1y)}{\phi((k-\delta_1)/\sigma_1 \sqrt{\theta_1} \exp[-y^2/(2\theta_1)] \phi(a_2 - b_2y)} \quad \frac{1-a}{a} \quad \text{if } y = dx_1 + x_2,$$

where

$$\begin{align*}
\mathbf{a}' &= (a_1, a_2) \\
\nu &= \mathbf{a}' \delta = d\delta_1 + \delta_2 \\
\theta_1 &= \mathbf{a}' \Sigma \mathbf{a} = (d\sigma_1)^2 + 2d\sigma_1\sigma_2d + \sigma_2^2 \\
\theta_2 &= \mathbf{a}' \mathbf{R} \mathbf{a} = d^2 + 2d^2p_2 + 1
\end{align*}$$

where

$$\begin{align*}
a_1 &= \frac{\kappa \sqrt{\theta_1}}{\sqrt{|\Sigma|}} - \frac{\mathbf{a}' \Sigma (\mu_2, -\mu_1)}{\sqrt{|\Sigma| \sqrt{\theta_1}}} \\
b_1 &= \frac{\sigma_1 (d + \rho_1 \sigma_2)}{\sqrt{|\Sigma| \sqrt{\theta_1}}} \\
a_2 &= \frac{\kappa \sqrt{\theta_2}}{\sqrt{|\mathbf{R}|}} \\
b_2 &= \frac{d + \rho_2}{\sqrt{|\mathbf{R}| \sqrt{\theta_2}}}
\end{align*}$$

(1)
proof:

Case A: In Population 1, \( X_1 \) has p.d.f.

\[
f_1(x_1|k) = \begin{cases} 
\frac{\exp[-(x_1-\delta_1)^2/(2\sigma_1^2)]}{\sigma_1\sqrt{2\pi}} \phi\left(\frac{x_1-\delta_1}{\sigma_1}\right) & \text{if } x_1 \leq k \\
0 & \text{otherwise}
\end{cases}
\]

and in Population 2 \( X_1 \) has p.d.f.

\[
f_2(x_1|k) = \begin{cases} 
\frac{\exp[-x_1^2/2]}{\sqrt{2\pi}} \phi(k) & \text{if } x_1 \leq k \\
0 & \text{otherwise}
\end{cases}
\]

Inequality (A) follows directly from the definition of the optimal decision rule given above.

Case B: \( Y = dX_1 + X_2 \).

Define \( U = X_1 - dX_2 \equiv c'X \).

Define \( W = \begin{bmatrix} y \\ u \end{bmatrix} = WX \), i.e. \( W = \begin{bmatrix} d & 1 \\ 1 & -d \end{bmatrix} = \begin{bmatrix} d' & c' \end{bmatrix} \).

Then \( X = W^{-1}W = (d'c')^{-1}Ww \).

So \( x_1 \leq k \) iff \( d'w \leq k(d'd) \); i.e. \( u \leq k(d'd) - dy \).

Now, in Population 2 \( x \sim N(0,R) \) truncated to \( \{x: x_1 \leq k\} \).

Thus, in Population 2 \( w \sim N(0,V) \) truncated to \( \{w: u \leq k(d'd) - dy\} \),

where \( V = WW' = \begin{bmatrix} \bar{d}'R_d & \bar{d}'R_c \\ \bar{d}'R_c & c'R_c \end{bmatrix} = \begin{bmatrix} \theta_2 & r \\ r & v \end{bmatrix} \). \( \tag{2} \)

So \( \phi(k) g_2(w|k) = \begin{cases} 
\frac{\exp\left[-\frac{y^2}{2\theta_2}\right]}{\sqrt{2\pi\theta_2}} \frac{\exp\left[-\frac{[u-y(r/\theta_2)]^2}{2(v-r^2/\theta_2)}\right]}{\sqrt{2\pi(v-r^2/\theta_2)}} & \text{if } u \leq k(d'd) - dy; \\
0 & \text{otherwise}
\end{cases} \)

Hence \( \phi(k) f_2(y|k) = \frac{\exp\left[-\frac{y^2}{2\theta_2}\right]}{\sqrt{2\pi\theta_2}} \phi\left(\frac{k(d'd) - y(d + r/\theta_2)}{\sqrt{v - r^2/\theta_2}}\right) \).

Now, (2) implies \( |\mathbf{V}| = |\mathbf{W}^2|R = (d'd)^2|R| \) and \( |\mathbf{V}| = \mathbf{v}\theta_2 - r^2 \).

Therefore \( \mathbf{v}\theta_2 - r^2 = (d'd)^2|R| \).

Also, \( d\theta_2 + r = (d'd)(d + \rho_2) \).
\[ \phi(k) f_2(y|k) = \frac{\exp\left(-\frac{y^2}{2\theta_2}\right)}{\sqrt{2\pi\theta_2}} \phi\left(\frac{k\sqrt{\theta_2}}{\sqrt{\Xi}} - y\frac{d+\rho_2}{\sqrt{\theta_2}\sqrt{\Xi}}\right). \]  

(3)

In Population 1 \( \mathbf{x} \sim N(\delta, \Sigma) \) truncated to \( \{x; x_1 \leq k\} \), and so \( w \sim N(\mathbf{w}\delta, \mathbf{V}) \) truncated to \( \{w; u\mathbf{k}(\mathbf{a}'\mathbf{d}) - \mathbf{d}y\} \),

where

\[
\begin{align*}
\mathbf{V} = \mathbf{WW}' &= \begin{pmatrix} \mathbf{d}'\mathbf{d} & \mathbf{d}'\xi \xi' \mathbf{c} \\ \mathbf{d}'\xi \xi' \mathbf{c} & \mathbf{c}'\xi \xi' \mathbf{c} \end{pmatrix} = \begin{pmatrix} \theta_1 & r \\ r & v \end{pmatrix} \\
\mathbf{W} = (\mathbf{a}'\delta, \mathbf{c}'\delta) &\equiv (\nu, \delta)
\end{align*}
\]  

(3.5)

So

\[ \phi\left(\frac{k-\delta}{\sigma_1}\right) f_1(y|k) = \frac{\exp\left(-\frac{(y-v)^2}{2\theta_1}\right)}{\sqrt{2\pi\theta_1}} \phi\left(\frac{k(\mathbf{a}'\mathbf{d}) - \mathbf{d}y - \mathbf{c}'\delta - r(y-v)}/\theta_1\right). \]  

(a)

It follows from (3.5) and the definition of \( \mathbf{W} \) that

\[ v\theta_1 - r^2 = (\mathbf{d}'\mathbf{d})^2/\Sigma. \]  

(b)

Also,

\[ \mathbf{d}\theta_1 + r = (\mathbf{d}'\mathbf{d})\mathbf{d}'\Sigma u_1; \]

(c)

and

\[ \mathbf{c}'\delta - v\theta_1 = (\theta_1\mathbf{c}' - \mathbf{r}\mathbf{d}')\delta/\theta_1, \]

(d)

where

\[ (\theta_1\mathbf{c}' - \mathbf{r}\mathbf{d})' = (\theta_1 - \mathbf{r}\mathbf{d}, -[\mathbf{d}\theta_1 + r]) \]

(e)

and

\[ \theta_1 - \mathbf{r}\mathbf{d} = (\mathbf{d}'\mathbf{d})\mathbf{d}'\Sigma u_2. \]

(f)

Substituting (b)-(f) into (a), we get

\[ \phi\left(\frac{k-\delta}{\sigma_1}\right) f_1(y|k) = \frac{\exp\left(-\frac{(y-v)^2}{2\theta_1}\right)}{\sqrt{2\pi\theta_1}} \phi\left(\frac{k\sqrt{\theta_1}}{\sqrt{\Sigma}} - \frac{\mathbf{d}'\Sigma u_2 - \mathbf{u}_1\delta}{{\sqrt{\theta_1}}/\sqrt{\Sigma}} - y\frac{\sigma_1(\sigma_1 d + \rho_1 \sigma_2)}{\sqrt{\theta_1}\sqrt{\Sigma}}\right). \]  

(4)

Then the optimal decision rule based on \( Y \) is to choose Population 1 iff

\[ \frac{f_1(y|k)}{f_2(y|k)} > \frac{1-\alpha}{\alpha}, \]

where the p.d.f.'s are given by (3) and (4), with definitions (1) substituted to simplify the arguments.

\[ \square \]

Inequalities (A) and (B) in the above theorem can be put in more enlightening form by taking natural logarithms.

**Corollary IV.1:** The optimal decision rule based on the linear function \( Y = dX_1 + X_2 \) can be written: Choose Population 1 iff \( U(Y) < 0 \), where
\[
\begin{align*}
\begin{cases}
U(y) = Q(y) + G(y), \\
Q(y) = D y^2 - 2A y + B
\end{cases}
\end{align*}
\]

with
\[
D = \theta_2 - \theta_1
\]
\[
A = v \theta_2
\]
\[
B = \theta_2 \left( v^2 - 2 \theta_1 \ln \left( \frac{a}{1-a} \left( \frac{\theta_2}{\theta_1} \right)^{1/2} \frac{\phi(k)}{\phi(k-\delta_1)} \right) \right)
\]

\[
G(y) = -2 \theta_1 \theta_2 g(y),
\]

with
\[
g(y) = \ln \frac{\phi(a_1 - b_1 y)}{\phi(a_2 - b_2 y)}
\]

and \(v, \theta_i, a_i, b_i, i=1,2\), are defined in (1).

**proof**: Take the natural logarithm of both sides of inequality (B) of the theorem.

The properties of the functions defined in (5) and the conditions under which \(U(y)\) is negative are studied in Sections 4.3-4.5.

Observe now that when \(Y=X_1\), inequality (A) of Theorem 1 holds iff

\[
Q(x_1) \equiv (\sigma_1^2 - 1)x_1^2 + 2 \delta_1 x_1 - \left[ \delta_1^2 - 2 \sigma_1^2 \ln \left( \frac{a}{1-a} \left( \frac{\phi(k)}{\phi(k-\theta_1)} \right) \right) \right] > 0. \tag{6}
\]

**THEOREM IV.2**: The optimal decision rule based on the linear function \(Y=X_1\) can be written as follows:

A.) If \(\sigma_1=1\) and

a.) \(\delta_1=0\), then \(X_1\) cannot distinguish between the two populations.

b.) \(\delta_1>0\), then choose Population 1 iff

\[
x_1 > \frac{\delta_1}{2} - \delta_1 \ln \left( \frac{a}{1-a} \frac{\phi(k)}{\phi(k-\theta_1)} \right),
\]

which critical value is less than \(k\) for all positive \(\delta_1\).

B.) If \(\sigma_1<1\), choose Population 1 iff

\[
\frac{\delta_1 - \sqrt{h}}{1-\sigma_1^2} < x_1 < \frac{\delta_1 + \sqrt{h}}{1-\sigma_1^2},
\]

where

\[
h \equiv \delta_1^2 - 2(1-\sigma_1^2) \ln \left( \frac{\sigma_1 \phi([k-\theta_1]/\sigma_1)}{(a/1-a) \phi(k)} \right) > 0. \tag{7}
\]
C.) If \( \sigma_1 > 1 \) and

a.) \( h < 0 \), always choose Population 1;

b.) \( h \geq 0 \), choose Population 1 iff

\[
\frac{\delta_1 + \sigma_1 \sqrt{h}}{1 - \sigma_1^2} \quad \text{or} \quad \frac{\delta_1 - \sigma_1 \sqrt{h}}{1 - \sigma_1^2}
\]

**proof:**

**Case A:** \( \sigma_1 = 1 \).

Then \( Q(x_1) \) is a linear function of \( x_1 \), and the results follow directly from (6).

Consider the "boundary" value

\[
B(\delta | k, a) = \left( \frac{\delta}{2} \right) - \delta^{-1} \ln \left( \frac{a}{1-a} \frac{\phi(k)}{\phi(k-\delta)} \right)
\]

If it happened that \( B(\delta | k, a) \) exceeded \( k \) for some \( \delta > 0 \), then the optimal decision rule based on \( X_1 \) would be, in effect, "Always choose Population 2", which decision rule is *not* optimal unless \( a = 1/2 \), since its misclassification probability exceeds the misclassification probability for the rule, "Always choose Population 1". Therefore it is useful, as a check, to show that \( B(\delta | k, a) < k \) for all positive \( \delta \) and for all \( k \geq 0, a \geq 1/2 \).

Because \( \xi(t) = \phi(t) / \phi(t) \) for all real \( t \), we can write

\[
B(\delta | k, a) = \left( \frac{\delta}{2} \right) - \delta^{-1} \ln \left( \frac{a}{1-a} \frac{\phi(k)}{\phi(k-\delta)} \right)
\]

\[
= \left( \frac{\delta}{2} \right) - \delta^{-1} \ln \left( \frac{a}{1-a} \frac{\xi(k-\delta)}{\xi(k)} \exp\left[-(k^2 - [k-\delta]^2)/2\right] \right)
\]

\[
= k - \delta^{-1} \ln \left( \frac{a}{1-a} \frac{\xi(k-\delta)}{\xi(k)} \right)
\]

Since \( \xi(t) \) increases as \( t \) decreases [Lemma II.8],

\( \xi(k-\delta) > \xi(k) \) for all \( \delta > 0 \) and \( k \geq 0 \).

Therefore

\[
\ln \left( \frac{a}{1-a} \frac{\xi(k-\delta)}{\xi(k)} \right) > \ln \left( \frac{a}{1-a} \right) \geq 0 \text{ for all } \delta > 0, k \geq 0, a \geq 1/2.
\]

Thus \( B(\delta | k, a) < k \) for all \( \delta > 0, k \geq 0, a \geq 1/2 \).

**Case B:** \( \sigma_1 < 1 \).

Now \( Q(x_1) \), defined in (6), is an arch-shaped quadratic with zeros

\[
x_1 = \frac{\delta_1 \pm \sigma_1 \sqrt{h}}{1 - \sigma_1^2}
\]
with \( h \) defined in (7).

When \( \sigma_1 < 1 \), \( h \) is strictly positive. To see this, define

\[
G(k, \delta, \sigma) \equiv \sigma \Phi \left( \frac{k-\delta}{\sigma} \right) - \Phi(k).
\]

(9)

For any given \( \delta > 0 \) and \( \sigma < 1 \), \( G(k, \delta, \sigma) \) attains two maxima in the domain of \( k \), i.e. \( k \geq 0 \); and when \( \delta = 0 \), \( G(k, 0, \sigma) \) is maximized when \( k = 0 \). For if we let

\[
G'(k|\delta, \sigma) = \frac{\delta}{\sigma} G(k, \delta, \sigma) = \sigma \Phi \left( \frac{k-\delta}{\sigma} \right) (1/\sigma) - \Phi(k) = \Phi \left( \frac{k-\delta}{\sigma} \right) - \Phi(k),
\]

then \( G'(k|\delta, \sigma) = 0 \) iff \( |(k-\delta)/\sigma| = k \), i.e. iff \( k = \delta/(1-\sigma) \).

When \( \sigma < 1 \), then \( \delta/(1-\sigma) \) is positive and so lies in the domain of \( k \).

Consider the behavior of the derivative \( G' \):

a.) \( 0 \leq k < \delta/(1+\sigma) \) implies \( (k-\delta)/\sigma < -k \leq 0 \) implies \( G'(k|\delta, \sigma) < 0 \).

b.) \( \delta/(1+\sigma) \leq k \leq \delta \) implies \( -k < (k-\delta)/\sigma \leq 0 \) implies \( G'(k|\delta, \sigma) > 0 \).

c.) \( \delta < k \leq 1/(1-\sigma) \) implies \( 0 < (k-\delta)/\sigma < k \) implies \( G'(k|\delta, \sigma) > 0 \).

d.) \( \delta/(1-\sigma) \leq k \) implies \( 0 < k < (k-\delta)/\sigma \) implies \( G'(k|\delta, \sigma) < 0 \).

Therefore, for all \( \delta \geq 0 \), \( 0 < \sigma < 1 \), and

for \( k \geq \delta/(1+\sigma) \), \( \max G(k, \delta, \sigma) = G(\delta/[1-\sigma], \delta, \sigma) = (\sigma-1) \Phi(\delta/[1-\sigma]) < 0 \);

for \( 0 \leq k < \delta/(1+\sigma) \), \( \max G(k, \delta, \sigma) = G(0, \delta, \sigma) = \sigma \Phi(-\delta/\sigma)-1/2 \leq (\sigma-1)/2 < 0 \).

Thus \( G(k, \delta, \sigma) \) is negative for all \( \delta \geq 0 \), \( \sigma < 1 \), and \( k \geq 0 \).

That is, \( \sigma \Phi([k-\delta]/\sigma) < \Phi(k) \), and so for all \( \alpha \geq 1/2 \),

\[
\sigma \Phi \left( \frac{k-\delta}{\sigma} \right) < \frac{\alpha}{1-\alpha} \Phi(k) \text{ for all } k \geq 0, \delta \geq 0, 0 < \sigma < 1.
\]

Applying this inequality to (7), we see that

\( h > 0 \) for all \( k \geq 0, \delta \geq 0, 0 < \sigma_1 < 1 \).

Therefore the zeros of \( Q(x_1) \), given in (8), are real and distinct, and so \( Q(x_1) > 0 \) in the open interval between them.
As in Case A, it is useful, as a check on the decision rule, to show that the left endpoint of the "choose Population 1" interval does not exceed \( k \).

Let
\[
L(\delta |k, \sigma, a) = \frac{\delta - \sigma \sqrt{\frac{h}{(1-\sigma^2)}}}{\sigma^2}
\]
where \( h \) is defined in (7) and subscripts are omitted for simplicity.

Since, as shown above, \( h > 0 \) for all \( k \geq 0, \delta \geq 0, 0 < \sigma < 1 \), it follows that
\[
L < \frac{\delta}{\sigma^2} \leq k \text{ for all } 0 \leq \delta \leq k(1-\sigma^2), \ k \geq 0.
\]  
(a)

For \( \delta > k(1-\sigma^2) \), let us suppose \( L(\delta |k, \sigma, a) \geq k \) for some choice of parameters.

Then
\[
[\delta - k(1-\sigma^2)]^2 \geq \sigma^2 \left( \delta^2 + 2(1-\sigma^2) \left( \ln \left( \frac{\sigma^2}{(1-\sigma^2)} \right) \frac{\phi(k)}{\sigma} \right) \right);
\]
so
\[
\delta^2 - 2\sigma^2 + k^2(1-\sigma^2) \geq 2\sigma^2 \left( \ln \left( \frac{\sigma^2}{(1-\sigma^2)} \right) \frac{\phi(k)}{\sigma} \right);
\]

\[
\ln \exp \left( \frac{1}{2} \left( \frac{(k-\delta)}{\sigma} + \frac{\phi(k)}{\sigma} \right) \right) \geq \frac{\alpha}{1-\alpha} \frac{\phi(k)}{\sigma} \quad \text{or}
\]

\[
1 \geq \frac{\alpha}{1-\alpha} \frac{\phi(k)}{\sigma} \frac{\phi \left( \frac{k-\delta}{\sigma} \right)}{\phi \left( \frac{\phi(k)}{\sigma} \right)} = \frac{\alpha}{1-\alpha} \frac{\xi \left( \frac{k-\delta}{\sigma} \right)}{\phi \left( \frac{\phi(k)}{\sigma} \right)} \xi(k) \quad \text{or}
\]

(b)

Now, \( \alpha / 2 \geq 1 / 2 \) implies \( \alpha / (1 - \alpha) > 1 / \sigma \).

Also, \( \xi(t) \) is a decreasing function of \( t \), and \( (k-\delta)/\sigma \leq k \)
for all \( \delta \geq k(1-\sigma) \), and so
\[
\frac{\xi \left( \frac{k-\delta}{\sigma} \right)}{\xi(k)} \geq 1 \text{ for all } \delta \geq k(1-\sigma).
\]

Since \( k(1-\sigma) < k(1-\sigma^2) \) for all \( \sigma < 1 \), the R.H.S. of (b) is
strictly greater than unity for all \( \delta > k(1-\sigma^2) \). Hence a contradic-
tion.

Therefore, \( L < k \) for all \( \delta > k(1-\sigma^2) \); and by (a) we have \( L < k \) for
all \( \delta \geq 0, k \geq 0, \alpha / 2, \text{ and } 0 < \sigma < 1 \).

Case C: \( \sigma > 1 \).

\( Q(x_1) \) is a U-shaped quadratic having as zeros the values defined in
(b). Thus, if \( h < 0 \), then \( Q(x_1) > 0 \) for all real \( x_1 \), and the optimal
decision rule is to ignore the sample and simply choose Population 1; whereas if \( h \geq 0 \), then \( q(x_1) > 0 \) for all \( x_1 \) not in the closed interval between the points defined in (8), and so the optimal decision rule selects Population 1 for all points outside this closed interval.

\[
\]

The conditions under which the function \( h \) is negative are studied in the next section.

4.2 When is it Optimal to Ignore the Data Based on \( X=X_1 \)?

In this section it is shown that for fixed values of \( \sigma > 1, \delta \geq 0, \) and \( k \geq 0 \), the function \( h \) defined in (7) is negative for sufficiently large \( \alpha \), and consequently the optimal decision rule based on \( X_1 \) is the trivial one, to choose Population 1 regardless of the data. Upper bounds on \( \alpha \) for the optimality of nontrivial decision rules are derived as functions of the parameters \( k \) and \( \sigma \) and of \( k \) alone.

**Lemma IV.1:** When \( \sigma > 1 \), \( h(k, \delta, \sigma, \alpha) \) is a strictly decreasing function of \( \alpha \). Consequently, given \( k \geq 0 \), \( \delta \geq 0 \), and \( \sigma > 1 \), the function \( h \) is negative for all \( \alpha > \alpha_0 \), where

\[
\alpha_0 \equiv (1 + 1/B)^{-1}\]

\[
B \equiv \frac{\sigma \phi \left( \frac{k-\delta}{\sigma} \right)}{\phi(k)} \exp \left( \frac{\delta^2}{2(\sigma^2-1)} \right)
\]

**Proof:** Differentiating (7), we get

\[
h'(\alpha | k, \delta, \sigma) = \frac{\partial}{\partial \alpha} h(k, \delta, \sigma, \alpha) = -\frac{2(\sigma^2-1)}{\alpha(1-\alpha)} < 0 \text{ for all } \sigma > 1.
\]

Now, \( h = 0 \) iff \( \delta^2 = 2(1-\sigma^2) \ln \left( \frac{\sigma \phi \left( \frac{k-\delta}{\sigma} \right)}{\phi(k)} \right) \) [from (7)]

iff \( \frac{\alpha}{1-\alpha} = B \).

Therefore, \( h < 0 \) for all \( \alpha > \alpha_0 = B/(B+1) = (1+1/B)^{-1}. \)

**Lemma IV.2:** Given \( \sigma > 1, k \geq 0, \) and \( \alpha \geq 1/2 \), the function \( h \) has a minimum with respect to \( \delta \), and this \( \delta \)-value does not depend on \( \alpha \). Specifically, the value of \( \delta \) satisfying the equation,
\[ \delta = \frac{\sigma^2 - 1}{\sigma} \xi \left( \frac{k-\delta}{\sigma} \right), \]  

(11)

minimizes \( h \) for the given values of \( k \) and \( \sigma \) and for all \( a \geq 1/2 \).

**proof:** \( h'(\delta | k, \sigma, a) = \frac{3}{3} h(k, \delta, \sigma, a) = 2 \left\{ \delta - \frac{\sigma^2 - 1}{\sigma} \left\{ \xi \left( \frac{k-\delta}{\sigma} \right) \right\} \right\} \),

which vanishes iff (11) holds. Note that this equation does not contain \( a \).

\[ h''(\delta | k, \sigma, a) = \frac{3}{3} h(k, \delta, \sigma, a) = 2 \left\{ 1 + \frac{\sigma^2 - 1}{\sigma^2} \xi' \left( \frac{k-\delta}{\sigma} \right) \right\} \]

\[ = 2 \left\{ 1 - (1 - 1/\sigma^2) \omega \left( \frac{k-\delta}{\sigma} \right) \right\} \] by Lemma II.9

\[ > 0 \] because by Corr. II.2, \( 0 < \omega(t) < 1 \) for all real \( t \).

Therefore \( h'(\delta | k, \sigma, a) \) is an increasing function of \( \delta \).

Furthermore, \( h'(0 | k, \sigma, a) = -2[(\sigma^2 - 1)/\sigma] \xi(k/\sigma) < 0 \).

There exists a value of \( \delta > 0 \) such that (11) holds, for \( h'(\delta | k, \sigma, a) \) can be written

\[ h'(\delta | k, \sigma, a) = 2\sigma \left\{ (\delta/\sigma) - (1 - 1/\sigma^2) \xi([k/\sigma] - [\delta/\sigma]) \right\}. \]  

(12)

Now, because \( \xi'(-t) < 1 \) (Corr. II.2) while the function \( g(t) = t \) has slope 1,

\[ \xi(-t) < t + \xi(0) = t + \sqrt{2/\pi} \] for all \( t > 0 \),

and so for all \( c \geq 0 \) and \( 0 < c < 1 \), and for all \( t > [(1-\varepsilon)/\varepsilon][c+\xi(0)] \),

\[ t - (1-\varepsilon)\xi(c-t) > ct - (1-\varepsilon)[c+\xi(0)] > 0. \]

Applying this result to (12), we get

\[ h'(\delta | k, \sigma, a) \geq 0 \] for all \( (\delta/\sigma) > \sigma^2 (1 - 1/\sigma^2) \xi([k/\sigma] + \xi(0)) \); i.e. for all \( \delta > (\sigma^2 - 1)(k + \sigma \sqrt{2/\pi}) \).

Therefore, there exists a \( \delta \) between 0 and \( (\sigma^2 - 1)(k + \sigma \sqrt{2/\pi}) \) such that \( h'(\delta | k, \sigma, a) = 0 \), and this \( \delta \)-value minimizes \( h(k, \delta, \sigma, a) \) for the given \( k \) and \( \sigma \).

\[ \square \]

**Lemma IV.3:** Given \( \sigma > 1 \) and \( k \geq 0 \), let \( \delta^* \) denote the unique \( \delta \) minimizing \( h(k, \delta, \sigma, a) \) for these values of \( \sigma \) and \( k \), and let \( a_{\delta^*} \) denote the corresponding value of \( a \) for which \( h = 0 \). That is,

\[ \left\{ \begin{array}{l} \delta^* = \delta^*(k, \sigma) = [(\sigma^2 - 1)/\sigma] \xi([k-\delta^*]/\sigma) \\ a_{\delta^*} = a_{\delta^*(k, \delta^*, \sigma)} = (1 + 1/B^*)^{-1} \quad \text{with} \\ B^* = B(k, \delta^*, \sigma) = \sigma \xi([k-\delta^*]/\sigma)/\phi(k) \end{array} \right\} \]  

(13)
If $\alpha \leq \alpha_0^*$, then $h(k, \delta, \sigma, \alpha) \geq 0$ for all $\delta \geq 0$.

**proof:** $h(k, \delta^*, \sigma, \alpha) < h(k, \delta, \sigma, \alpha)$ for all $0 \leq \delta \leq \delta^*$, by Lemma 2.

$h(k, \delta^*, \sigma, \alpha) \geq 0$ for all $\alpha \leq \alpha_0^*$, by Lemma 1.

\[\square\]

**TABLE IV.1**

Values of $\delta^*$ and $\alpha_0^*$ for Selected Values of $k$ and $\sigma$.

(In each cell the top number is $\delta^*$.)

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.00</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>2.00</th>
<th>3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>0.016</td>
<td>0.013</td>
<td>0.010</td>
<td>0.008</td>
<td>0.006</td>
<td>0.004</td>
<td>0.003</td>
<td>0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>0.510</td>
<td>0.501</td>
<td>0.501</td>
<td>0.502</td>
<td>0.502</td>
<td>0.502</td>
<td>0.502</td>
<td>0.502</td>
<td>0.503</td>
</tr>
<tr>
<td>1.05</td>
<td>0.083</td>
<td>0.067</td>
<td>0.054</td>
<td>0.041</td>
<td>0.031</td>
<td>0.022</td>
<td>0.016</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>0.504</td>
<td>0.505</td>
<td>0.506</td>
<td>0.507</td>
<td>0.508</td>
<td>0.509</td>
<td>0.511</td>
<td>0.512</td>
<td></td>
</tr>
<tr>
<td>1.10</td>
<td>0.172</td>
<td>0.140</td>
<td>0.112</td>
<td>0.087</td>
<td>0.066</td>
<td>0.049</td>
<td>0.035</td>
<td>0.016</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>0.508</td>
<td>0.510</td>
<td>0.511</td>
<td>0.513</td>
<td>0.515</td>
<td>0.516</td>
<td>0.518</td>
<td>0.521</td>
<td>0.523</td>
</tr>
<tr>
<td>1.25</td>
<td>0.474</td>
<td>0.392</td>
<td>0.319</td>
<td>0.254</td>
<td>0.198</td>
<td>0.150</td>
<td>0.112</td>
<td>0.057</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>0.518</td>
<td>0.522</td>
<td>0.525</td>
<td>0.528</td>
<td>0.532</td>
<td>0.536</td>
<td>0.540</td>
<td>0.547</td>
<td>0.554</td>
</tr>
<tr>
<td>1.50</td>
<td>1.094</td>
<td>0.918</td>
<td>0.759</td>
<td>0.619</td>
<td>0.498</td>
<td>0.395</td>
<td>0.308</td>
<td>0.179</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>0.530</td>
<td>0.535</td>
<td>0.541</td>
<td>0.547</td>
<td>0.554</td>
<td>0.561</td>
<td>0.568</td>
<td>0.580</td>
<td>0.595</td>
</tr>
<tr>
<td>2.00</td>
<td>2.749</td>
<td>2.345</td>
<td>1.984</td>
<td>1.665</td>
<td>1.388</td>
<td>1.149</td>
<td>0.945</td>
<td>0.627</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>0.544</td>
<td>0.552</td>
<td>0.561</td>
<td>0.571</td>
<td>0.581</td>
<td>0.592</td>
<td>0.603</td>
<td>0.622</td>
<td>0.649</td>
</tr>
<tr>
<td></td>
<td>0.556</td>
<td>0.567</td>
<td>0.579</td>
<td>0.592</td>
<td>0.606</td>
<td>0.620</td>
<td>0.635</td>
<td>0.663</td>
<td>0.705</td>
</tr>
<tr>
<td>5.00</td>
<td>20.641</td>
<td>18.071</td>
<td></td>
<td>8.488</td>
<td>5.424</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.578</td>
<td>0.590</td>
<td></td>
<td>0.691</td>
<td>0.744</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.00</td>
<td></td>
<td></td>
<td></td>
<td>39.17</td>
<td>27.448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.705</td>
<td>0.765</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65.06</td>
<td>0.770</td>
</tr>
</tbody>
</table>

Brief study of this table, together with Figure 1, suggests (at least) two interesting possibilities:

1.) There may be a simple functional relationship between $(\delta^*/\sigma)$ and $\sigma$ for each fixed $k$.

2.) If $\alpha_0^*$ is considered to be a function of $\sigma$, given fixed $k$, then

$$\lim_{\sigma \to \infty} \alpha_0^* < 1$$

for each $k$.

The first of these possibilities is explored in Lemmas 4-7. The second is investigated in Lemma 8 and Table 3.
Figure IV.1: Graphs of $\delta^\%$ versus $\sigma$ for Selected Values of $K$. 
**Lemma IV.4:** Given \( \sigma > 1 \) and \( k > 0 \). If \( \delta^* > k \), then the following equation holds:

\[
(\delta^*/\sigma) = (1 - 1/\sigma^2)[R([\delta^*/\sigma]-[k/\sigma])^{-1},
\]

(14)

where \( R(t) \equiv [1-\Phi(t)]/\phi(t) \), \( t > 0 \), is Mills' ratio.

**proof:** From (13) we get the equation,

\[
(\delta^*/\sigma) = (1 - 1/\sigma^2)\xi([k/\sigma]-[\delta^*/\sigma])
\]

By Lemma II.7, \( R(t) = 1/[\xi(-t)] \) for all \( t > 0 \).

\[ \square \]

**Lemma IV.5:** For a given \( \sigma > 1 \) and \( k > 0 \), \( \delta^* > k \) iff

\[
k < \sqrt{2/\pi}((\sigma^2-1)/\sigma) = 0.7978((\sigma^2-1)/\sigma)
\]

or, alternatively, \( \sigma > \sqrt{2\pi k^2/8} + \sqrt{1 + \pi k^2/8} \).

**proof:** By definition, \( \delta^* \) is the \( \delta \)-value such that \( h'(\delta|k,\sigma) \), given by (12), equals zero.

It was shown in the proof of Lemma 2 that \( h'(\delta|k,\sigma) \) is an increasing function of \( \delta \).

Therefore, if \( \delta^* > k \), then

\[
h'(k|k,\sigma) = 2\sigma([k/\sigma] - (1 - 1/\sigma^2)\xi(0)) < h'(\delta^*|k,\sigma) = 0,
\]

which implies that

\[
k < \frac{\sigma^2-1}{\sigma} \xi(0) = \frac{\sigma^2-1}{\sigma} \sqrt{2/\pi}.
\]

(15)

Conversely, if (15) holds, then \( h'(k|k,\sigma) < 0 \), which implies \( \delta^* > k \).

Now, (15) holds iff

\[
\sigma > \frac{k\sqrt{\pi} + \sqrt{\pi k^2 + 8}}{2\sqrt{2}}
\]

\[ \square \]

**Table IV.2**

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( k_0 \equiv \sqrt{2/\pi}((\sigma^2-1)/\sigma) )</th>
<th>( \sigma )</th>
<th>( k_0 )</th>
<th>( \sigma )</th>
<th>( k_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>0.0159</td>
<td>4.00</td>
<td>2.9921</td>
<td>15.00</td>
<td>11.92</td>
</tr>
<tr>
<td>1.10</td>
<td>0.0523</td>
<td>4.50</td>
<td>3.4132</td>
<td>20.00</td>
<td>15.92</td>
</tr>
<tr>
<td>1.23</td>
<td>0.3590</td>
<td>5.00</td>
<td>3.8298</td>
<td>25.00</td>
<td>19.92</td>
</tr>
<tr>
<td>1.50</td>
<td>0.6649</td>
<td>6.00</td>
<td>4.6543</td>
<td>30.00</td>
<td>23.91</td>
</tr>
<tr>
<td>1.75</td>
<td>0.9404</td>
<td>7.00</td>
<td>5.4712</td>
<td>35.00</td>
<td>27.90</td>
</tr>
<tr>
<td>2.00</td>
<td>1.1968</td>
<td>8.00</td>
<td>6.2833</td>
<td>40.00</td>
<td>31.90</td>
</tr>
<tr>
<td>2.50</td>
<td>1.6756</td>
<td>9.00</td>
<td>7.0923</td>
<td>45.00</td>
<td>35.89</td>
</tr>
<tr>
<td>3.00</td>
<td>2.1277</td>
<td>10.00</td>
<td>7.8991</td>
<td>50.00</td>
<td>39.88</td>
</tr>
<tr>
<td>3.50</td>
<td>2.5646</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Lemma IV.6:** If \( k=0 \), then for all \( \sigma > 1 \),
\[
\frac{\sqrt{\sigma^2-1}}{\sqrt{\pi-2}} \left( 1 - \frac{\pi}{(4-\pi)+2(\pi-2)\sigma^2} \right)^{1/2} < \frac{\delta^*}{\sigma} \leq \frac{\sqrt{\sigma^2-1}}{\sqrt{1 + \pi/[2(\sigma^2-1)]}}.
\]

Consequently, for "large",
\[
0.936\sqrt{\sigma^2-1} \leq (\delta^*/\sigma) < \sqrt{\sigma^2-1},
\]

where \( f(t) \leq g(t) \) iff \( f(t)[1-o(1)] < g(t) \) as \( t \to \infty \). \hfill (15.5)

**proof:** When \( k=0 \), \( \delta^* > k = 0 \) for all \( \sigma > 1 \), by Lemma 5; and so, by Lemma 4, (14) holds.

Since \( (\delta^*/\sigma) \) could be relatively small, let us consider bounds for \( R(t) \), \( t > 0 \), stated by Boyd [5]:
\[
\frac{\pi}{\sqrt{t^2+2\pi} + (\pi-1)t} < R(t) < \frac{\pi}{(\pi-2)[\sqrt{t^2+2\pi/(\pi-2)^2} + 2t/(\pi-2)]}.
\]

In somewhat simplified notation (14) can be written
\[
X R(X) = W, \hfill (17)
\]
with
\[
\begin{cases}
W = 1-(1/\sigma^2) \in (0,1) \\
X = \delta^*/\sigma
\end{cases} \hfill (18)
\]

From (17) and the lower bound of (16) it follows that
\[
\frac{\pi X}{\sqrt{X^2 + 2\pi} + (\pi-1)X} < X R(X) = W,
\]
i.e. \([W+(1-W)\pi]X < W\sqrt{X^2+2\pi} \)
\[
\text{i.e. } (1-W)[2W+\pi(1-W)]X^2 < 2W^2;
\]
\[
\text{i.e. } 0 < X < W \frac{\sqrt{2}}{(1-W)[2W+\pi(1-W)]} = \sqrt{\sigma^2-1} \frac{\sqrt{2}(\sigma^2-1)}{\sqrt{2}(\sigma^2-1)+\pi}. \hfill (19)
\]

From (17) and the upper bound of (16) it follows that
\[
W = X R(X) < \frac{\pi X}{2X + \sqrt{2\pi+X^2(\pi-2)^2}};
\]
i.e. \( W^2[2\pi+X^2(\pi-2)^2] < (\pi-2W)^2X^2 \)
\[
\text{i.e. } 2\pi W^2 < X^2[(\pi-2W)-W^2(\pi-2)^2], \text{ which is positive because}
\]
\[
0<W<1 \text{ implies } \pi-2W > \pi-2 > W(\pi-2) > 0;
\]
\[
\text{i.e. } 2W^2 < X^2[\pi-4W+W^2] = X^2(1-W)[\pi(1+W)-4W]; \hfill (20)
\]
\[
\text{i.e. } X > \frac{W\sqrt{2}}{(1-W)[\pi(1+W)-4W]} = \frac{\sqrt{\sigma^2-1}2(\sigma^2-1)}{\sqrt{2(\pi-2)\sigma^2+(4-\pi)}};
\]
\[ (\delta^*/\sigma) > \frac{\sqrt{\sigma^2-1}}{\sqrt{\pi-2}} \left( 1 - \frac{\pi}{2(\pi-2)\sigma^2+(4-\pi)} \right)^{1/2}. \]

\[ \square \]

**Lemma IV.7:** If \( k>0 \), then for all \( \sigma > \sqrt{\pi k^2/8} + \sqrt{1 + \pi k^2/8} \), the following inequalities hold:

\[ L_0[-(1+b) + \sqrt{1+t^2}] < (\delta^*/\sigma) < U_0[-u(1+a) + \sqrt{1+u^2}], \]

where \( U_0 \) and \( L_0 \) are the upper and lower bounds, respectively, for \( (\delta^*/\sigma) \) when \( k=0 \), which are given in Lemma 6, and

\[
\begin{cases}
  b = \frac{(4-\pi)}{[\sigma^2(\pi-2)]} ; & a = \frac{(\pi-2)}{\sigma^2} \\
  t = k\sqrt{\pi-2}/(2\sqrt{1+b/2}) ; & u = k/(2\sqrt{1+a/2})
\end{cases}
\]

Consequently, for "large" \( \sigma \), the bounds on \( (\delta^*/\sigma) \) are approximately

\[ (\sqrt{\sigma^2-1}\sqrt{\pi-2}) g(k\sqrt{\pi-2}/2) ; (\delta^*/\sigma) < \sqrt{\sigma^2-1} g(k/2), \]

where \( g(t) = \sqrt{1+t^2} - t \) and \( \xi \) is defined in (15.5).

**Proof:** By Lemmas 4 and 5, \( t \) is positive.

Therefore Boyd's bounds (16) may be applied to \( R(\cdot) \).

Using the notation defined in (18) and defining

\[ Z = k/\sigma \in (0, 0.7978W) \text{ because } \delta^* \leq k \text{ implies } k \leq 0.7978(\sigma^2-1)/\sigma, \quad (22) \]

we can write (14) as

\[ X R(X-Z) = W. \quad (23) \]

**Upper Bound:**

From (23) and the lower bound of (16) it follows that

\[ \frac{\pi X}{(\pi-1)(X-Z)+\sqrt{(X-Z)^2+2\pi}} < X R(X-Z) = W; \]

i.e. \( X[W+\pi(1-W)] + W(\pi-1)Z < W\sqrt{(X-Z)^2+2\pi} \);

i.e. \( X^2((1-W)[2W+\pi(1-W)]) + 2WZX[2W+\pi(1-W)-1] - W^2[2-Z^2(\pi-2)] < 0. \quad (24) \)

Since the L.H.S. of (24) is a U-shaped quadratic, \( AX^2+2BX+C \), with \( A, B, C \) all positive, (24) is satisfied iff \( X \) lies in the interval having endpoints \((-B+\sqrt{B^2+4AC})/A. \) Because this interval contains 0, but Lemma 5 restricts \( X \) to the domain, \( X>0 \), (24) can be satisfied iff
\[ Z < X < \frac{-WZ(V-1) + W\sqrt{Z^2(V-1)^2 + (1-W)W[2-Z^2(\pi-2)]}}{(1-W)V} \]

with \( V \equiv 2W + \pi(1-W) \) \hfill (25)

\[ = \frac{W}{\sqrt{(1-W)V}} \left\{ \frac{-Z(V-1)}{\sqrt{V(1-W)}} + \left\{ 2 + \frac{Z^2}{(1-W)V} \right\}^{1/2} \right\} \]

\[ = U_o \left\{ \frac{-Z(V-1)}{\sqrt{2V(1-W)}} + \left[ 1 + \frac{Z^2}{2V(1-W)} \right]^{1/2} \right\} , \] \hfill (26)

where \( U_o \equiv \sqrt{2W^2/[V(1-W)]} \) is the upper bound on \( X \) when \( k=0 \) [cf. (19) and (25)]. Returning to the original parameters, we have

\[ V = 2(\sigma^2 - 1)/\sigma^2 + \pi/\sigma^2 , \]

so that \( u \equiv \sqrt{Z^2/[2V(1-W)]} = \sqrt{(k/2)^2 + (1 + a/2)} \),

with \( a \equiv (\pi-2)/\sigma^2; \)

and \( V-1 = [2(\sigma^2 - 1) + \pi - a]/\sigma^2 = 1+a. \)

Thus (26) can be written

\[ U_o \{-u(1+a) + \sqrt{1+u^2} \} = U_o \{g(u) - au\} . \]

**Lower Bound:**

From (23) and the upper bound of (16) it follows that

\[ W = X R(X-Z) < \frac{\pi X}{\sqrt{2\pi + (\pi-2)^2}} \]

\[ \sqrt{2\pi + (\pi-2)^2 (X-Z)^2 + 2(X-Z)^2} \]

i.e. \( W^2[2\pi + (\pi-2)^2(X-Z)^2] < [X(\pi-2W) + 2WZ]^2 \); or

\[ X^2[\pi-2W^2 + 2X(2WZ(\pi-2W) + WZ(\pi-2)]^2 + W^2[4\pi^2 - 2\pi + 2\pi^2(\pi-2)^2}] \]

\[ > 0, \]

where the coefficient of \( X^2 \) is clearly positive;

i.e. \( X^2(1-W)[\pi(1+\pi) - 4W] + 2WZX[2-W(\pi-4\pi)] > 0 \). \hfill (27)

Since the L.H.S. of (27) is a quadratic in \( X \), i.e. \( AX^2 + 2BX - C \), with \( A, B, C \) all positive as a consequence of the restrictions on \( W \) and \( Z \), and since \( X \) must be positive, (27) can be satisfied iff

\[ X > (-B + \sqrt{B^2 + 4AC})/A ; \]

i.e. \[ X > \frac{-WZ[E-(\pi-2)] + \sqrt{(WZ)^2[E-(\pi-2)]^2 + W^2(1-W)[E[2-Z^2(4-\pi)]]} }{(1-W)E} \]

\[ \text{with } E \equiv \pi(1+\pi) - 4W; \] \hfill (27.5)
\[ X > L_0 \left\{ \frac{-Z[E-(\pi-2)]}{\sqrt{2(1-W)E}} + \left[ 1 + \frac{((\pi-2)Z)^2}{2E(1-W)} \right]^{1/2} \right\}, \tag{29} \]

where \( L_0 = \sqrt{2W^2/(1-W)E} \) is the lower bound on \( X \) when \( k=0 \) [cf. (20) and (28)]. In the original parameters

\[ E = 2(\pi-2) + (4-\pi)/\sigma^2, \]

so that

\[ t = \sqrt{[(\pi-2)Z]^{2}/[2E(1-W)]} = k\sqrt{\pi-2}/(2\sqrt{1+b/2}) \]

with \( b \equiv (4-\pi)/[\sigma^2(\pi-2)] \); and

\[ [(E-(\pi-2))/(\pi-2)] = 1+b. \]

Therefore (29) can be written

\[ L_0 \left\{ -t(1+b) + \sqrt{1+t^2} \right\} = L_0 \left\{ g(t) - bt \right\}. \]

Let us check to see whether this lower bound exceeds \( Z \), the lower bound required for the valid application of Boyd's bounds. Using (27.5), we have

\[ -WZ[E-(\pi-2)] + W\sqrt{Z^2[E-(\pi-2)]^2 + (1-W)E[2Z^2(4-\pi)]} > Z \]

\[ (1-W)E \]

iff

\[ Z^2[(\pi-2)W^2 - W^2(\pi-2)^2] < 2W^2(1-W)E \]

iff

\[ Z^2(1-W)E < 2W^2(1-W)E \]

iff

\[ 0 < Z < \frac{W\sqrt{2/\nu}}{W} = \xi(0) \]

iff

\[ 0 < k < \sigma[(\sigma^2-1)/\sigma^2] \xi(0). \]

This last inequality is satisfied because of the given condition on \( \sigma \) and Lemma 5.

Since \( \lim a = \lim b = 0 \) as \( \sigma \to \infty \), the coefficients of \( U_o \) and \( L_o \) tend to \( g(k/2) \) and \( g([k/2]/\sqrt{4-\pi}) \), respectively, as \( \sigma \to \infty \), where \( g(t) = \sqrt{1+t^2} + t \).

From Lemma 6 we see that for "large" \( \sigma \),

\[ L_o \approx \frac{\sqrt{\sigma^2-1}}{\sqrt{\pi-2}} \quad \text{and} \quad U_o \approx \frac{\sqrt{\sigma^2-1}}{\sqrt{\pi-2}}. \]

In summary, then, Lemmas 6 and 7 suggest that for fixed \( k \geq 0 \) and for large \( \sigma \),

\[ (\delta^#/\sigma) \approx c(k)\sqrt{\sigma^2-1}, \]

where \( c(k) \) is a constant between 0 and 1 which depends on \( k \); and, furthermore, \( c(k) \) may be approximately \( g(k/2) = \sqrt{1+(k/2)^2} - (k/2). \)
Consider now the behavior of \( a_0^* \) when \( \sigma \) is large. Recall [cf. (10) and (13)] that \( a_0^* \) is the value of \( \alpha \) for which \( h(k, \delta^*, \sigma, \alpha) = 0 \), where \( \delta^* = \delta^*(k, \sigma) \) is the unique \( \delta \)-value minimizing \( h(k, \delta, \sigma, \alpha) \) for the given values of \( k \) and \( \sigma \).

**Lemma IV.8:** If \( k \geq 2 \) and \( \sigma > (k/2)^{\sqrt{\pi/2}} + \sqrt{1 + \pi k^2/8} \) are given, then under the assumption that \( (\delta^*/\sigma) = c(k, \sigma)\sqrt{\sigma^2 - 1} \), \( 0 < c(k, \sigma) \leq 1 \), the function \( a_0^* \) has a unique minimum when

\[
c(k, \sigma) = g([k/2]^{\sqrt{1 - \frac{1}{\sigma^2}}}).
\]

**Proof:** Recall from Lemma 3 that \( a_0^* = (1 + 1/B^*)^{-1} \), where

\[
B^* = B(k, \delta^*, \sigma) = \left\{ \sigma \left( \frac{k-\delta}{\sigma} \right) + \phi(k) \right\} \exp\left( \frac{\delta^*}{2(\sigma^2 - 1)} \right)
\]

and \( (\delta^*/\sigma) = (1 - 1/\sigma^2) \xi([k/\sigma] - [\delta^*/\sigma]). \)

It follows from these facts and the fact that \( \xi(t) = \phi(t) \div \phi(t) \) that

\[
B^* = \left\{ \sigma(1 - 1/\sigma^2) \frac{\phi([k/\sigma] - [\delta^*/\sigma])}{(\delta^*/\sigma)} \frac{\phi(k)}{\exp\left( 1/2 \right)} \right\} \exp\left\{ \frac{(\delta^*/\sigma)}{\sqrt{\sigma^2 - 1}} + k\sqrt{1 - 1/\sigma^2} \right\} \exp\left( (\delta^*/\sigma)^2/2(\sigma^2 - 1) \right). \tag{30}
\]

Suppose that \( (\delta^*/\sigma) = c\sqrt{\sigma^2 - 1} \) with \( 0 < c \leq 1 \).

Substituting this into (30), we get

\[
B^* = (1/c)^{\sqrt{1 - 1/\sigma^2}} \xi(k) \exp\left( (c + k\sqrt{1 - 1/\sigma^2})^2/2 \right).
\]

Considering \( B^* \) to be a function of \( c \), given \( k \) and \( \sigma \), we differentiate it:

\[
B^*(c|k, \sigma) = \sqrt{1 - 1/\sigma^2} \xi(k) \frac{c^2 + ck\sqrt{1 - 1/\sigma^2} - 1}{c^2} \exp\left( \frac{(c+k\sqrt{1 - 1/\sigma^2})^2}{2} \right), \tag{31}
\]

which vanishes iff \( c = g([k/2]^{\sqrt{1 - 1/\sigma^2}}). \)

Now, because the quadratic on the R.H.S. of (31) is U-shaped, it is negative for all positive \( c < g([k/2]^{\sqrt{1 - 1/\sigma^2}}) \) and positive for all larger \( c \); and since this quadratic is multiplied by a quantity which is positive for all positive \( c \), \( B^*(c|k, \sigma) \) is positive or negative according as \( c \) exceeds \( g([k/2]^{\sqrt{1 - 1/\sigma^2}}) \) or not.

Therefore \( B^*(c|k, \sigma) \) is minimized when \( c = g([k/2]^{\sqrt{1 - 1/\sigma^2}}). \)

Now, \( a_0^* = (1 + 1/B^*)^{-1} = B^*/(B^*+1) \), which is a strictly in-
creasing function of $B^*$.

Therefore, $a_o^*$ is minimized when $B^*$ is minimized.

□

Lemmas 3 and 6-8 suggest that even if $\sigma$ is very large, the optimal decision rule based on $X_1$ will not be trivial (i.e. "ignore the data") as long as $\alpha<\alpha_o(k)$, where $k$ is the truncation point and

$$
\alpha_o(k) = \left[1 + 1/B(k)\right]^{-1}, \text{ with}
$$

$$
B(k) = \frac{\xi(k)}{g(k/2)} \exp\left\{[k+g(k/2)]^2/2\right\}.
$$

This formula for $B(k)$ is obtained by substituting $(\delta^*/\sigma)=g(k/2)\sqrt{\sigma^2-1}$ into (30). It is easy to compute as long as $k$ is not too large, say $k<25$. For larger $k$, use the following formula, obtained by substituting $\xi(k) = \phi(k)\phi(k) = \phi(k)$:

$$
B(k) = \frac{\exp\left\{g(k/2)[2k+g(k/2)]/2\right\}}{\sqrt{2\pi} \ g(k/2)}.
$$

Table 3 shows values of $\alpha_o(k)$ for selected values of $k$. In the third column are the values of $\alpha_o^*(k,\sigma)$ for the corresponding $k$ and the largest $\sigma$ computed in Table 1.

TABLE IV.3: Values of $\alpha_o(k)$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\alpha_o(k)$</th>
<th>$\alpha_o^*(k,\sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.5681</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.5732</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.5813</td>
<td>0.5666 (σ=3)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5969</td>
<td>0.5785 (σ=3)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.6141</td>
<td>0.5916 (σ=2)</td>
</tr>
<tr>
<td>1.00</td>
<td>0.6328</td>
<td>0.6057 (σ=3)</td>
</tr>
<tr>
<td>1.25</td>
<td>0.6523</td>
<td>0.6204 (σ=3)</td>
</tr>
<tr>
<td>1.50</td>
<td>0.6722</td>
<td>0.6351 (σ=3)</td>
</tr>
<tr>
<td>2.00</td>
<td>0.7109</td>
<td>0.6629 (σ=3)</td>
</tr>
<tr>
<td>3.00</td>
<td>0.7741</td>
<td>0.7046 (σ=3)</td>
</tr>
<tr>
<td>5.00</td>
<td>0.8468</td>
<td>0.7046 (σ=3)</td>
</tr>
<tr>
<td>7.00</td>
<td>0.8846</td>
<td></td>
</tr>
<tr>
<td>10.00</td>
<td>0.9160</td>
<td></td>
</tr>
<tr>
<td>15.00</td>
<td>0.9422</td>
<td></td>
</tr>
<tr>
<td>20.00</td>
<td>0.9560</td>
<td></td>
</tr>
<tr>
<td>25.00</td>
<td>0.9645</td>
<td></td>
</tr>
<tr>
<td>50.00</td>
<td>0.9819</td>
<td></td>
</tr>
<tr>
<td>100.00</td>
<td>0.9909</td>
<td></td>
</tr>
<tr>
<td>200.00</td>
<td>0.9954</td>
<td></td>
</tr>
</tbody>
</table>
4.3 Behavior of the Function $Q(y)$.

Recall, from Corollary 1, that the optimal decision rule based on the linear function $Y = dX_1 + X_2$ is to choose Population 1 iff

$$U(y) = Q(y) + G(y) < 0,$$

where

$$Q(y) = Dy^2 - 2Ay + B,$$

with

$$\begin{align*}
D &= \theta_2 - \theta_1 = (1 - \sigma_1^2)d^2 + 2(\rho_2 - \rho_1 \sigma_1 \sigma_2) d + (1 - \sigma_2^2) \\
A &= \nu \theta_2 = (d \delta_1 + \delta_2)(d^2 + 2\rho_2 d + 1) \\
B &= \theta_2 \{\nu^2 - 2\theta_1 \ln[(\alpha/(1-\alpha))\nu \theta_2 / \theta_1 \phi(k)/\phi([k-\delta_1]/\sigma_1)]\}.
\end{align*}$$

(32)

In this section we shall consider how the shape of the quadratic $Q(y)$ is affected by changes in $d$ and the parameters $\pi' = (\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha, k)$.

The next lemma follows immediately from (32).

**Lemma IV.9**: The quadratic $Q(y)$ depends on $\alpha$ only through the constant term $B$, and the coefficient of the $y^2$-term is independent of $\delta$.

**Lemma IV.10**: If $d$ and all the other parameters are held fixed, then $Q(y)$ decreases as $\alpha$ increases.

**Proof**: When $\alpha = 1/2$, then

$$B = \theta_2 \{\nu^2 - 2\theta_1 \ln[\sqrt{\theta_2 / \theta_1} \phi(k)/\phi([k-\delta_1]/\sigma_1)]\} \equiv B_0.$$

For $\alpha < 1/2$, $B = B_0 - 2\theta_1 \theta_2 \ln[\alpha/(1-\alpha)]$, which is a decreasing function of $\alpha$.

Since $Q(y)$ depends on $\alpha$ only through $B$, it is also a decreasing function of $\alpha$.

**Lemma IV.11**: In Tables 4-6 are given the necessary and sufficient conditions for $Q(y)$ to attain each of its possible shapes. [The word NONE implies that only the basic restrictions on the parameters assumed throughout this chapter are required: namely, for $i=1,2$, $\delta_i \geq 0$, $\sigma_i > 0$, $|\rho_i| < 1$, and $d$ is real.] The numbers $d_o, d_1, d_2, \rho'$, and $\rho''$ are the following functions of the parameters comprising $\Sigma$ and $R$:

$$\begin{align*}
\left\{ \begin{array}{l}
d_o = (\sigma_2^2 - 1)/[2(\rho_2 - \rho_1 \sigma_2)] \\
((d_1, d_2)) = [(\rho_1 \sigma_2 \sigma_2 - \rho_2) \pm \sqrt{(\rho_2 - \rho_1 \sigma_1 \sigma_2)^2 - (1 - \sigma_1^2)(1 - \sigma_2^2)}]/(1 - \sigma_1^2) \\
((\rho'', \rho')) = \rho_1 \sigma_1 \sigma_2 \pm \sqrt{(1 - \sigma_1^2)(1 - \sigma_2^2)}
\end{array} \right\}
\end{align*}$$

(33)
A.) Q(y) is a constant function iff one of the following sets of conditions holds:

<table>
<thead>
<tr>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>d</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>where . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1&lt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>NONE</td>
<td>$\delta_2=-\delta_1d_1$</td>
<td>$i=1$ or $2$.</td>
</tr>
<tr>
<td>$\sigma_1=1$</td>
<td>$\sigma_2=1$</td>
<td>NONE</td>
<td>NONE</td>
<td>d</td>
<td>NONE</td>
<td>$\delta_2=-\delta_1d_1$</td>
<td>$i=1$ or $2$.</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>NONE</td>
<td>d</td>
<td>NONE</td>
<td>$\delta_2=-\delta_1d_1$</td>
<td>$i=1$, or $2$.</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&gt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>NONE</td>
<td>$\delta_2=-\delta_1d_1$</td>
<td>$i=1$ or $2$.</td>
</tr>
</tbody>
</table>

B.) Q(y) is a linear function (of non-zero slope) iff the above sets of conditions on $\sigma_1, \sigma_2, \rho_1, \rho_2, d$, and $\delta_1$ do hold but the corresponding conditions on $\delta_2$ do not.

C.) Q(y) is a U-shaped quadratic iff one of the following sets of conditions holds:

<table>
<thead>
<tr>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>d</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1&lt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1=1$</td>
<td>$\sigma_2=1$</td>
<td>NONE</td>
<td>NONE</td>
<td>d</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>$d_1, d_2$</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&gt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>NONE</td>
<td>NONE</td>
</tr>
</tbody>
</table>

D.) Q(y) is an arch-shaped quadratic iff one of the following conditions holds:

<table>
<thead>
<tr>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>d</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1&lt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>$d(a_1, a_2)$</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1=1$</td>
<td>$\sigma_2=1$</td>
<td>NONE</td>
<td>NONE</td>
<td>d</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&lt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>$d(a_1, a_2)$</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1&gt;1$</td>
<td>$\sigma_2&gt;1$</td>
<td>NONE</td>
<td>$\rho_2(\rho', \rho''')$</td>
<td>d</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$\sigma_2$</td>
<td>$\rho_1$</td>
<td>$\rho_2$</td>
<td>$d$</td>
<td>$\delta_1$</td>
<td>$\delta_2$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma_1 &lt; 1$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
<td>$\rho_2 \geq \rho_1 \sigma_2$</td>
<td>$d &lt; d_0$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
</tr>
<tr>
<td>$\sigma_1 \geq 1$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
<td>$\rho_2 \notin (\rho', \rho'')$</td>
<td>$d \notin [d_2, d_1]$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
</tr>
<tr>
<td>$\sigma_2 &lt; 1$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
<td>$\rho_2 \notin (\rho', \rho'')$</td>
<td>$\text{n}$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
</tr>
<tr>
<td>$\sigma_2 &gt; 1$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
<td>$\rho_2 \notin (\rho', \rho'')$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
<td>$\text{NONE}$</td>
</tr>
</tbody>
</table>

**Proof:** From (32) we see that

1. when $\sigma_1 < 1$, D is a U-shaped quadratic in $d$;
2. when $\sigma_1 = 1$, D is a linear function of $d$;
3. when $\sigma_1 > 1$, D is an arch-shaped quadratic in $d$.

Furthermore,

4. when $\sigma_1 = 1$, $D = 0$ iff $d = d_1$ or $d = d_2$, defined in (33);
5. when $\sigma_1 = 1$, $D = 0$ iff $d = d_0$, defined in (33), provided $\rho_2 \neq \rho_1 \sigma_2$;
6. when $\sigma_1 = 1$ and $\rho_2 = \rho_1 \sigma_2$, then D is a constant function of $d$.

Moreover, $d_1$ and $d_2$ are real iff $(\rho_2 - \rho_1 \sigma_2) \geq (1-\sigma_1^2)(1-\sigma_2^2)$, which can happen iff one of the following three conditions holds:

7. $\sigma_1 < 1 \leq \sigma_2$;
8. $\sigma_2 < 1 \leq \sigma_1$;
9. Conditions (7) and (8) do not hold, and $\rho_2 \notin (\rho', \rho'')$, defined in (33).

From (32) we see also that, because $\theta_2 > 0$ for all $d$, $A = 0$ iff

10. $\delta_2 = -\delta_1 d$.

Recalling the definition of $Q(y)$, i.e. $Q(y) = Dy^2 - 2Ay + B$, we see that

A.) $Q(y)$ will be a constant function iff $D = A = 0$, which can happen iff
   a.) Condition (4) and one of conditions (7)-(9) hold, together with Condition (10); or
   b.) Conditions (5) and (10) hold.
   [Note that when Condition (6) holds, then the requirement that $D = 0$ implies $\sigma_1 = \sigma_2 = 1$ and $\rho_1 = \rho_2$, i.e. $\Sigma = R$. This violates Basic Assumption #3.]
B.) $Q(y)$ will be a linear function of non-zero slope iff $D = 0$
and \( A \neq 0 \), which can happen iff

a.) Condition (4) and one of Conditions (7)-(9) hold but Condition (10) does not;

b.) Condition (5) holds but Condition (10) does not.

C.) \( Q(y) \) will be a U-shaped quadratic iff \( D > 0 \). From Conditions (1)-(6) we see that \( D > 0 \) for

a.) all \( d \) when \( \sigma_1 < 1 \) and \( d_1 \) and \( d_2 \) are imaginary;

b.) \( df[d_1,d_2] \) when \( \sigma_1 < 1 \) and one of conditions (7)-(9) holds;

c.) all \( d \) when \( \sigma_2 < 1 = \sigma_1 \) and \( \rho_2 = \rho_1 \sigma_2 \);

d.) \( d < 0 \) or \( d > d_0 \) according as \( \rho_2 < 0 > \rho_1 \sigma_2 \) when \( \sigma_1 = 1 \);

e.) \( d_2 < d < d_1 \) when \( \sigma_1 > 1 \) and one of Conditions (7)-(9) holds.

D.) \( Q(y) \) is an arch-shaped quadratic iff \( D < 0 \). The conditions on \( d \) are analogous to (a)-(e) above.

\( \square \)

1.4 Behavior of the Function \( G(y) \).

In order to study the function \( G(y) \), let us collect those definitions pertaining to it:

\[
\begin{align*}
G(y) &= -2\theta_1 \theta_2 g(y) \\
\theta_1 &= \sigma_1^2 d^2 + 2\rho_1 \sigma_1 \sigma_2 d + \sigma_2^2 \\
\theta_2 &= d^2 + 2\rho_2 d + 1 \\
g(y) &= \ln[\phi(a_1-b_1 y) + \phi(a_2-b_2 y)] \\
\frac{a_1}{\sqrt{\theta_1}} &= \frac{\sqrt{\Sigma}}{\sqrt{\theta_1}} - \frac{[d^2 \Sigma \left( u_2, -u_1 \right)]}{(\sqrt{\Sigma})^2} \\
\frac{a_2}{\sqrt{\theta_2}} &= \frac{\sqrt{\Sigma}}{\sqrt{\theta_2}} \\
b_1 &= \frac{[\sigma_1 (\sigma_1 d + \rho_1 \sigma_2)]}{(\sqrt{\Sigma})^2} \\
b_2 &= \frac{(d + \rho_2)}{(\sqrt{\Sigma})^2}
\end{align*}
\]  

(34)

**Lemma IV.12:** \( G(y) \) is independent of \( a \) and depends on \( \delta \) only through \( a_1 \).

**Proof:** Follows immediately from (34).

\( \square \)

**Lemma IV.13:** If \( b_1 \neq b_2 \), then \( G(y) = 0 \) iff \( y = y_0 = (a_2-a_1)/(b_2-b_1) \). If \( b_1 = b_2 \), there is no real \( y \) such that \( G(y) = 0 \) unless \( a_1 = a_2 \), in which case \( G(y) \equiv 0 \).
proof: Since $\theta_i > 0$, $i=1,2$, we have $G(y)=0$ iff $g(y)=0$, iff $\phi(a_1-b_1y) = \phi(a_2-b_2y)$ iff $a_1-b_1y = a_2-b_2y$.

Hence the result. \hfill \Box

Since $G(y)$ is merely a negative multiple of the function $g(y)$, it is sufficient to study the shape of $g(y)$ in order to discover how $G(y)$ behaves.

4.4.1 Behavior of $g(y)$.

**Lemma IV.14:** The first and second derivatives of $g(y)$ are
\[
\begin{align*}
g'(y) &= b_2 \xi(a_2-b_2y) - b_1 \xi(a_1-b_1y) \\
g''(y) &= b_2^2 \omega(a_2-b_2y) - b_1^2 \omega(a_1-b_1y)
\end{align*}
\]
where $\xi(t) \equiv \phi(t) + \phi(t)$ and $\omega(t) = -\xi'(t) = \xi(t)[t+\xi(t)]$ are the functions studied in some detail in Chapter II.

**Proof:**
\[
g'(y) = \frac{d}{dy}[\ln \phi(a_1-b_1y) - \ln \phi(a_2-b_2y)]
\]
\[
= -b_1[\phi(a_1-b_1y) + \phi(a_1-b_1y)] + b_2[\phi(a_2-b_2y) + \phi(a_2-b_2y)].
\]
\[
g''(y) = \frac{d}{dy} g'(y) = -b_2^2 \xi'(a_2-b_2y) + b_1^2 \xi'(a_1-b_1y).
\]

**Lemma IV.15:** The limits as $y \to \pm\infty$ of $g(y)$, $g'(y)$, $g''(y)$ are given in Table 7.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\lim_{y \to -\infty} g(y)$</th>
<th>$\lim_{y \to +\infty} g(y)$</th>
<th>$\lim_{y \to -\infty} g'(y)$</th>
<th>$\lim_{y \to +\infty} g'(y)$</th>
<th>$\lim_{y \to -\infty} g''(y)$</th>
<th>$\lim_{y \to +\infty} g''(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_2 &lt; b_1 &lt; 0$</td>
<td>$\mp\infty$ (or $0$)</td>
<td>$-\infty$</td>
<td>$-\infty$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &gt; 0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$b_1 = b_2 &lt; 0$</td>
<td>$\ell n \phi(a_1-a_2)$</td>
<td>$0$</td>
<td>$b(a_1-a_2)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$b_1 &lt; b_2 &lt; 0$</td>
<td>$\ell n \phi(a_1)$</td>
<td>$0$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &lt; 0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$b_1 &lt; b_2 &lt; 0$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$-\xi'(a_2)$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &lt; 0$</td>
</tr>
<tr>
<td>$b_1 = 0 &lt; b_2$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &gt; 0$</td>
</tr>
<tr>
<td>$0 &lt; b_1 &lt; b_2$</td>
<td>$-\ell n \phi(a_2)$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &gt; 0$</td>
</tr>
<tr>
<td>$0 &lt; b_1 = b_2$</td>
<td>$0$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$b(a_1-a_2)$</td>
<td>$0$</td>
<td>$0$</td>
</tr>
<tr>
<td>$0 &lt; b_2 &lt; b_1$</td>
<td>$-\ell n \phi(a_1)$</td>
<td>$+\infty$</td>
<td>$0$</td>
<td>$-\xi'(a_2)$</td>
<td>$-\infty$</td>
<td>$0$</td>
</tr>
<tr>
<td>$0 &lt; b_2 &lt; b_1$</td>
<td>$-\ell n \phi(a_2)$</td>
<td>$-\infty$</td>
<td>$-\infty$</td>
<td>$-\infty$</td>
<td>$0$</td>
<td>$b_2^2 - b_1^2 &gt; 0$</td>
</tr>
</tbody>
</table>
proof: Let us obtain these results in ascending order of difficulty, discussing first \( g''(y) \), then \( g'(y) \), and finally \( g(y) \).

Recall from Lemma II.8 and Theorem II.5 that \( \xi(t) \) and \( \omega(t) \) are strictly decreasing functions, that both tend to zero as \( t \to +\infty \), and that \( \xi(t) \to +\infty \) and \( \omega(t) \to 1 \) as \( t \to -\infty \).

A.) From Lemma 14 we have \( g''(y) = b_2^2 \omega(a_2-b_2 y) - b_1^2 \omega(a_1-b_1 y) \).

Consequently, for Cases (1)-(4) and (9), with \( b_1 \leq 0 \) and \( b_2 \leq 0 \),

\[
\lim_{y \to +\infty} g''(y) = 0 \quad \text{and} \quad \lim_{y \to -\infty} g''(y) = b_2^2 - b_1^2.
\]

For Cases (6)-(8) and (11)-(12), with \( b_1 \geq 0 \) and \( b_2 \geq 0 \),

\[
\lim_{y \to -\infty} g''(y) = 0 \quad \text{and} \quad \lim_{y \to +\infty} g''(y) = b_2^2 - b_1^2.
\]

For Case (5), with \( b_1 < 0 < b_2 \),

\[
\lim_{y \to -\infty} g''(y) = -b_1^2, \quad \lim_{y \to +\infty} g''(y) = b_2^2.
\]

For Case (10), with \( b_2 < 0 < b_1 \),

\[
\lim_{y \to -\infty} g''(y) = b_2^2, \quad \lim_{y \to +\infty} g''(y) = -b_1^2.
\]

B.) From Lemma 14 we have \( g'(y) = b_2 \xi(a_2-b_2 y) - b_1 \xi(a_1-b_1 y) \).

For Cases (1)-(4) and (9), with \( b_1 \leq 0 \), \( i=1,2 \), \( \lim_{y \to +\infty} g'(y) = 0 \).

For Cases (6)-(8) and (11)-(12), with \( b_1 \geq 0 \), \( i=1,2 \), \( \lim_{y \to -\infty} g'(y) = 0 \).

For Case (5), with \( b_1 < 0 < b_2 \),

\[
\lim_{y \to -\infty} g'(y) = (-b_1)(+\infty) = +\infty, \quad \lim_{y \to +\infty} g'(y) = b_2(+\infty) = +\infty.
\]

For Case (10), with \( b_2 < 0 < b_1 \),

\[
\lim_{y \to -\infty} g'(y) = b_2(+\infty) = -\infty, \quad \lim_{y \to +\infty} g'(y) = (-b_1)(+\infty) = -\infty.
\]

To obtain the missing limit:

For Cases (4),(6),(9), and (11), with \( b_i = 0 \) for one \( i \), the remaining limit is \( +\infty \) according to the sign of the non-zero coefficient.

For Cases (1),(3),(7), and (12), with the \( b \)'s unequal and of like sign, we can write:

\[
g'(y) = b_2 \xi(a_2-b_2 y) \left\{ 1 - \frac{b_1 \xi(a_1-b_1 y)}{b_2 \xi(a_2-b_2 y)} \right\}.
\]
\[ \ell \lim g'(y) = b_2(\mp\infty) \left\{ 1 - \left[ \frac{\ell \lim }{b_2} \frac{\omega(a_1-b_1y)}{\ell \lim \omega(a_2-b_2y)} \right] \right\} \\
= (\pm\infty) \text{sgn}[(b_2^2-b_1^2)/b_2]. \]

For Cases (2) and (8), with \( b_1=b_2=b \), we observe that

\[ g'(y) = b\{\xi([a_1-by]-[a_1-a_2]) - \xi(a_1-by)\}. \]

\[ \ell \lim g'(y) = b(a_1-a_2), \text{ by Lemma II.10.} \]

C.) By definition, \( g(y) = \ell \ln\{\phi(a_1-b_1y)+\phi(a_2-b_2y)\}. \)

For Cases (1)-(3), (7), (8), and (12), when \( b_1 \) and \( b_2 \) are of like sign,

\[ g(y) = \ell \ln(1/1) = 0 \text{ as } a_1-b_1y \to \pm\infty, i=1,2. \]

To obtain the missing limit for each of these six cases, we observe that when \( a_1-b_1y<0 \) for \( i=1 \) and \( 2 \), we may use the definition of \( \xi(t) \) and the relationship between \( \xi(t) \) and Mills' ratio to write:

\[ g(y) = \ell \ln \left( \frac{\rho_{b_1y-a_1}}{\rho_{b_2y-a_2}} \frac{\phi_{b_1y-a_1}}{\phi_{b_2y-a_2}} \right) \]  

\[ = \ell \ln \frac{R_{b_1y-a_1}}{R_{b_2y-a_2}} - [(b_1^2-b_2^2)y^2-2(a_1b_1-a_2b_2)y+](a_1^2-a_2^2)/2, \]

so that for large values of \( b_1y-a_1 \)

\[ g(y) = \ell \ln \frac{b_2y-a_2}{b_1y-a_1} - [(b_1^2-b_2^2)y^2-2(a_1b_1-a_2b_2)y+(a_1^2-a_2^2)]/2. \]

Consequently, for Cases (1), (3), (7), and (12), when \( b_1 \) and \( b_2 \) are unequal and of like sign,

\[ g(y) \to (\pm\infty) \cdot \text{sgn}[(b_2^2-b_1^2)/2] \text{ as } a_1-b_1y \to -\infty, i=1,2. \]

And for Cases (2) and (8), where \( b_1=b_2, \)

\[ g(y) \to (\pm\infty) \cdot \text{sgn}(a_1-a_2) \text{ as } a_1-b_1y \to -\infty, i=1,2. \]

Now, for Cases (5) and (10), where one \( b \) is positive and the other is negative,

\[ g(y) = \ell \ln \phi(a_1-b_1y) - \ell \ln \phi(a_2-b_2y) \]

\[ + \pm\infty \text{ according as } a_1-b_1y \pm\infty. \]

For Cases (6) and (9), where \( b_1=0, \)

\[ g(y) = \ell \ln \phi(a_1) - \ell \ln \phi(a_2-b_2y) \]

\[ + \ell \ln \phi(a_1) \text{ or } (\pm\infty) \text{ according as } a_2-b_2y \pm\infty. \]
Similarly, for Cases (4) and (11), where \( b_2 = 0 \),
\[
g(y) = -\ln \Psi(a_2) \text{ or } (-\infty) \text{ according as } a_1 - b_1 y + \infty.
\]
\[\square\]

On the basis of Lemmas 13 and 15 we can sketch the graph of \( g(y) \) for each of the twelve cases (cf. Figure 3).

Since the shape of \( g(y) \) and hence that of \( G(y) \) depends upon the relative values of \( b_1 \) and \( b_2 \), which are functions of \( d \) and the parameters \( \sigma_i \) and \( \rho_i \), \( i = 1, 2 \), we would like to know how \( b_1 \) and \( b_2 \) vary with respect to one another as these parameters, especially \( d \), change.

\[4.4.2 \text{ Behavior of } b_1 \text{ and } b_2.\]

**Lemma IV.16:** For any given set of parameters \( \pi^i=(\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha, k) \) satisfying the Basic Assumptions, it is true that

1. \( b_1(d|\pi) \) is a strictly increasing function of \( d \), \( i = 1, 2 \);
2. \[ b_1(d|\pi) = (\sigma_1 \sqrt{1 - \rho_1^2})^{-1} \text{ and } b_2(d|\pi) > 1/\sqrt{1 - \rho_2^2} \text{ as } d \to \infty. \]
3. \( b_1(d|\pi) = 0 \text{ iff } d = -\rho_1 \sigma_2/\sigma_1 \text{, and } b_2(d|\pi) = 0 \text{ iff } d = -\rho_2 \).

**Proof:** From (34) we have
\[
b_1(d|\pi) = \frac{\sigma_2 d + \rho_1 \sigma_1 \sigma_2}{\sqrt{\Sigma|\sigma_1 |^2 d^2 + 2 \rho_1 \sigma_1 \sigma_2 d + \sigma_2^2}}, \quad \frac{\theta_1'(d|\pi)}{2\sqrt{\Sigma|\theta_1(d|\pi)}}.
\]
\[
b_2(d|\pi) = \frac{d + \rho_2}{\sqrt{\Sigma|d^2 + 2 \rho_2 d + 1}} = \frac{\theta_2'(d|\pi)}{2\sqrt{\Sigma|\theta_2(d|\pi)}},
\]

where \( \theta_1'(d|\pi) \) denotes the first partial derivative of \( \theta_1 \) with respect to \( d \), \( i = 1, 2 \).

Since \( \theta_1(d|\pi) \) is positive for all \( d \), Statement (3) follows immediately from these definitions.

Now, we can write
\[
\theta_i(d|\pi) = (t_id)^2 + 2t_i v_i d + v_i^2, \quad i = 1, 2,
\]
where \( t_1 = \sigma_1 \), \( v_1 = \sigma_2 \), \( t_2 = v_2 = 1 \) are all positive.

Then
\[
\begin{align*}
\{ & \theta_i'(d|\pi) = 2t_i(t_i d + v_i \rho_i) \\
& \theta_i''(d|\pi) = 2t_i^2 
\}
\end{align*}, \quad i = 1, 2.
\]

Let \( C_i = 1/(2t_i v_i \sqrt{1 - \rho_i^2}) > 0, \quad i = 1, 2. \)

A.) For \( i = 1, 2 \),
\[
\begin{align*}
b_1'(d|\pi) &= C_i[\theta_1 \theta_1'' - (\theta_1')^2/2] \theta_1^{-3/2} \\
&= 2C_i(t_i v_i)^2 (1 - \rho_i^2)[\theta_i(d|\pi)]^{-3/2}.
\end{align*}
\]
Case 1: $b_1 < b_2 < 0$

Case 2: $b_1 = b_2 < 0$

Case 3: $b_1 < b_2 < 0$

Case 4: $b_1 < 0 = b_2$

Case 5: $b_1 < 0 < b_2$

Case 6: $b_1 = 0 < b_2$

Case 7: $0 < b_1 < b_2$

Case 8: $0 < b_1 = b_2$

Case 9: $b_1 < 0 = b_1$

Case 10: $b_1 < 0 < b_1$

Case 11: $b_1 = 0 < b_1$

Case 12: $0 < b_2 < b_1$

Figure IV.3: The 12 Basic Graphs of Function $g(y)$
which is positive for all \( d \), since \( \theta_i > 0 \) for all \( d \).

Therefore, \( b_i(d|\pi) \) is a strictly increasing function of \( d \), \( i=1,2 \).

B.) For \( b_i(\pi) \neq 0 \) we can also write

\[
b_i(d|\pi) = \frac{2tiCi(tid + v_i\pi)}{\sqrt{(tid)^2 + 2tivipid + v_i^2}}
\]

\[
= \frac{\text{sgn } \theta_i'(d|\pi)(1 + \frac{v_i^2(1-\rho_i^2)}{(tid + v_i\pi)^2})^{-1/2}}{v_i\sqrt{1-\rho_i^2}}
\]

which tends to \( \pm (v_i\sqrt{1-\rho_i^2})^{-1} \) as \( d \to \pm \infty \).

\[\square\]

To see how \( b_1(d|\pi) \) and \( b_2(d|\pi) \) vary with respect to each other as \( d \) increases from \( -\infty \) to \( +\infty \), we first studied the simplest case possible, namely, \( \sigma_1 = \sigma_2 = 1 \) and \( \rho_1 \neq \rho_2 \). As this proved to be quite messy, we did not attempt more general cases.

\textit{Lemma IV.17:} For any set of parameters \( \pi'=(\delta_1,\delta_2,\sigma_1,\sigma_2,\rho_1,\rho_2,a,k) \) satisfying the Basic Assumptions and the condition that \( \sigma_1 = \sigma_2 = 1 \), the relative values of \( b_1(d|\pi) \) and \( b_2(d|\pi) \) are given in Table 9. The values in the interval \((-1,1)\) of \( \rho' \) and \( \rho'' \), defined by

\[
((\rho'',\rho')) = (\rho_2 \pm 2\sqrt{3 - 2\rho_2^2})/3
\]

are given in Table 8 for selected values of \( \rho_2 \).

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\( \rho_2 \) & \( \rho'' \) & \( \rho' \) \\
\hline
-1.00 & 0.3333 & 0.333 & -1.0000 \\
-0.95 & 0.4121 & 0.35 & -0.9899 \\
-0.90 & 0.4832 & 0.40 & -0.9580 \\
-0.85 & 0.5480 & 0.45 & -0.9239 \\
-0.80 & 0.6077 & 0.50 & -0.8874 \\
-0.75 & 0.6629 & 0.55 & -0.8484 \\
-1/\sqrt{2} & 1/\sqrt{2} & 0.60 & -0.8066 \\
-0.70 & 0.7142 & 0.65 & -0.7620 \\
-0.65 & 0.7620 & 0.70 & -0.7142 \\
-0.60 & 0.8066 & 1/\sqrt{2} & -1/\sqrt{2} \\
-0.55 & 0.8484 & 0.75 & -0.6629 \\
-0.50 & 0.8874 & 0.80 & -0.6077 \\
-0.45 & 0.9239 & 0.85 & -0.5480 \\
-0.40 & 0.9580 & 0.90 & -0.4832 \\
-0.35 & 0.9899 & 0.95 & -0.4121 \\
-0.333 & 1.0000 & 1.00 & -0.3333 \\
\hline
\end{tabular}
\end{table}
<table>
<thead>
<tr>
<th>$b_1$</th>
<th>$b_2$</th>
<th>Values of $d$ for which $b_1 = b_2$</th>
<th>Range of $d$ for which $b_1 &gt; b_2$</th>
<th>Range of $d$ for which $b_1 &lt; b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty, -1/2)</td>
<td>(-\infty, -b_2)</td>
<td>$d^* &gt; 1$</td>
<td>((-\infty, -d^*))</td>
<td>((-\infty, -d))</td>
</tr>
<tr>
<td>(-\infty, -b_2)</td>
<td>(-\infty, -b_2)</td>
<td>$d^* = 1$</td>
<td>((-\infty, -d^*))</td>
<td>((-\infty, -d))</td>
</tr>
<tr>
<td>(-1 &lt; b_1 &lt; -b_2) and possibly $d^*$ and $d$ satisfying one of conditions (1)-(3) below.</td>
<td>((-\infty, -d^*)) if (1) holds; ((-\infty, -d)) otherwise.</td>
<td>((-\infty, -d^*)) if (2) holds; ((-\infty, -d)) otherwise.</td>
<td>((-\infty, -d^*)) if (3) holds; ((-\infty, -d)) otherwise.</td>
<td>((-\infty, -d^*)) if (3) holds; ((-\infty, -d)) otherwise.</td>
</tr>
</tbody>
</table>

**LIST OF CONDITIONS**
1. $d^* < d^* = \min(-1, -d^*)$.
2. $\max(-2, -b_2) < d^* < d^* = 1$.
3. $d^* < d^* < d^* = -1$.
proof: When $\sigma_1=\sigma_2=1$,
\[ b_1^2(d|\pi)-b_2^2(d|\pi) = \frac{(d+\rho_1)^2}{(1-\rho_1^2)\theta_1(d|\pi)} - \frac{(d+\rho_2)^2}{(1-\rho_2^2)\theta_2(d|\pi)} \leq 0 \]
according as
\[ (1-\rho_2^2)(d+\rho_1)^2\theta_2(d|\pi) - (1-\rho_1^2)(d+\rho_2)^2\theta_1(d|\pi) \geq 0. \]
Now, the L.H.S. of this, after substituting the appropriate quadratic for $\theta_1$ and $\theta_2$, collecting terms, and simplifying, can be written
\[ (\rho_1^2-\rho_2^2)d^4 + 2(\rho_1^2-\rho_2^2)(\rho_1+\rho_2)d^3 + 2(\rho_1^2-\rho_2^2)(1+2\rho_1\rho_2)d^2 + \]
\[ + 2(\rho_1-\rho_2)(1+\rho_1\rho_2)^2d + (\rho_1^2-\rho_2^2). \]
(35)

Case 1: $\rho_1 = -\rho_2 (\neq 0)$.

(35) equals $4\rho_1 d$ in this case. Omitting the argument $(d|\pi)$ for the sake of simplicity, we have, then,
\[ b_1^2-b_2^2 \geq 0 \iff \rho_1 d \geq 0. \]
(36)

Lemma 16 and statement (36) together imply:

1.) If $\rho_1<0$, then
   a.) $b_1^2\geq b_2^2 \iff d\leq 0$;
   b.) $b_1\leq 0 \iff d\leq|\rho_1|$;
   c.) $b_2\leq 0 \iff d\leq\rho_1$.

Consequently,
   i.) $b_1>b_2\leq 0$ for all $d\leq\rho_1$;
   ii.) $b_1\leq 0<b_2$ for all $d \in (\rho_1,|\rho_1|]$;
   iii.) $0>b_1<b_2$ for all $d \in (|\rho_1|,\infty)$;
   iv.) $b_1<b_2$ for all real $d$.

2.) If $\rho_1>0$, then by analogous reasoning we have $b_1>b_2$ for all real $d$.

Case 2: $\rho_1^2\neq \rho_2^2$.

The quantity $\rho_1^2-\rho_2^2$ can be factored out front in (35), so that
\[ b_1^2\geq b_2^2 \iff (\rho_1^2-\rho_2^2)H(d|\pi) \geq 0, \]
(37)
where
\[ H(d|\pi) = d^4 + 2(\rho_1+\rho_2)d^3 + 2(1+2\rho_1\rho_2)d^2 + \]
\[ + 2[(1+\rho_1\rho_2)^2/(\rho_1+\rho_2)]d + 1, \]
(38)
$\equiv H(d)$ for short.

[1.] $H(0)=1$ and $H(d) \to +\infty$ as $d \to +\infty$ regardless of the values of $\rho_1$ and $\rho_2$. 
[2.] One zero of $H(d)$ lies between $-\rho_1$ and $-\rho_2$. However, for this value of $d$ we have $b_1 = -b_2$, a case of no special interest to us (cf. the twelve cases determining the shape of $g(y)$).

**proof:** From (37) we have

\[
d = -\rho_1 \text{ implies } b_1^2 = 0 < b_2^2 \\
\text{implies } (\rho_1^2 - \rho_2^2)H(-\rho_1) < 0; \\
d = -\rho_2 \text{ implies } b_1^2 = 0 = b_2^2 \\
\text{implies } (\rho_1^2 - \rho_2^2)H(-\rho_2) > 0.
\]

It follows from Lemma 16-(1),(3) that $b_1 = -b_2$.

[3.] If $\rho_1 + \rho_2 > 0$, then $H(-1) < 0 < H(+1)$; if $\rho_1 + \rho_2 < 0$, then $H(+1) < 0 < H(-1)$. Thus, there exists one zero, $d^\#$, of $H(d)$ such that

\[
\begin{cases} 
  d^\# < -1 & \text{if } \rho_1 + \rho_2 > 0 \\
  d^\# > +1 & \text{if } \rho_1 + \rho_2 < 0 
\end{cases}
\]

For $d^\#$, then, $b_1 = b_2$ and is positive or negative according as $\rho_1 + \rho_2$ is negative or positive.

**proof:** By (38) $H(+1) = 2(l + \rho_1)^2(l + \rho_2)^2/(\rho_1 + \rho_2)$

and $H(-1) = -2(l - \rho_1)^2(l - \rho_2)^2/(\rho_1 + \rho_2)$.

It follows from Lemma 16-(1),(3) that $b_1 = b_2$ when $d = d^\#$ and that $b_1(d^\#) < 0$ when $d^\# < -1$, whereas $b_1(d^\#) > 0$ when $d^\# > +1$, $i=1,2$.

To learn more about the behavior of $H(d)$, consider its derivatives:

\[
H'(d) = 4d^3 + 6(\rho_1 + \rho_2)d^2 + 4(l + 2\rho_1 \rho_2)d + 2(1 + \rho_1 \rho_2)^2/(\rho_1 + \rho_2); \\
H''(d) = 12d^2 + 12(\rho_1 + \rho_2)d + 4(l + 2\rho_1 \rho_2).
\]

[4.] $H''(0) \leq 0$ iff $\rho_1 \rho_2 \leq -1/2$.

[5.] $H''(d) \rightarrow +\infty$ as $d \rightarrow +\infty$ regardless of the values of $\rho_1$ and $\rho_2$.

[6.] When $\rho_1$ and $\rho_2$ are values such that $H''(d) \geq 0$ for all $d$, then $H'(d)$ is a strictly increasing function of $d$. This fact, together with [1], implies that $H(d)$ has a unique minimum; and since $H(d) < 0$ for $d = l$ or $d = -l$, there must be exactly two real zeros of $H(d)$. These two are given in [2] and [3].

[7.] $H''(d) \geq 0$ for all $d$ iff one of the following sets of conditions holds:
a.) \(-1 < \rho_2 < -1/3\) and \(-1 < \rho_1 < \rho'' < 1\);
b.) \(-1/3 \leq \rho_2 \leq 1/3\);
c.) \(1/3 < \rho_2 < 1\) and \(-1 < \rho' \leq \rho_1 < 1\).

**proof:** The zeros of \(H''(d)\) are
\[
-(\rho_1 + \rho_2) \pm \sqrt{[3\rho_1^2 - 2\rho_1 \rho_2 - (4 - 3\rho_2^2)]/3} \div 2,
\]
which are imaginary iff the quantity under the radical is negative, i.e. iff \((\rho'', \rho')\).

Now, \(\rho' \leq -1\) iff \(-1 \leq \rho_2 \leq 1/3\), and \(\rho'' \geq +1\) iff \(-1/3 \leq \rho_2 \leq 1\).

[8.] In some cases we need to know how \(\rho'\) and \(\rho''\) compare in size with \(-\rho_2\):

a.) \(\rho'' \leq -\rho_2\) iff \(-1 < \rho_2 \leq -1/\sqrt{2} \approx -0.707;\)
b.) \(\rho' \geq -\rho_2\) iff \(1/\sqrt{2} \leq \rho_2 < 1\).

**proof:** \(\rho'' \leq -\rho_2\) iff \(-\rho_2 = |\rho_2|\) and \(2\sqrt{3 - 2\rho_2^2} \leq 4|\rho_2|\)
iff \(-1 < \rho_2 \leq -1/\sqrt{2}\).

\(\rho' \geq -\rho_2\) iff \(\rho_2 > 0\) and \(-4\rho_2 \leq -2\sqrt{3 - 2\rho_2^2}\)
iff \(1/\sqrt{2} \leq \rho_2 < 1\).

On the basis of statements [3] and [6]-[8] the relative values of \(b_1\) and \(b_2\) can be determined for all real \(d\) and for all possible choices of the parameters \(\rho_1\) and \(\rho_2\) except when \(\rho_1 \rho_2\) is "near" \(-1\). It appears that the only way to determine analytically whether \(H(d)\) can have four real zeros is to solve for them, a nontrivial task, in view of the complexity of the coefficients. However, if \(H(d)\) does have four real zeros, we have the following information about the extra two zeros, \(d'\) and \(d''\), say:

[9.] They can occur only when \(\rho_1 \rho_2\) is "near" \(-1\); specifically,
(from [6] and [7]):
a.) \(-1 < \rho_2 < -1/3\) and \(\rho'' < \rho_1 < 1\); or
b.) \(1/3 < \rho_2 < 1\) and \(-1 < \rho_1 < \rho'\).

[10.] Both must be located in exactly one of the intervals:
a.) \((-\infty, \min(d^*, -1))\)
b.) \((-1, \min(-\rho_1, -\rho_2))\)
c.) \((\min(-\rho_1, -\rho_2), \max(-\rho_1, -\rho_2))\)
d.) \((\max(-\rho_1, -\rho_2), +1)\)
e.) \((\max(d^*, +1), +\infty)\).
proof: Since \( H(-\rho_1) \) and \( H(-\rho_2) \) are opposite in sign, there must be an odd number of zeros between them. Since, by \([2] \) and \([3] \), \( H(-1) \) and \( H(\min\{-\rho_1,-\rho_2\}) \) have the same sign and \( H(+1) \) and \( H(\max\{-\rho_1,-\rho_2\}) \) have the same sign, there must be an even number of zeros in each of the intervals between them. Consequently, \(|d'|<1 \) iff \(|d''|<1 \); and \(d' \) and \(d'' \) both occur in one of the three middle intervals. If \( H(-1)<0<H(+1) \), there must be an odd number of zeros less than \(-1 \) and an even number of zeros exceeding \(+1 \). Similarly, if \( H(+1)<0<H(-1) \), there must be an odd number of zeros exceeding \(+1 \) and an even number of zeros less than \(-1 \).

If there are three zeros less than \(-1 \), we define \(d^* \) to be the one closest to \(-1 \); and if there are three zeros exceeding \(+1 \), we define \(d^* \) to be the one closest to \(+1 \).

Consequently, \(|d'|>1 \) iff \(|d''|>1 \), and both \(d' \) and \(d'' \) are in one of the intervals, (\(-\infty, \min\{d^*,-1\}\)) or (\(\max\{d^*,+1\}, +\infty\)).

[11.] They are of no interest if they lie in the middle interval, since \(b_1 \) and \(b_2 \) are of opposite sign for all \(d \) between \(-\rho_1 \) and \(-\rho_2 \).

\[
\begin{array}{|c|c|c|}
\hline
\text{For } & \text{Relative Values of } b_1 \text{ and } b_2 & \text{Shape of } g(y) \\
\hline
-\infty < d < d^* < -1 & b_2 < b_1 < 0 & \text{Case 1} \\
d = d^* & b_1 = b_2 < 0 & \text{Case 2} \\
d^* < d < -\rho_2 & b_1 < b_2 < 0 & \text{Case 3} \\
d = -\rho_2 & b_1 < 0 = b_2 & \text{Case 4} \\
-\rho_2 < d < -\rho_1 & b_1 < 0 < b_2 & \text{Case 5} \\
d = -\rho_1 & b_1 = 0 < b_2 & \text{Case 6} \\
-\rho_1 < d < +\infty & 0 < b_1 < b_2 & \text{Case 7} \\
\hline
\end{array}
\]

We can use Table 9 and Figure 3 to see how \(g(y)\) changes shape as \(d\) increases, given any parameter set \(\pi'\) with \(\sigma_1=\sigma_2=1\). For example, suppose \(0 < \rho_2 \leq 1/\sqrt{2} \) and \(-\rho_2 < \rho_1 < \rho_2 \).
4.4.3 Behavior of \( a_1 \) and \( a_2 \) when \( k=0 \).

When \( k=0 \), we have by definition (34)

\[
\begin{align*}
\begin{cases}
a_1(d|\pi) &= -\frac{\delta_1(\delta_2 \sigma_1 - \delta_1 \rho_1 \sigma_2) + \sigma_2(\delta_2 \rho_1 \sigma_1 - \delta_1 \rho_1 \sigma_2)}{\sqrt{\prod_1(\sigma_1^2 d^2 + 2 \rho_1 \sigma_1 \sigma_2 d + \sigma_2^2)}} \\
a_2(d|\pi) &= 0 \text{ for all } d
\end{cases}
\end{align*}
\]

(39)

**Lemma IV.18:** Let \( \pi \) be any parameter set satisfying the Basic Assumptions and the condition \( k=0 \). Then the following statements hold:

1.) If \( \delta = 0 \), then \( a_1 = a_2 = 0 \) for all \( d \);

2.) If \( \delta_1 = 0 < \delta_2 \), then \( a_1 = \delta_2 b_1 \), which is a strictly increasing function of \( d \) such that \( a_1 = \pm(\delta_2/\sigma_2)\sqrt{1-\rho_1^2} \) as \( d \to \pm\infty \), and \( a_1 = 0 = a_2 \) if \( \rho_1 = 0 \).

3.) If \( \delta_1 > 0 \) and \( (\delta_2/\sigma_2) = (\rho_1 \delta_1/\sigma_1) \), then

\[
a_1 = -\sigma_2(\delta_1/\sigma_1)\sqrt{(1-\rho_1^2)/\theta_1(d|\pi)}
\]

which is strictly negative for all \( d \), has a minimum value of \( -\delta_1/\sigma_1 \) when \( d = -\delta_2/\delta_1 \), and tends to zero as \( d \to \pm\infty \). Therefore, \( a_1 < a_2 \) for all \( d \).

[Note that this case cannot occur if \( \rho_1 < 0 \).]

4.) If \( \delta_1 > 0 \) and \( (\delta_2/\sigma_2) \neq (\rho_1 \delta_1/\sigma_1) \), then \( a_1 \) has a minimum value of

\[
-(\delta_1/\sigma_1)(1 + [\rho_1 - (\delta_2/\sigma_2)/(\sigma_2 \delta_1)]^2(1-\rho_1^2)^{-1})^{1/2}
\]

when \( d = -\delta_2/\delta_1 \) and approaches \( \pm[(\delta_2/\sigma_2) - \rho_1(\delta_1/\sigma_1)]\sqrt{1-\rho_1^2} \) as \( d \to \pm\infty \). Furthermore, \( a_1 = 0 = a_2 \) iff

\[
d = (\sigma_2/\sigma_1)((\delta_1/\sigma_1) - \rho_1(\delta_2/\sigma_2)) \pm[(\delta_2/\sigma_2) - \rho_1(\delta_1/\sigma_1)],
\]

which is positive iff \( -1 < \rho_1 < \min\{(\delta_1 \sigma_1)/(\sigma_1 \delta_2), (\delta_2 \sigma_1)/(\sigma_2 \delta_1)\} \).

**Proof:** Statement (1) follows directly from (39).

Statement (2) follows directly from (39), (34), and Lemma 16.

A.) If \( \delta_1 > 0 \) and \( (\delta_2/\sigma_2) = (\rho_1(\delta_1/\sigma_1)) \), then from (39) we have

\[
a_1 = -\sigma_2(\delta_1/\sigma_1)\sqrt{1-\rho_1^2}/\theta_1(d|\pi)
\]

which is strictly negative for all \( d \) and has a unique minimum at \( d = -\sigma_1(\sigma_2/\sigma_1) \), which is the value minimizing \( \theta_1(d|\pi) \).

By assumption, though, \( \rho_1(\sigma_2/\sigma_1) = (\delta_2/\delta_1) \). [Note that this equation cannot hold if \( \rho_1 < 0 \), since a Basic Assumption is that \( \delta_i \geq 0, i=1,2 \).]

Thus, \( \min_{d} a_1(d|\pi) = a_1(-\delta_2/\delta_1|\pi) = -(\delta_1/\sigma_1) \).

Because \( \theta_1(d|\pi) \to +\infty \) as \( d \to \pm\infty \), \( a_1(d|\pi) \to 0 \) as \( d \to \pm\infty \).
B.) If \( \delta_1 > 0 \) and \( (\delta_2 / \sigma_2) = \rho_1 (\delta_1 / \sigma_1) \), then \( a_1 (d|\pi) \) may be written

\[
a_1 (d|\pi) = \frac{A^* d + B^*}{\sqrt{\sum_{1} \sqrt{\theta_1 (d|\pi)}}}
\]

with

\[
A^* = \sigma_1^2 \delta_2 - \rho_1 \sigma_1 \delta_2 \delta_1 \equiv s_1 \delta_2 - s_2 \delta_1 \\
B^* = \rho_1 \sigma_1 \sigma_2 \delta_2 - \sigma_2^2 \delta_1 \equiv s_2 \delta_2 - s_3 \delta_1
\]

so that the derivative with respect to \( d \) is

\[
a_1' (d|\pi) = \frac{A^* (s_1 d^2 + 2 s_2 d + s_3) - (A^* d + B^*) (s_1 d + s_2)}{\sqrt{\sum_{1} \sqrt{\theta_1 (d|\pi)}}^3}
\]

\[
= \frac{\sqrt{\sum_{1}} (s_1 \delta_1 + s_1 \delta_2)}{\left(\sqrt{\theta_1 (d|\pi)}\right)^3} \leq 0 \quad \text{according as} \quad d \leq -\delta_2 / \delta_1,
\]

since \( \theta_1 (d|\pi) > 0 \) for all \( d \).

Therefore, \( a_1 \) has a unique minimum at \( d = -\delta_2 / \delta_1 \), and this minimum value is

\[
a_1 (-\delta_2 / \delta_1 | \pi) = \frac{-s_1 \delta_2^2 + 2 s_2 \delta_1 \delta_2 - s_2 \delta_1^2}{\sqrt{\sum_{1} \sqrt{s_1 \delta_2^2 - 2 s_2 \delta_1 \delta_2 + s_3 \delta_1^2}}}
\]

\[
= -\left(\delta_1 / \sigma_1\right) \left[ 1 + \frac{\left(\frac{\delta_2 / \sigma_2}{\delta_1 / \sigma_1}\right)^2}{1 - \rho_1^2} \right]^{1/2}
\]

To find limits as \( d \to \pm \infty \), write

\[
a_1 (d|\pi) = \frac{\delta_2 - \rho_1 \delta_1 (\text{sgn } d) \left[ 1 - \sigma_2 \left(\frac{\delta_1 / \sigma_1 - \delta_1 (\delta_2 / \sigma_2)}{\delta_1 (\delta_2 / \sigma_2) - \delta_1 (\delta_1 / \sigma_1)}\right) d^{-1}\right]}{\sqrt{1 - \rho_1^2}} \left\{ \left[ (\delta_2 / \sigma_2) - \rho_1 (\delta_1 / \sigma_1) \right] \sqrt{1 + 2 \rho_1 (\sigma_2 / \sigma_1) d^{-1} + (\sigma_2 / \sigma_1)^2 d^{-2}} \right\}
\]

which tends to \( [(\delta_2 / \sigma_2) - \rho_1 (\delta_1 / \sigma_1)] / \sqrt{1 - \rho_1^2} \) as \( d \to \pm \infty \).

Also, \( a_1 = 0 \) iff

\[
d = -B^* / A^* = \frac{\sigma_2 [\delta_1 / \sigma_1 - \rho_1 (\delta_2 / \sigma_2)]}{\sigma_1 [\delta_2 / \sigma_2 - \rho_1 (\delta_1 / \sigma_1)]} \equiv d_0.
\]

When \( 0 < (\delta_2 / \sigma_2) < \rho_1 (\delta_1 / \sigma_1) \) then

\( d_0 > 0 \) implies \( (\delta_1 / \sigma_1) < \rho_1 (\delta_2 / \sigma_2) \), which is a contradiction.

When \( (\delta_2 / \sigma_2) > \rho_1 (\delta_1 / \sigma_1) \), i.e. when \( \rho_1 < (\delta_2 / \sigma_2) \), then

\( d_0 > 0 \) iff \( (\delta_1 / \sigma_1) > \rho_1 (\delta_2 / \sigma_2) \), i.e. \( \rho_1 < (\delta_1 / \sigma_1) \).

Hence the result. \( \square \)
4.5 Behavior of the Decision Function $U(y)$.

Recall, from Corollary 1, that the optimal decision rule based on the linear function $Y = dX_1 + X_2$ is to choose Population 1 iff $U(y) < 0$, where $U(y) = Q(y) + G(y)$, the sum of the functions studied in the previous two sections.

**Lemma IV.19:** Let $\pi' = (\delta_1, \delta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha, k)$ be any parameter set satisfying the Basic Assumptions, and let $d$ be any real number. (1) $U(y)$ depends on $\alpha$ only through $B$, the constant term of the quadratic $Q(y)$. (2) If $d$ and all other parameters are held fixed, then at every point $y$, $U(y)$ decreases as $\alpha$ increases.

**Proof:** Statement (1) follows from Lemmas 9 and 12. Statement (2) follows from Lemma 10 and the definition of $U(y)$. \[ \square \]

**Theorem IV.3:** Let $\pi$ be any parameter set satisfying the Basic Assumptions, and let $d$ be any real number. The limiting values of the function $U(y)$ as $y \to \pm \infty$ are given in Table 10, where $C_i = C_i(d|\pi)$, $i=1,4,$

\[
\begin{align*}
C_1 &= \theta_2 - \sigma_2^2(1-\rho_1^2) = \min_{d} \theta_2(d|\pi) - \min_{d} \theta_1(d|\pi) = D + \theta_1 \theta_2 b_1^2 \\
C_2 &= (1-\rho_2^2) - \theta_1 = \min_{d} \theta_2(d|\pi) - \min_{d} \theta_1(d|\pi) = D - \theta_1 \theta_2 b_2^2 \\
C_3 &= (1-\rho_2^2) - \sigma_2^2(1-\rho_1^2) = \min_{d} \theta_2(d|\pi) - \min_{d} \theta_1(d|\pi) = D + \theta_1 \theta_2(b_1^2 - b_2^2) \\
C_4 &= A + b62(a_1-a_2), \text{ where } b_1=b_2=b \text{ and } \theta_1=\theta_2=0
\end{align*}
\]

and the remaining parametric functions are defined in (1) and (5).

Note that
\[
\begin{align*}
C_1 &= 0 \iff \rho_2^2 \geq 1-\sigma_2^2(1-\rho_1^2) \text{ and } d = -\rho_2 \pm \sqrt{\sigma_2^2(1-\rho_1^2) - (1-\rho_2^2)}; \\
C_2 &= 0 \iff \rho_2^2 \leq 1-\sigma_2^2(1-\rho_1^2) \text{ and } d = -\rho_2 \pm \sqrt{(1-\rho_2^2)^2 - \sigma_2^2(1-\rho_1^2)} / \sigma_1; \\
C_3 &= 0 \iff \rho_2^2 = 1-\sigma_2^2(1-\rho_1^2)
\end{align*}
\]

**Proof:** Since $Q(y) = Dy^2 - 2Ay + B$, we can construct the following table:

<table>
<thead>
<tr>
<th>D</th>
<th>A</th>
<th>$\lim_{y \to -\infty} Q(y)$</th>
<th>$\lim_{y \to +\infty} Q(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D &lt; 0$</td>
<td>$-$</td>
<td>$-\infty$</td>
<td>$-\infty$</td>
</tr>
<tr>
<td>$D &gt; 0$</td>
<td>$+$</td>
<td>$+\infty$</td>
<td>$+\infty$</td>
</tr>
<tr>
<td>$D = 0$</td>
<td>$A &lt; 0$</td>
<td>$-\infty$</td>
<td>$+\infty$</td>
</tr>
<tr>
<td></td>
<td>$A = 0$</td>
<td>$B$</td>
<td>$B$</td>
</tr>
<tr>
<td></td>
<td>$A &gt; 0$</td>
<td>$+\infty$</td>
<td>$-\infty$</td>
</tr>
</tbody>
</table>

TABLE IV.11
<table>
<thead>
<tr>
<th>Case (b₁,b₂)</th>
<th>D</th>
<th>A</th>
<th>( \lim \limits_{y \to -} U )</th>
<th>( \lim \limits_{y \to +} U )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-a</td>
<td>b₂ &lt; b₁ &lt; 0</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D = 0</td>
<td>A &gt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>2-a</td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>3-a</td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>4-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>5-a</td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>6-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>7-a</td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>8-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>9-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>10-a</td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>11-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td>12-a</td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td>13-a</td>
<td>b₁ = b₂</td>
<td>D = 0</td>
<td>A = 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &lt; b₂</td>
<td>D &lt; 0</td>
<td>A &lt; 0</td>
<td>( C_1 \neq 0 )</td>
</tr>
<tr>
<td></td>
<td>b₁ &gt; b₂</td>
<td>D &gt; 0</td>
<td>A &gt; 0</td>
<td>( (+) \cdot C_1 )</td>
</tr>
</tbody>
</table>
Since \( G(y) = -2\theta_1 \theta_2 g(y) \), a negative multiple of \( g(y) \), we have from Lemma 15:

**TABLE IV.12**

<table>
<thead>
<tr>
<th>((b_1, b_2))</th>
<th>((a_1, a_2))</th>
<th>(\lim_{y \to -\infty} G(y))</th>
<th>(\lim_{y \to +\infty} G(y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_1 &gt; b_2 &lt; 0)</td>
<td>(-\infty)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(a_1 &lt; a_2)</td>
<td>(\infty)</td>
<td>(-\infty)</td>
<td>0</td>
</tr>
<tr>
<td>(a_1 = a_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(a_1 &gt; a_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>0</td>
</tr>
<tr>
<td>(b_1 &lt; 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>2(\theta_1 \theta_2 \ln (a_2))</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 &lt; 0 = b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 &lt; 0 = b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 &lt; 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 &lt; 0 = b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 &lt; 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
</tbody>
</table>

The limits of \( U(y) \) follow directly from the entries in Tables 11 and 12, except for the "difficult" cases (marked by an asterisk) when one of the functions, \( Q(y) \), \( G(y) \), approaches \( \infty \) and the other tends to \( -\infty \). We can determine \( \lim U(y) \) for these cases in various ways.

1.) If \( D \neq 0 \), then we can use L'Hôpital's Rule to obtain the missing limit:

\[
\lim U(y) = \lim \left\{ y^2 \left[ D - \frac{2A}{y} + \frac{B}{y^2} - 2\theta_1 \theta_2 \frac{g(y)}{y^2} \right] \right\}
\]

\[
= \left\{ \begin{array}{ll}
(\infty) & \left[D - 2\theta_1 \theta_2 \lim [g(y)/y^2]\right] \\
(\infty) & \left[D - 2\theta_1 \theta_2 \lim g''(y)\right] \text{ if } |\lim g'(y)| = \infty \\
(\infty) & \left(\infty\right) \cdot D \text{ if } |\lim g'(y)| < \infty
\end{array} \right. \quad (41)
\]

From Lemma 15 we can distinguish the following cases:

A.) \( 0 < |\lim g'(y)| < \infty \) iff \( b_1 = b_2 \).

Therefore, if \( b_1 = b_2 \) and \( D \neq 0 \), then \( \lim U(y) = D \cdot (\infty) \).

From this equation we can calculate the missing limit for cases (a) and (b) of (2) and (8).

B.) \( |\lim g'(y)| = \infty \) and \( \lim g''(y) = b_2^2 - b_1^2 \) iff \( b_1 b_2 > 0 \) and \( b_1 \neq b_2 \).

By (41), then, the missing limit can be found for case (a) or
(b) of (1), (3), (7), and (12):

\[ \lim U(y) = (+\infty)[D + \theta_1 \theta_2 (b_1 + b_2)] = (+\infty)C_3 \]

because by definition

\[ D + \theta_1 \theta_2 (b_1 + b_2) = (\theta_2 - \theta_1) + \theta_1 \theta_2 \left[ \frac{\sigma_2 (\sigma_1 + \rho \sigma_2)}{\sigma_1 \sigma_2^2 (1 - \rho_1^2)} - \frac{(d + \rho_2)^2}{\sigma_2 (1 - \rho_2^2)} \right] \]

\[ = \theta_1 \theta_2 \left[ (\sigma_2^2 (1 - \rho_1^2))^{-1} - (1 - \rho_2^2)^{-1} \right] \]

\[ = \alpha C_3 \equiv (1 - \rho_2^2) - \sigma_2^2 (1 - \rho_1^2). \]

Note that \( C_3 = \min \frac{\theta_2 (d | \pi)}{d} - \min \frac{\theta_1 (d | \pi)}{d} \), which is independent of \( d \).

C.) \[ \lim g'(y) = +\infty \text{ and } \lim g''(y) = -b_1^2 \] only if \( b_1 \neq 0 \) and \( b_1 b_2 \leq 0 \).

By (h1) and Table 7, the missing limit for case (a) of (4), (5), (10), and (11) is:

\[ \lim U(y) = (+\infty)[D + \theta_1 \theta_2 b_1^2] = (+\infty)C_1 \]

because by definition

\[ D + \theta_1 \theta_2 b_1^2 = (\theta_2 - \theta_1) + \theta_1 \theta_2 \left[ \frac{\sigma_2 (\sigma_1 + \rho \sigma_2)}{\sigma_1 \sigma_2^2 (1 - \rho_1^2)} \right] \]

\[ = \theta_1 \left[ \frac{\sigma_2}{\sigma_2^2 (1 - \rho_1^2)} - 1 \right] \]

\[ = \alpha C_1 \equiv \theta_2 - \sigma_2^2 (1 - \rho_1^2) = \theta_2 (d | \pi) - \min \frac{\theta_1 (d | \pi)}{d}. \]

D.) \[ \lim g'(y) = +\infty \text{ and } \lim g''(y) = b_2^2 \] only if \( b_2 \neq 0 \) and \( b_1 b_2 \leq 0 \).

So the missing limit for case (b) of (5), (6), (9), and (10) is

\[ \lim U(y) = (+\infty)[D - \theta_1 \theta_2 b_2^2] = (+\infty)C_2 \]

because by definition

\[ D - \theta_1 \theta_2 b_2^2 = (\theta_2 - \theta_1) - \theta_1 \theta_2 \left[ \frac{(d + \rho_2)^2}{\sigma_2 (1 - \rho_2^2)} \right] \]

\[ = \theta_2 \left[ 1 - \frac{\theta_1}{1 - \rho_2^2} \right] \]

\[ = \alpha C_2 \equiv (1 - \rho_2^2) - \theta_1 = \min \frac{\theta_2 (d | \pi)}{d} - \theta_1 (d | \pi). \]

2.) If \( D = 0 \) and \( A \neq 0 \), we can apply L'Hôpital's Rule to get:

\[ \lim U(y) = \lim \left\{ y \left[ -2A + \frac{B}{y} - 2\theta_1 \theta_2 g(y) \right] \right\} \]

\[ = -(\lim y) \{ A + \theta_2 \lim [g(y)/y] \} \text{ since } \theta_1 = \theta_2 \text{ when } D = 0. \]

\[ = -(\lim y) \{ A + \theta_2 \lim g'(y) \}. \]  \hspace{1cm} (42)

A.) If \( |\lim g'(y)| = \infty \), then by (42)
\[ \begin{align*}
\left\{ \begin{array}{l}
\lim_{y \to -\infty} U(y) = \lim_{y \to -\infty} g'(y) \\
\lim_{y \to +\infty} U(y) = -\lim_{y \to +\infty} g'(y)
\end{array} \right. 
\end{align*} \]

From this pair of equations and Lemma 15 we obtain all the missing values except those for cases (c) and (i) of (2) and (8).

B.) When \( b_1=b_2=b \), we have from (42) and Lemma 15
\[ \lim_{y \to -\infty} U(y) = -(\lim_{y \to -\infty} y)[A + \theta^2(a_1-a_2)] = -(\lim_{y \to -\infty} y) \cdot C_4. \]

The conditions under which \( C_i \) vanishes, \( i=1-3 \), follow immediately from (40) and the definition of \( \theta_1 \), \( i=1,2 \).

To get an idea of the way in which the shape of \( U(y) \) changes as \( d \) changes, it is helpful to consider a special case of the parameters \( \pi \).

4.5.1 Behavior of \( U(y) \) When \( \delta'=(1,0) \), \( a_1=a_2=1 \), \( \delta_1 = -2/3 \), \( \delta_2 = +1/3 \), \( a=1/2 \), and \( k=0 \).

For this special case, the problem is to identify an individual selected from a 50-50 mixture of two populations with the distributions sketched in Figure 4. [For the computations of these and similar contours, see the Appendix.]

The shape of \( U(y) \) depends on the shapes of \( Q(y) \) and \( G(y) \), which, in turn, depend on the values of the functions:

\[
\left\{ \begin{array}{l}
\theta_1(d|\pi) = d^2 + 2(-2/3)d + 1 \\
\theta_2(d|\pi) = d^2 + 2(1/3)d + 1 \\
D(d|\pi) = \theta_2-\theta_1 = 2d \\
b_1(d|\pi) = (d - 2/3)[(\sqrt{5}/3)\theta_1] = (3d-2)/\sqrt{5} \theta_1 \\
b_2(d|\pi) = (d + 1/3)[(2\sqrt{2}/3)\theta_2] = (3d+1)/\sqrt{5} \theta_2 \\
a_1(d|\pi) = [d(0 + 2/3) + (0-1)][(\sqrt{5}/3)\theta_1] = (2d-3)/\sqrt{5} \theta_1 \\
a_2(d|\pi) = 0 \\
A(d|\pi) = d\theta_2 \\
B(d|\pi) = \theta_2[d^2 - 2\theta_1 \ln(\phi(0)/\phi(-1)) - \theta_1 \ln(\theta_2/\theta_1)]
\end{array} \right. \] (43)

In order to find the \( d \)-intervals for which each of the cases in Table 10 applies, we must know the \( d \)-intervals for which the functions, \( D,A,b_1,b_2, a_1, \) and \( a_2 \) are positive and negative. From (43) we can easily ascertain:
1.) \( D \leq 0 \) according as \( d \leq 0 \).

2.) There exists a number \( d^* > 1 \) such that \( b_1 \geq b_2 \) iff \( d \geq d^* \). [cf. Table 9.] By trial and error we ascertain that

\[
d^* = 1.38294964.
\]

The relationships between \( b_1 \) and \( b_2 \) for various values of \( d \) are shown in Table 13. [cf. Lemmas 16, 17.]

### Table IV.13

<table>
<thead>
<tr>
<th>( d )</th>
<th>((b_1, b_2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; -1/3 )</td>
<td>( b_1 &lt; b_2 &lt; 0 )</td>
</tr>
<tr>
<td>(-1/3 &lt; d &lt; -1/3 )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
</tr>
<tr>
<td>(-1/3 &lt; d &lt; 2/3 )</td>
<td>( b_1 &lt; 0 &lt; b_2 )</td>
</tr>
<tr>
<td>( d = 2/3 )</td>
<td>( b_1 = 0 = b_2 )</td>
</tr>
<tr>
<td>( 2/3 &lt; d &lt; d^* )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
</tr>
<tr>
<td>( d = d^* )</td>
<td>( 0 &lt; b_1 = b_2 )</td>
</tr>
<tr>
<td>( d^* &lt; d &lt; +\infty )</td>
<td>( 0 &lt; b_2 &lt; b_1 )</td>
</tr>
</tbody>
</table>

3.) \( A < 0 \) according as \( d \geq 0 \).

4.) \( a_1 > a_2 (= 0) \) according as \( d \leq 1.5 \).

We also need to know the \( d \)-values for which the coefficients, \( C_1, C_2, C_3 \), are positive, negative, and zero. [We do not have to worry about \( C_4 \) in this special case, for \( A = 0 \) whenever \( D = 0 \), i.e., iff \( d = 0 \).]

5.) \( C_3 = (1 - \rho_2^2) - \sigma_2^2 (1 - \rho_1^2) = \rho_1^2 - \rho_2^2 = 1/3 \).

6.) \( C_1 = \theta_2 (1 - \rho_1^2) = d^2 + (2/3)d + (2/3)^2 \), for which we know, by Theorem 3,

a.) \( C_1 \) cannot vanish, since \( \rho_2^2 < 1 - \sigma_2^2 (1 - \rho_1^2) = \rho_1^2 \);
b.) therefore, \( C_1 > 0 \) for all \( d \).

7.) \( C_2 = (1 - \rho_2^2) - \theta_1 = -d^2 + (4/3)d - (1/3)^2 \), which can vanish:

a.) \( C_2 = 0 \) iff \( d = (2/3) \pm \sqrt{1/3} \);
b.) \( C_2 > 0 \) iff \( (2 - \sqrt{3})/3 < d < (2 + \sqrt{3})/3 \), approx. \( 0.069 < d < 1.244 \);
c.) \( C_2 < 0 \) otherwise.

Applying the information in (1)-(7) to Table 10, we can tabulate \( \lim U(y) \) as a function of \( d \), as shown in Table 14.

\( y \to \pm \infty \)
TABLE IV.14

<table>
<thead>
<tr>
<th>d</th>
<th>((b_1, b_2))</th>
<th>D</th>
<th>(C_2)</th>
<th>(\lim U(y)) (y \to -\infty)</th>
<th>(\lim U(y)) (y \to +\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; -1/3)</td>
<td>(b_1 &lt; b_2 &lt; 0)</td>
<td>D&lt;0</td>
<td></td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(d = -1/3)</td>
<td>(b_1 &lt; 0 = b_2)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(-1/3 &lt; d &lt; 0)</td>
<td>(b_1 &lt; 0 &lt; b_2)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(d = 0)</td>
<td>(D=0)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(0 &lt; d &lt; 0.09)</td>
<td>(C_2 &lt; 0)</td>
<td>D&gt;0</td>
<td></td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(0.09 &lt; d &lt; 2/3)</td>
<td>(C_2 &gt; 0)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>(d = 2/3)</td>
<td>(b_1 = 0 &lt; b_2)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>(2/3 &lt; d &lt; 1.38)</td>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>(d = d^*)</td>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>(1.38 &lt; d &lt; +\infty)</td>
<td>(0 &lt; b_2 &lt; b_1)</td>
<td></td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

This obviously can be summarized more neatly!

TABLE IV.14-A

<table>
<thead>
<tr>
<th>d</th>
<th>(\lim U(y)) (y \to -\infty)</th>
<th>(\lim U(y)) (y \to +\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; 0.089)</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(0.089 &lt; d &lt; +\infty)</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

We have not determined \(\lim U(y)\) as \(y \to +\infty\) when \(d = -(\rho_1 + \sqrt{\rho_1^2 - \rho_2^2})\)\#0.089, but let us examine the graphs of \(U(y)\) for the values, \(d = 0.08\) and \(d = 0.09\), to see whether to worry about this point.

On the following pages are presented twelve graphs, one for each of the \(d\)-values,

\[
\{-1.0 \quad -1/3 \quad -0.1 \quad 0 \quad 0.01 \quad 0.08 \quad 0.09 \quad 0.5 \quad +2/3 \quad -\rho_1 \quad 1.0 \quad 1.383 \quad (\# d^*) \quad 5.0\}
\]

which represent each of the ten cases in Table 14 plus two values close to the boundary point 0.089. Each graph shows \(Q(y)\), \(G(y)\), and \(U(y)\) over the range of \(y\)-values for which \(G(y)\) could be reliably calculated.

While they constitute no proof as such, these graphs suggest that the "Choose Population 1" decision region is a half-line, \(y > y_0\), for all \(d < 0.089\) and then becomes an interval, \(y_1 < y < y_2\), for all \(d > 0.089\). However, for \(d\) close to 0.089, \(y_2\) is so large that there is virtually no probability for the \(y\)-values exceeding \(y_2\). For all practical purposes, then, it really doesn't matter what the limit of \(U(y)\) is as \(y\to+\infty\) when \(d = -(\rho_1 + \sqrt{\rho_1^2 - \rho_2^2})\).

A word needs to be said about the last graph (the one corresponding to the value \(d = 5\)), for it is a confusing graph to say the least. In this
case, the numerical values of the functions involved are more enlightening than their graphs:

1.) \( Q(y) \) is a U-shaped quadratic having these properties:
   a.) \( Q(y) < 0 \) for \(-2.527 < y < 31.86;\)
   b.) \( \min_{y} Q(y) = Q(14.67) = -2,956.14.\)

2.) \( G(y) \) is roughly J-shaped, as the tabulated figures show:

<table>
<thead>
<tr>
<th>( y )</th>
<th>( Q(y) )</th>
<th>( G(y) )</th>
<th>( U(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-6.0</td>
<td>1314.96</td>
<td>-3 \times 10^{-7}</td>
<td>1314.96</td>
</tr>
<tr>
<td>-5.0</td>
<td>911.63</td>
<td>-10^{-4}</td>
<td>911.63</td>
</tr>
<tr>
<td>-4.0</td>
<td>528.30</td>
<td>-0.02</td>
<td>528.28</td>
</tr>
<tr>
<td>-3.5</td>
<td>344.13</td>
<td>-0.15</td>
<td>343.98</td>
</tr>
<tr>
<td>-3.0</td>
<td>164.96</td>
<td>-0.98</td>
<td>163.98</td>
</tr>
<tr>
<td>-2.75</td>
<td>77.26</td>
<td>-2.31</td>
<td>74.95</td>
</tr>
<tr>
<td>-2.625</td>
<td>33.87</td>
<td>-3.46</td>
<td>30.42</td>
</tr>
<tr>
<td>-2.5625</td>
<td>12.29</td>
<td>-4.20</td>
<td>8.09</td>
</tr>
<tr>
<td>-2.50</td>
<td>-9.20</td>
<td>-5.10</td>
<td>-14.30</td>
</tr>
<tr>
<td>-2.375</td>
<td>-51.96</td>
<td>-7.40</td>
<td>-59.36</td>
</tr>
<tr>
<td>-2.25</td>
<td>-91.41</td>
<td>-10.57</td>
<td>-101.98</td>
</tr>
<tr>
<td>-2.125</td>
<td>-136.55</td>
<td>-14.86</td>
<td>-151.41</td>
</tr>
<tr>
<td>-2.00</td>
<td>-178.37</td>
<td>-20.57</td>
<td>-198.94</td>
</tr>
<tr>
<td>-1.50</td>
<td>-342.54</td>
<td>-64.49</td>
<td>-407.03</td>
</tr>
<tr>
<td>-1.00</td>
<td>-501.70</td>
<td>-157.81</td>
<td>-659.51</td>
</tr>
<tr>
<td>-0.50</td>
<td>-655.87</td>
<td>-305.19</td>
<td>-961.06</td>
</tr>
<tr>
<td>0.00</td>
<td>-805.04</td>
<td>-477.53</td>
<td>-1282.57</td>
</tr>
<tr>
<td>0.50</td>
<td>-949.20</td>
<td>-621.64</td>
<td>-1570.84</td>
</tr>
<tr>
<td>1.00</td>
<td>-1088.37</td>
<td>-684.42*</td>
<td>-1772.79</td>
</tr>
<tr>
<td>1.50</td>
<td>-1222.54</td>
<td>-626.53</td>
<td>-1849.07*</td>
</tr>
<tr>
<td>2.00</td>
<td>-1351.70</td>
<td>-422.82</td>
<td>-1774.52</td>
</tr>
<tr>
<td>2.50</td>
<td>-1475.87</td>
<td>-58.12</td>
<td>-1533.99</td>
</tr>
<tr>
<td>2.5625</td>
<td>-1491.04</td>
<td>-0.72</td>
<td>-1491.76</td>
</tr>
<tr>
<td>2.625</td>
<td>-1506.13</td>
<td>59.36</td>
<td>-1446.77</td>
</tr>
<tr>
<td>2.75</td>
<td>-1536.08</td>
<td>187.59</td>
<td>-1348.47</td>
</tr>
<tr>
<td>2.875</td>
<td>-1565.71</td>
<td>326.65</td>
<td>-1239.06</td>
</tr>
<tr>
<td>3.00</td>
<td>-1595.04</td>
<td>476.62</td>
<td>-1118.42</td>
</tr>
<tr>
<td>3.50</td>
<td>-1709.20</td>
<td>1186.87</td>
<td>-522.33</td>
</tr>
<tr>
<td>3.75</td>
<td>-1764.41</td>
<td>1608.94</td>
<td>-155.47</td>
</tr>
<tr>
<td>3.8125</td>
<td>-1778.02</td>
<td>1721.49</td>
<td>-56.53</td>
</tr>
<tr>
<td>3.875</td>
<td>-1791.55</td>
<td>1836.86</td>
<td>45.32</td>
</tr>
<tr>
<td>4.00</td>
<td>-1818.37</td>
<td>2076.10</td>
<td>257.73</td>
</tr>
<tr>
<td>4.50</td>
<td>-1922.54</td>
<td>3146.08</td>
<td>1223.54</td>
</tr>
<tr>
<td>5.00</td>
<td>-2021.70</td>
<td>4415.35</td>
<td>2393.64</td>
</tr>
</tbody>
</table>

The geometry of the decision regions will be discussed in Subsection 4.5.3.
Figure IV.5
\( \alpha = -1 \)

\[
\begin{align*}
\delta_1 &= 1 \\
\delta_2 &= 0 \\
\epsilon_1 &= \epsilon_2 = 1 \\
\rho_1 &= -\frac{3}{2} \\
\rho_2 &= \frac{1}{2} \\
\alpha_1 &= \frac{1}{2} \\
\kappa_1 &= 0
\end{align*}
\]
Figure IV.6
\[ a = \sqrt[3]{3} = -f_a \]

<table>
<thead>
<tr>
<th>( a )</th>
<th>( b )</th>
<th>( c )</th>
<th>( d )</th>
<th>( e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( -\sqrt{3} )</td>
<td>( \sqrt{3} )</td>
</tr>
<tr>
<td>( x )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ G(y) \]
\[ U(y) \]
\[ Q(y) \]
Figure IV.7
\(d = -0.1\)

| \(\delta_1\) | 1 |
|\(\delta_2\) | 0 |
|\(\theta_1 = \theta_2\) | 1 |
|\(\phi = -\frac{\pi}{3}\) | |
|\(\beta = \frac{1}{3}\) | |
|\(\alpha = 1/2\) | |
|\(x = 0\) | |
Figure IV.8
\( d = 0 \)

\[ \begin{align*}
\delta_1 &= 1 \\
\delta_2 &= 0 \\
\alpha_n &= \frac{1}{2} \\
\rho_1 &= -\frac{2}{3} \\
\rho_2 &= \frac{1}{3} \\
\alpha &= \frac{1}{2} \\
\kappa &= 0
\end{align*} \]
\textbf{Figure IV.9}
\[ d = 0.01 \]
\begin{align*}
\delta_1 &= 1 \\
\delta_2 &= 0 \\
\beta_1 &= \beta_2 = 1 \\
\beta_1 &= -\frac{2}{3} \\
\beta_2 &= +\frac{1}{3} \\
\alpha &= \frac{1}{2} \\
\chi &= 0
\end{align*}
Figure IV.10

$d = 0.08$

\[
\begin{array}{|c|}
\hline
\delta_1 = 1 \\
\delta_2 = 0 \\
\sigma_1 \sigma_2 = 1 \\
\rho_1 = -\sqrt{3} \\
\rho_2 = +\sqrt{3} \\
\alpha = \frac{1}{2} \\
\kappa = 0 \\
\hline
\end{array}
\]

$Q(y)$

$\sim G(y)$

$\sim U(y) \to -\infty$
Figure IV.11
\[ d = 0.09 \]

\[
\begin{array}{|c|}
\hline
\delta_1 &=& \frac{1}{2} \\
\delta_2 &=& 0 \\
\sigma &=& 0 \\
\lambda_1 &=& -\frac{2}{3} \\
\lambda_2 &=& +\frac{1}{3} \\
\alpha &=& \frac{1}{2} \\
\kappa &=& 0 \\
\hline
\end{array}
\]
Figure IV.12

\[ d = 0.5 \]

\[ \begin{array}{|c|c|}
\hline
\psi & 1 \\
\bar{\rho} & 0 \\
\theta & 1 \\
\rho & -\frac{2}{3} \\
\rho_3 & \frac{1}{3} \\
\alpha & \frac{1}{2} \\
\kappa & 0 \\
\hline
\end{array} \]
Figure IV.13
\[ \alpha = \frac{2}{3} = -\beta_1 \]

| \( \delta_1 \) | 1 |
| \( \delta_2 \) | 0 |
| \( \sigma_1 = \sigma_2 \) | 1 |
| \( \beta_2 = -\frac{2}{3} \) |
| \( \beta_2 = +\frac{1}{3} \) |
| \( \alpha = \frac{1}{2} \) |
| \( \chi = 0 \) |
Figure IV.14
\[ d = +1 \]

\[
\begin{array}{c|c}
\delta_1 & 1 \\
\delta_2 & 0 \\
P & 1 \\
\rho & -2/3 \\
\sigma & 41/3 \\
\alpha & 1/2 \\
K & 0 \\
\end{array}
\]
Figure IV.15
\[ d = 1.583 \]
\[ [b_x(d) = b_2(d)] \]
Figure IV.16
\( d = 5 \)

\[
\begin{align*}
\delta_1 &= 1 \\
\delta_2 &= 0 \\
\gamma &= 1 \\
\rho &= -\frac{2}{3} \\
\beta &= -\frac{1}{3} \\
\alpha &= \frac{1}{2} \\
\chi &= 0 
\end{align*}
\]

U(y) and Q(y)

\[\rightarrow\]

\[\Theta(y)\rightarrow\]

\[\leftarrow U(y)\]
4.5.2 The Limits of \( U(y) \) as \( y \to \infty \) When \( k=0, \sigma_1=\sigma_2=1, \) and \( \rho_1<\rho_2 \).

In this subsection we generalize somewhat on the preceding special case in order to study the effects of the relative values of \( \rho_1 \) and \( \rho_2 \) on the tails of the function \( U(y) \). As before, we need to know the forms of the functions,

\[
\begin{align*}
\theta_i(d|\pi) &= d^2 + 2\rho_1d + 1, \quad i=1,2; \\
D(d|\pi) &= \theta_2 - \theta_1 = 2(\rho_2 - \rho_1)d; \\
b_i(d|\pi) &= (d+\rho_1)/(\sqrt{1-\rho_1^2}\sqrt{\theta_1(d|\pi)}), \quad i=1,2;
\end{align*}
\]

which depend only on the parameters of the covariance matrices, and

\[
\begin{align*}
a_1(d|\pi) &= [(\delta_2 - \rho_1\delta_1)d - (\delta_1 - \rho_1\delta_2)]/(\sqrt{1-\rho_1^2}\sqrt{\theta_1(d|\pi)}); \\
a_2 &= 0; \\
A &= \theta_2(\delta_1d + \delta_2);
\end{align*}
\]

which also depend on \( \kappa \) and \( \delta \), and

\[
B = \theta_2\left\{\gamma^2 - 2\theta_1\ln\left[\frac{\alpha}{1-\alpha}\frac{\sqrt{\theta_2}}{\sqrt{\theta_1}\Phi(-\delta_1)}\right]\right\},
\]

which depends on all the parameters. In order to use Table 10, we need to know the \( d \)-values for which these functions are positive, negative, and zero.

[1.] \( D \geq 0 \), according as \( d \geq 0 \), since \( \rho_1<\rho_2 \).

[2.] Concerning the relative values of \( b_1 \) and \( b_2 \): By Lemma 16, the relationships shown in Table 16 apply for all \( \rho_1<\rho_2 \). From Table 9 we can construct Tables 17-25 and the following Index to show the relationships between \( b_1 \) and \( b_2 \) for values of \( d \) not between \( -\rho_2 \) and \( -\rho_1 \):

<table>
<thead>
<tr>
<th>INDEX</th>
</tr>
</thead>
</table>

**TABLE IV.16**

<table>
<thead>
<tr>
<th>( b_1, b_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_2 )</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( -1&lt;\rho_2\leq0 )</td>
</tr>
<tr>
<td>( -1&lt;\rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( 0&lt;\rho_2\leq1/3 )</td>
</tr>
<tr>
<td>( -\rho_2&lt;\rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( -\rho_2&lt;\rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( -1&lt;\rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
<tr>
<td>( \rho_1&lt;\rho_2 )</td>
</tr>
</tbody>
</table>

**TABLE IV.17**

<table>
<thead>
<tr>
<th>( b_1, b_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_2 )</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( -\infty&lt;d&lt;\rho_2 )</td>
</tr>
<tr>
<td>( -\rho_2&lt;d&lt;\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
<tr>
<td>( d=\rho_2 )</td>
</tr>
<tr>
<td>( d=\rho_1 )</td>
</tr>
</tbody>
</table>
[3.] The behavior of $A$ depends on $\delta$:

a.) If $\delta=0$, then $A=0$.

b.) If $\delta_1=0<\delta_2$, then $A$ is a quadratic in $d$ such that $A>0$ for all $d$.

c.) If $\delta_1>0$, then $A \leq 0$ according as $d \leq -\left(\delta_2/\delta_1\right)$.

[4.] The behavior of $a_1$ is more complicated still. Since $k=0$, we have from Lemma 18:

a.) If $\delta=0$, then $a_1=a_2=0$ for all $d$.

b.) If $\delta_1=0<\delta_2$, then $a_1=\delta_2b_1$.

c.) If $\delta_1>0$ and $\delta_2=\rho_1\delta_1$, which can happen iff $\rho_1>0$, then $a_1=0=a_2$ for all $d$.

d.) If $\delta_1>0$ and $\delta_2<\rho_1\delta_1$, which can happen iff $\rho_1>0$, then $a_1 \leq 0=a_2$ according as $d \geq \left(\delta_1-\rho_1\delta_2\right)/(\delta_2-\rho_1\delta_1)$.

e.) If $\delta_1>0$ and $\delta_2<\rho_1\delta_1$, then $a_1 \leq 0=a_2$ according as $d \leq \left(\delta_1-\rho_1\delta_2\right)/(\delta_2-\rho_1\delta_1)$.

Finally, we must consider the values of the coefficients $C_i$, $i=1-3$:

[5.] $C_3 = \rho_1^2-\rho_2^2$, which is

a.) positive iff $-1 < \rho_1 < -|\rho_2|$;

b.) zero iff $\rho_1 = -\rho_2$;

c.) negative otherwise.

[6.] $C_1 = \theta_2-(1-\rho_1^2) = d^2 + 2\rho_2d + \rho_1^2$, which is

a.) zero iff

i.) $C_3=0$ and $d=\rho_1$; i.e. $d=\rho_1=-\rho_2$; or

ii.) $C_3<0$ and $d = -\rho_2 \pm \sqrt{\rho_2^2-\rho_1^2} \equiv (\rho_2**,\rho_2**)$;

b.) negative iff $C_3<0$ and $\rho_2** < d < \rho_2**$;

c.) positive otherwise.

[7.] $C_2 = (1-\rho_2^2)-\theta_1 = -(d^2 + 2\rho_1d + \rho_2^2)$, which is

a.) zero iff

i.) $C_3=0$ and $d=\rho_2$; i.e. $d=\rho_2=-\rho_1$; or

ii.) $C_3>0$ and $d = -\rho_1 \pm \sqrt{\rho_1^2-\rho_2^2} \equiv (\rho_1**,\rho_1*)$;
b.) positive iff $C_3 > 0$ and $\rho_1^* < d < \rho_1^{**}$;  
c.) negative otherwise.

[8.] To partition the real line, we need to know how $\rho_1^*$ and $\rho_1^{**}$ compare with zero and $-\rho_j$, \(i,j=1,2, i \neq j\).

a.) Both $\rho_1^*$ and $\rho_1^{**}$ have the same sign as $-\rho_1$ if $\rho_j \neq 0$.

b.) If $\rho_j = 0$, then $\rho_1^*$ and $\rho_1^{**}$ are equal to zero and $-2\rho_1$.

c.) $\rho_2^* < -\rho_2 < -\rho_1 < \rho_1^{**}$.

d.) $\rho_1^* \leq -\rho_2$ iff $\rho_2 \leq 0$.

e.) $\rho_2^{**} \geq -\rho_1$ iff $\rho_1 \geq 0$.

**proof:**

By a well-known theorem and the definitions above of $C_1$, $i=1,2$, we have

$$\rho_1^*\rho_1^{**} = \rho_j^2 \geq 0.$$  

This establishes (a) and (b).

Part (c) follows directly from the definitions of $\rho_1^*$ and $\rho_2^{**}$ and the fact that $\rho_1 < \rho_2$.

$$\rho_1^* \leq -\rho_2 \text{ iff } \rho_2 - \rho_1 \leq \sqrt{\rho_1^2 - \rho_2^2}$$  
$$\text{iff } (\rho_2 - \rho_1)^2 \leq -\rho_2 - \rho_1(\rho_2 + \rho_1)$$  
$$\text{iff } \rho_2 - \rho_1 \leq -(\rho_2 + \rho_1)$$  
$$\text{iff } \rho_2 \leq 0.$$  

$$\rho_2^{**} \geq -\rho_1 \text{ iff } \sqrt{\rho_2^2 - \rho_1^2} \geq \rho_2 - \rho_1 \text{ ()}>0\text{ iff } (\rho_2 - \rho_1)(\rho_2 + \rho_1) \geq (\rho_2 - \rho_1)^2 \text{ iff } \rho_1 \geq 0.$$  

On the basis of [1]-[8] we can distinguish seven cases.

**Case I:** $-1 < \rho_2 < 0$.

For all $\rho_1 < \rho_2$ we have $C_3 > 0$, $C_4 > 0$ for all $d$, Tables 16 and 17 give the relative values of $b_1$ and $b_2$, and $C_2 > 0$ for all $d$, for $[-\rho_2, -\rho_1]$, since $\rho_1^* < -\rho_2 < -\rho_1 < \rho_1^{**}$. Consequently, the limits of $U(y)$ can be tabulated as a function of $d$ as shown in Table 26.

(i) When $d = 0$, $U(y) = \begin{cases} (-\infty) \text{ if } A \neq 0 \\ B \text{ if } A = 0 \end{cases}$ as $y \to +\infty$.  

\[ \begin{array}{|c|c|}
\hline
\text{d} & (b_1, b_2) \\
\hline
-\infty < d < d^* & b_2 < b_1 < 0 \\
\hline
-\infty < d < d^* & b_1 = b_2 < 0 \\
\hline
-\infty < d < d^* & b_2 < b_1 < 0 \\
\hline
-\infty < d < d^* & b_1 = b_2 < 0 \\
\hline
-\infty < d < d^* & b_2 < b_1 < 0 \\
\hline
-\infty < d < d^* & b_1 = b_2 < 0 \\
\hline
-\infty < d < d^* & b_2 < b_1 < 0 \\
\hline
-\infty < d < d^* & b_1 = b_2 < 0 \\
\hline
\end{array} \]
From [3] we see that A=0 iff δ2=0; and A cannot be negative when d=0. Moreover, by definition
\[ B = \delta_2^2 - 2\ln \left( \frac{\alpha}{1-\alpha} \frac{1/2}{\phi(-\delta_1)} \right) \] when d=0.

Therefore, \( U(y) \rightarrow \begin{cases} B \leq 0 & \text{if } \delta_2=0 \\ -\infty & \text{if } \delta_2>0 \end{cases} \) as \( y \rightarrow +\infty \).

<table>
<thead>
<tr>
<th>d</th>
<th>(b1, b2)</th>
<th>D</th>
<th>Clm U(y) ( y \rightarrow -\infty )</th>
<th>Clm U(y) ( y \rightarrow +\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; 0)</td>
<td>b1 &lt; b2 &lt; 0</td>
<td>D &lt; 0</td>
<td>+\infty</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>d = 0</td>
<td>b1 &lt; b2 &lt; 0</td>
<td>D = 0</td>
<td>+\infty</td>
<td>see (i)</td>
</tr>
<tr>
<td>0 &lt; d &lt; -p2</td>
<td>b1 &lt; b2 = 0</td>
<td>D &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d = -p2</td>
<td>b1 &lt; b2 = 0</td>
<td>C2 &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>-p2 &lt; d &lt; -p1</td>
<td>b1 &lt; b2 &lt; b2</td>
<td>C2 &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d = -p1</td>
<td>b1 = b2 &lt; b2</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>-p1 &lt; d &lt; d*</td>
<td>0 &lt; b1 &lt; b2</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d = d*</td>
<td>0 &lt; b1 &lt; b2</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d* &lt; d &lt; +\infty</td>
<td>0 &lt; b2 &lt; b1</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

**Case 2:** \( \rho_2 = 0 \).

For all \( \rho_1 < \rho_2 \) we have \( C_3 > 0 \), \( C_1 > 0 \) for all \( d \), Tables 16 and 17 give the values of \( b_1 \) and \( b_2 \), and \( \rho_1 \neq 0 \) so that \( C_2 > 0 \) for all \( d \in (0, -\rho_1) \). The limits of \( U(y) \) are shown in Table 27.

<table>
<thead>
<tr>
<th>d</th>
<th>(b1, b2)</th>
<th>D</th>
<th>Clm U(y) ( y \rightarrow -\infty )</th>
<th>Clm U(y) ( y \rightarrow +\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; 0)</td>
<td>b1 &lt; b2 &lt; 0</td>
<td>D &lt; 0</td>
<td>+\infty</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>d = 0</td>
<td>b1 &lt; b2 &lt; 0</td>
<td>D = 0</td>
<td>+\infty</td>
<td>see (ii)</td>
</tr>
<tr>
<td>0 &lt; d &lt; -p1</td>
<td>b1 &lt; b2 &lt; b2</td>
<td>D &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d = -p1</td>
<td>b1 &lt; b2 &lt; b2</td>
<td>C2 &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>-p1 &lt; d &lt; d*</td>
<td>0 &lt; b1 &lt; b2</td>
<td>C2 &gt; 0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d = d*</td>
<td>0 &lt; b1 = b2</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>d* &lt; d &lt; +\infty</td>
<td>0 &lt; b2 &lt; b1</td>
<td>+\infty</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

(\textit{ii}) By the same reasoning as in (i), we have A=0 iff δ2=0 and A>0 otherwise when d=0. Table 10 shows, then,

\[ U(y) \rightarrow \begin{cases} B - 2(\ln 2) < 0 & \text{if } \delta_2=0 \\ -\infty & \text{if } \delta_2>0 \end{cases} \] as \( y \rightarrow +\infty \) when d=0.

**Case 3:** \( \rho_1 = -\rho_2 < 0 \).

Then \( C_3 = 0 \), \( C_1 = 0 \) iff \( d = -\rho_2 \) and is positive otherwise, \( C_2 = 0 \) iff \( d = -\rho_1 \) and is negative otherwise, and Tables 16 and 18 apply.

The limits of \( U(y) \) are shown in Table 28.
TABLE IV.28

<table>
<thead>
<tr>
<th>d</th>
<th>(b₁,b₂)</th>
<th>D</th>
<th>C₁</th>
<th>( \lim U(y) ) ( y+\to-\infty )</th>
<th>( \lim U(y) ) ( y+\to+\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty&lt;\text{d}&lt;\rho₂)</td>
<td>b₁&lt; b₂&lt;0</td>
<td>D&lt;0</td>
<td>C₁=0</td>
<td>( \text{df. iiia} )</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>( \text{d}=\rho₂)</td>
<td>b₁&lt; b₂=0</td>
<td></td>
<td>C₁=0</td>
<td>( \text{df. iiib} )</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(-\rho₂&lt;\text{d}&lt;0)</td>
<td>b₁&lt; b₂&lt;\rho₂</td>
<td>D=0</td>
<td>C₁&lt;0</td>
<td>( +\infty )</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>( \text{d}&lt;\rho₁)</td>
<td>b₁= b₂=0</td>
<td>D&gt;0</td>
<td>C₂&lt;0</td>
<td>( +\infty )</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(-\rho₁&lt;\text{d}&lt;\infty)</td>
<td>0&lt; b₁&lt; b₂</td>
<td></td>
<td>C₂=0</td>
<td>( +\infty )</td>
<td>( \text{df. iiia} )</td>
</tr>
</tbody>
</table>

(iii) When one of the functions \( C_i, i=1-3 \), is zero, we can rewrite \( U(y) \) as follows:

\[
U(y) = [Dy^2-2Ay+B] - 2\theta_1\theta_2 \ln \left( \frac{\phi(a_1-b_1y)}{\xi(a_1-b_1y)} \frac{\xi(a_2-b_2y)}{\phi(a_2-b_2y)} \right) 
\]

\[
= [D+\theta_1\theta_2(b_1^2-b_2^2)]y^2 - 2[A+\theta_1\theta_2(a_1b_1-a_2b_2)]y + 
+ [B+\theta_1\theta_2(a_2^2-a_1^2)] - 2\theta_1\theta_2 \ln \left( \frac{\xi(a_2-b_2y)}{\xi(a_1-b_1y)} \right) .
\]

If the coefficient of \( y^2 \) is zero and \( k=0 \), then \( U(y) \) can be written:

\[
U(y) = -2[A+\theta_1\theta_2a_1b_1]y + [B+\theta_1\theta_2a_2^2] - 2\theta_1\theta_2 \ln \left( \frac{\xi(-b_2y)}{\xi(a_1-b_1y)} \right) . (44)
\]

When \( C_3=0 \), then equation (44) applies for all \( d \) because

\[
D+\theta_1\theta_2(b_1^2-b_2^2) = C_3 \text{ for all } d
\]

by (40).

Let \( C_5 \) denote the coefficient of \( y \) in (44).

From [3] and [4] we see that \( C_5 \) depends in complex fashion on the relative values of the parameters, \( \rho_1, \rho_2, \delta_1, \delta_2, \) and \( d \).

Using the definitions and Lemma 18, we have for these means:

\( \delta=0 \) implies \( A=a_1=0 \) implies \( C_5=0 \)

implies \( U(y) = B - 2\theta_1\theta_2 \ln \left( \frac{\xi(-b_2y)}{\xi(-b_1y)} \right) 
\]

\[
= -2\theta_1\theta_2 \ln \left( \frac{a \sqrt{\frac{\theta_2}{\theta_1}} \xi(-b_2y)}{\xi(-b_1y)} \right) .
\]

\( \delta=\delta_2 \), i.e. \( \delta_1=0<\delta_2 \), implies \( A=\delta_2 \theta_2 \) and \( a_1=\delta_2 b_1 \)

implies \( C_5 = -2\delta_2 \theta_2 (1+\theta_2 b_1^2) < 0 \).

\( \delta=\delta_1 \), i.e. \( \delta_1>0>\delta_2 \), implies

\[
C_5 = -2 \left( \delta_1 \theta_2 d + \theta_1 \theta_2 \frac{d + \rho_1}{\sqrt{1-\rho_1^2}} \frac{-\delta((\rho_1 d+1))}{\sqrt{1-\rho_1^2}} \right) = \frac{2 \rho_1 \theta_1 \theta_2}{1-\rho_1^2} .
\]
Also, \( \rho_1 < 0 < \delta_1 \) implies
\[
\begin{cases} 
\bar{a}_1 > 0 \text{ according as } d > \frac{\delta_1 + \rho_2 \delta_2}{\delta_2 + \rho_2 \delta_1} , \text{ which is positive;} \\
A < 0 \text{ according as } d < \frac{-(\delta_2/\delta_1)}{1}, \text{ non-positive.}
\end{cases}
\]
\[
\delta = \frac{1}{1}, \text{ i.e. } \delta_1 = \delta_2 = 0, \text{ implies}
\]
\[
C_5 = -2 \left( \delta \theta_2 (d+1) + \theta_1 \theta_2 \frac{d+\rho_1}{\sqrt{1-\rho_1^2}} \frac{\delta (1-\rho_1) \delta - \delta (1-\rho_1)}{\sqrt{1-\rho_2^2}} \right)
\]
\[
= \frac{-2 \delta \theta_1 \theta_2}{1+\rho_1} < 0.
\]

For these four types of mean vectors, then, our information about \( C_5 \) is summarized in Table 29, where \( \delta > 0 \).

(a) The limits of \( U(y) \) are given in terms of \( C_3 \)--and so are undefined when \( C_3 = 0 \)—in Table 10 only when \( b_1 \) and \( b_2 \) are unequal and of like sign. However, in that case, the limit of the third term of (44) as \( b_2 y \to +\infty \) is proportional to
\[
\lim \left[ \ln \frac{\xi(-b_2 y)}{\xi(a_1-b_1 y)} \right] = \ln \left[ \frac{\lim_{b_2 \to \infty} b_2 \omega(-b_2 y)}{\lim_{b_1 \to \infty} b_1 \omega(a_1-b_1 y)} \right] = \ln(b_2/b_1),
\]
which is finite. Consequently, \( \lim U(y) \) depends on \( C_5 \). From Table 29 we can compute the missing limits for the four special mean vectors:
- When \( \delta = 0 \), \( \lim U(y) = B - 2 \theta_1 \theta_2 \ln(b_2/b_1) = -2 \theta_1 \theta_2 \ln\left(\frac{\alpha \delta + \rho_2}{1-\alpha \delta - \rho_2}\right) \).
- When \( \delta = \delta u_1 \) or \( \delta u_2 \) or \( \delta_1 \), \( \delta > 0 \), then
  \( U(y) \to +\infty \) as \( y \to -\infty \) when \( b_1 < 0 \), \( i=1,2 \);
  \( U(y) \to -\infty \) as \( y \to +\infty \) when \( b_1 > 0 \), \( i=1,2 \).

Note that for \( d > \rho_2 \), \( \lim U(y) \) is negative; but for \( d < -\rho_2 \),
\[
\lim U(y) \leq 0 \text{ according as } d \leq -\rho_2/(2\alpha - 1).
\]

(b) When \( d = -\rho_2 = \rho_1 \), then \( C_1 = 0 \) and the limit of \( U(y) \) as \( y \to -\infty \) is undefined in Table 10. In (44) we see that as \( y \to -\infty \), the third term tends to
\[
2 \theta_1 \theta_2 \left[ \ln \left( \lim \xi(a_1-b_1 y) \right) - \ln \xi(0) \right] = +\infty.
\]
From this and (44) and Table 29 we conclude that for these four special mean vectors, \( U(y) \to +\infty \) as \( y \to -\infty \).
(a) When \( d = -\rho_1 = \rho_2 \), then \( C_2 = 0 \) and the limit of \( U(y) \) as \( y \to +\infty \) is undefined in Table 10. In (44) we see that as \( y \to +\infty \), the third term tends to

\[ -2812 \varepsilon \{ \lim \ln \xi(-b_2y) \} - \ln \xi(a_1) \] \( = -\infty \).

From this and (44) and Table 29 we conclude that for these four special mean vectors, \( U(y) \to -\infty \) as \( y \to +\infty \).

**Case 4: \( \rho_1 < -\rho_2 < 0 \).**

Then \( C_3 > 0 \), \( C_1 > 0 \) for all \( d \), and \( \rho_1* > -\rho_2 \) so that for \( d \in [-\rho_2, -\rho_1] \)

\( C_2 \leq 0 \) according as \( d \leq \rho_1* (> 0) \).

**(A)** If Table 17 applies, the limits of \( U(y) \) are given in Table 30-A.

<table>
<thead>
<tr>
<th>( d )</th>
<th>( (b_1, b_2) )</th>
<th>( D )</th>
<th>( \lim_{y \to -\infty} U(y) )</th>
<th>( \lim_{y \to +\infty} U(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; -\rho_2 )</td>
<td>( b_1 &lt; b_2 &lt; 0 )</td>
<td>D&lt;0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( d = -\rho_2 )</td>
<td>( b_1 &lt; b_2 = 0 )</td>
<td></td>
<td>-\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>(-\rho_2 &lt; d &lt; 0 )</td>
<td>( b_1 &lt; 0 &lt; b_2 )</td>
<td>D=0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( 0 &lt; d &lt; a )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
<td>D&lt;0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( \rho_1^* &lt; d &lt; -\rho_1 )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
<td>D&gt;0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( d = -\rho_1 )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>(-\rho_1 &lt; d &lt; a )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td>D&gt;0</td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>( d = d^* )</td>
<td>( 0 &lt; b_1 = b_2 )</td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>( d^* &lt; d &lt; +\infty )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

**(B)** If Table 20 applies, then Table 30-A gives the limits of \( U(y) \) for all \( d \geq -\rho_2 \). The limits for \( d < -\rho_2 \) are given in Table 30-B.

<table>
<thead>
<tr>
<th>( d )</th>
<th>( (b_1, b_2) )</th>
<th>( D )</th>
<th>( \lim_{y \to -\infty} U(y) )</th>
<th>( \lim_{y \to +\infty} U(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; d' )</td>
<td>( b_1 &lt; b_2 &lt; 0 )</td>
<td>D&lt;0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( d = d' )</td>
<td>( b_1 = b_2 &lt; 0 )</td>
<td></td>
<td>-\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( d' &lt; d &lt; d'' )</td>
<td>( b_2 &lt; b_1 &lt; 0 )</td>
<td>D&gt;0</td>
<td>+\infty</td>
<td>-\infty</td>
</tr>
<tr>
<td>( d = d'' )</td>
<td>( b_1 = b_2 &lt; 0 )</td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
<tr>
<td>( d'' &lt; d &lt; -\rho_2 )</td>
<td>( b_1 &lt; b_2 &lt; 0 )</td>
<td></td>
<td>+\infty</td>
<td>+\infty</td>
</tr>
</tbody>
</table>

**(C)** If Table 21 or Table 22 applies, then Table 30-A gives the limits of \( U(y) \) for all \( d \), since \( U(y) \to +\infty \) as \( y \to +\infty \) as long as \( C_3, D, b_1, \) and \( b_2 \) are all positive.

**Case 5: \( -\rho_2 < \rho_1 < 0 \).**

Then \( C_3 < 0 \), \( C_2 < 0 \) for all \( d \), \( \rho_2** < -\rho_1 \) so that for \( d \in [-\rho_2, -\rho_1] \),
\( C_1 \leq 0 \) according as \( d \leq \rho_2^{**} \) \(<0\).

(A) If Table 19 applies, then Table 31-A gives the limits of \( U(y) \).

<table>
<thead>
<tr>
<th>( d )</th>
<th>( (b_1, b_2) )</th>
<th>( D )</th>
<th>( C_1 )</th>
<th>( \lim U(y)_{y-&gt;-\infty} )</th>
<th>( \lim U(y)_{y-&gt;+\infty} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; d^* )</td>
<td>( b_2 &lt; b_1 &lt; 0 )</td>
<td>( D &lt; 0 )</td>
<td>(-\infty )</td>
<td>(-\infty )</td>
<td></td>
</tr>
<tr>
<td>( d = d^* )</td>
<td>( b_2 = b_1 &lt; 0 )</td>
<td>(-\infty )</td>
<td>(-\infty )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d^* &lt; d &lt; -\rho_2 )</td>
<td>( b_1 &lt; b_2 &lt; 0 )</td>
<td>(-\infty )</td>
<td>(-\infty )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d = -\rho_2 )</td>
<td>( b_1 &lt; b_2 = 0 )</td>
<td>(-\infty )</td>
<td>(-\infty )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-\rho_2 &lt; d &lt; \rho_2^{**} )</td>
<td>( b_1 &lt; 0 &lt; b_2 )</td>
<td>( C_1 &lt; 0 )</td>
<td>(-\infty )</td>
<td>(-\infty )</td>
<td></td>
</tr>
<tr>
<td>( \rho_2^{**} &lt; d &lt; 0 )</td>
<td>( b_1 &lt; 0 &lt; b_2 )</td>
<td>( D = 0 )</td>
<td>( C_1 &lt; 0 )</td>
<td>( D &gt; 0 )</td>
<td></td>
</tr>
<tr>
<td>( d = 0 )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
<td>( C_1 &gt; 0 )</td>
<td>( +\infty )</td>
<td>(-\infty )</td>
<td></td>
</tr>
<tr>
<td>( 0 &lt; d &lt; -\rho_1 )</td>
<td>( b_1 &lt; b_2 )</td>
<td>( D &gt; 0 )</td>
<td>( +\infty )</td>
<td>(-\infty )</td>
<td></td>
</tr>
<tr>
<td>( d = -\rho_1 )</td>
<td>( b_1 = 0 &lt; b_2 )</td>
<td>( +\infty )</td>
<td>(-\infty )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-\rho_1 &lt; d &lt; +\infty )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td>( +\infty )</td>
<td>(-\infty )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(B) If Table 23 or Table 24 applies, then Table 31-A gives the limits of \( U(y) \) for all \( d \), since \( U(y) \to -\infty \) as \( y \to +\infty \) as long as \( C_3 \), \( D \), \( b_1 \), and \( b_2 \) are all negative.

(C) If Table 25 applies, then Table 31-A gives the limits of \( U(y) \) for all \( d \leq -\rho_1 \), and Table 31-B gives them for \( d > -\rho_1 \).

<table>
<thead>
<tr>
<th>( d )</th>
<th>( (b_1, b_2) )</th>
<th>( D )</th>
<th>( \lim U(y)_{y-&gt;-\infty} )</th>
<th>( \lim U(y)_{y-&gt;+\infty} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\rho_1 &lt; d &lt; d' )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td>( D &gt; 0 )</td>
<td>( +\infty )</td>
<td>(-\infty )</td>
</tr>
<tr>
<td>( d = d' )</td>
<td>( 0 &lt; b_1 = b_2 )</td>
<td>( +\infty )</td>
<td>(+\infty )</td>
<td></td>
</tr>
<tr>
<td>( d' &lt; d &lt; d'' )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td>(+\infty )</td>
<td>(+\infty )</td>
<td></td>
</tr>
<tr>
<td>( d = d'' )</td>
<td>( 0 &lt; b_1 = b_2 )</td>
<td>(+\infty )</td>
<td>(+\infty )</td>
<td></td>
</tr>
<tr>
<td>( d'' &lt; d &lt; +\infty )</td>
<td>( 0 &lt; b_1 &lt; b_2 )</td>
<td>(+\infty )</td>
<td>(-\infty )</td>
<td></td>
</tr>
</tbody>
</table>

Case 6: \( \rho_1 = 0 < \rho_2 \).

Then \( C_3 < 0 \), \( C_2 < 0 \) for all \( d \), \( \rho_2^{**} = 0 \) so that for \( d > \rho_2^{*} \) \( C_1 \leq 0 \) according as \( d \leq 0 \), and Table 19 applies. The limits of \( U(y) \) are given in Table 32.

(iv) By the same reasoning as in (i) we have, when \( d = 0 \), \( A = 0 \) iff \( \delta_2 = 0 \) and \( A > 0 \) otherwise. From Table 10, then, as \( y \to -\infty \),

\[
U(y) \to \begin{cases} 
B - 2a_1\rho_2 \ln \psi(a_1) = -2\ln\{a/[2(1-a)]\} & \text{if } \delta_2 = 0; \\
+\infty & \text{if } \delta_2 > 0.
\end{cases}
\]
Case 7: \(0 < \rho_1 < \rho_2\).

Then \(C_3 < 0, C_2 < 0\) for all \(d\), \(\rho_2^{**} > 0 > -\rho_1\) so that \(C_1 < 0\) for all \(d \in [-\rho_2, -\rho_1]\), and Table 19 applies. The limits of \(U(y)\) are given in Table 33.

### Table IV.33

<table>
<thead>
<tr>
<th>(d)</th>
<th>((b_1, b_2))</th>
<th>(D)</th>
<th>(\lim U(y))</th>
<th>(\lim U(y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-\infty &lt; d &lt; d^*)</td>
<td>(b_2 &lt; b_1 &lt; 0)</td>
<td>(D &lt; 0)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
</tr>
<tr>
<td>(d^* = d^0)</td>
<td>(b_1 = b_2 &lt; 0)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td></td>
</tr>
<tr>
<td>(d^* &lt; d &lt; -\rho_2)</td>
<td>(b_1 &lt; b_2 &lt; 0)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td></td>
</tr>
<tr>
<td>(d = -\rho_2)</td>
<td>(b_1 = b_2 = 0)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td></td>
</tr>
<tr>
<td>(-\rho_2 &lt; d &lt; -\rho_1)</td>
<td>(b_1 &lt; 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td></td>
</tr>
<tr>
<td>(d = -\rho_1)</td>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(-\infty)</td>
<td>(-\infty)</td>
<td></td>
</tr>
<tr>
<td>(d = 0)</td>
<td>(b_1 = 0 &lt; b_2)</td>
<td>(D = 0)</td>
<td>(\lim ) (U(y)) (\lim ) (U(y))</td>
<td></td>
</tr>
<tr>
<td>(0 &lt; d &lt; +\infty)</td>
<td>(0 &lt; b_1 &lt; b_2)</td>
<td>(D &gt; 0)</td>
<td>(\lim ) (U(y)) (\lim ) (U(y))</td>
<td></td>
</tr>
</tbody>
</table>

\((v)\) By the same reasoning as in \((i)\) and \((iv)\), when \(d = 0\)

\[ U(y) = \begin{cases} B = -2ln\left[\left(\frac{c}{1-c}\right)\frac{1}{2}\frac{1}{1-c}\right] & \text{if } \delta_2 < 0; \\ +\infty & \text{if } \delta_2 > 0. \end{cases} \]

as \(y \to -\infty\).

The following theorem summarizes the information obtained in the above seven cases.

**Theorem IV.4:** For any set of parameters \(\pi\) such that \(k = 0, \sigma_1 = \sigma_2 = 1\), and \(\rho_1 < \rho_2\), Table 34 gives the range of \(d\)-values for which each of the four major relationships between

\[\lim U(y) \quad \text{and} \quad \lim U(y)\]

holds. The small number in the upper left-hand corner refers to the list of exceptions below. The numbers, \(\rho_1^{*}\) and \(\rho_2^{**}\) are defined:

\[\begin{cases} \rho_1^{*} = -\rho_1 - \sqrt{\rho_1^2 - \rho_2^2} \\ \rho_2^{**} = -\rho_2 + \sqrt{\rho_2^2 - \rho_1^2} \end{cases}\]
<table>
<thead>
<tr>
<th>Relative Values of $\rho_1$ and $\rho_2$</th>
<th>$\lim_{y \to -\infty} U(y) &lt; 0$</th>
<th>$\lim_{y \to +\infty} U(y) &gt; 0$</th>
<th>$\lim_{y \to -\infty} U(y) &gt; 0$</th>
<th>$\lim_{y \to +\infty} U(y) &lt; 0$</th>
<th>Both limits are negative</th>
<th>Both limits are positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1 &lt; \rho_2 &lt; 0$</td>
<td>$d \leq 0$</td>
<td>$d = 0$</td>
<td>$d &lt; \rho_1$</td>
<td>$d &gt; 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1 &lt; -\rho_2 &lt; 0$</td>
<td>$d &lt; \rho_1$</td>
<td>$d &gt; 0$</td>
<td>$d &gt; \rho_1^*$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1 = -\rho_2 &lt; 0$</td>
<td>$-\infty &lt; d &lt; +\infty$ or $d &gt; -\frac{\rho_2}{2a-1}$</td>
<td>$d &lt; -\frac{\rho_2}{2a-1}$</td>
<td>$d &gt; 0$</td>
<td>$d &lt; 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-\rho_2 &lt; \rho_1 &lt; 0$</td>
<td>$d &gt; \rho_2^*$</td>
<td>$d &lt; \rho_2^* &lt; 0$</td>
<td>$d &gt; 0$</td>
<td>$d &lt; 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1 = 0 &lt; \rho_2$</td>
<td>$d = 0$</td>
<td>$d &lt; 0$</td>
<td>$d &gt; 0$</td>
<td>$d &lt; 0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0 &lt; \rho_1 &lt; \rho_2$</td>
<td>$d = 0$</td>
<td>$d &lt; 0$</td>
<td>$d &gt; 0$</td>
<td>$d &lt; 0$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Exceptions:**

1. If $\delta = 0$ and $a = 1/2$, then $U(y) \to 0$ as $y \to +\infty$ when $d = 0$.

2. If there exist $d' < d'' < -\rho_2$ such that $b_1 = b_2$ for these values of $d$, then both limits are negative for $d' \leq d \leq d''$.

3. Results have not been obtained for a general mean vector $\hat{\delta}$. For four special mean vectors we have shown that

   If $\hat{\delta} = \delta U_1$ or $\delta U_2$ or $\delta \delta_1$, $\delta > 0$, then $\lim_{y \to -\infty} U(y) > 0$ and $\lim_{y \to +\infty} U(y)$ for all $d$.

   If $\hat{\delta} = 0$, then $\lim_{y \to +\infty} U(y) < 0$ for all $d$, and $\lim_{y \to -\infty} U(y) \leq 0$ according as $d < -\frac{\rho_2}{2a-1}$.

4. If there exist $-\rho_1 < d' < d''$ such that $b_1 = b_2$ for these values of $d$, then both limits are positive for $d' \leq d \leq d''$.

5. If $\delta_2 = 0$, then $\lim_{y \to -\infty} U(y) \leq 0$ according as $\alpha \leq 2/3$ when $d = 0$.

6. If $\delta = 0$ and $a = 1/2$, then $U(y) \to 0$ as $y \to +\infty$ when $d = 0$.

From Table 34 we can draw some rather interesting conclusions concerning the partitioning of the original $X$-plane into decision regions.
4.5.3 Geometrical Implications of the Limiting Values of $U(y)$.

It must be emphasized at the outset that knowledge of the behavior of $U(y)$ for large values of $|y|$ tells us nothing about the behavior of $U(y)$ for $y$ closer to zero. In particular, we must be concerned about the number of zeros of the function $U(y)$: although the graphs presented in Subsection 4.5.1 show $U(y)$ either as a decreasing function or as a function of increasing slope, it doesn't take much imagination to see the possibilities for the existence of local maxima and minima. The results of a numerical study of $U(y)$ for moderate-sized values of $y$ are given in the next subsection.

The limiting values of $U(y)$ do indicate, however, the nature of the partition of the $X$-plane into decision regions based on the function $Y = dx_1 + x_2$:

[1.] If the limits of $U(y)$ as $y \to +\infty$ and as $y \to -\infty$ are nonzero and of opposite sign, then the $X$-plane is partitioned into two half-planes by the line $x_2 = y^* - dx_1$, where $U(y^*) = 0$.

[2.] If both limits are positive, then the "choose Population 1" region is a strip between the parallel lines having slope $-d$ and $x_2$-intercepts $y^*$ and $y^0$, which are zeros of $U(y)$.

[3.] If both limits are negative, then the "choose Population 1" region is:

a.) the whole $X$-plane if $U(y)$ is strictly negative;
b.) the complement of the strip including the parallel lines,

\[ x_2 = y^* - dx_1 \quad \text{and} \quad x_2 = y^0 - dx_1, \]

where $y^*$ and $y^0$ are zeros of $U(y)$.

Any additional zeros in an interval about the origin must satisfy [4] or [5]:

[4.] If there is an odd number of these zeros, one of them must be a local maximum or minimum of $U(y)$. This zero does not affect the partition defined by the limiting values of $U(y)$, since it merely defines a line within a partition cell in such a way that all points in the cell that are not on the line remain in the original decision region. The points on the line may change from the "choose Population 1" region to the "choose Population 2" region, but there is zero probability attached to them. The remaining pairs of zeros satisfy [5].
[5.] If there is an even number of additional zeros, they define parallel lines which partition the cell into an odd number of strips: the first, third, etc. comprising the same decision region as the original cell, while the second, fourth, etc. unite with the unpartitioned cell(s) to comprise the alternative decision region.

Considering [1]-[5] and the entries in Table 34, we may conclude that when $k=0$, $\sigma_1=\sigma_2=1$, $\rho_1<\rho_2$, and . . .

[A.] $\rho_1^2>\rho_2^2$, then (cf. Figures 17 and 18 for examples of the relative positions of the two populations comprising the mixture) the optimal partition of the $x$-plane based on a family of lines with . . .

a.) non-negative slope (i.e. $d\leq 0$) is a simple two-cell partition, as in [1], or a refinement thereof, as in [5].

b.) negative slope (i.e. $d>0$) is a three-cell partition with the center strip being the "choose Population 1" region, as in [2], or a refinement thereof, as in [5].

[B.] $\rho_1^2<\rho_2^2$, then (cf. Figure 19) the optimal partition of the $x$-plane based on a family of lines with . . .

a.) non-positive slope (i.e. $d\geq 0$) is a simple two-cell partition, as in [1], or a refinement thereof, as in [5].

b.) positive slope (i.e. $d<0$) is a three-cell partition with the center strip being the "choose Population 2" region, as in [3], or a refinement thereof, as in [5].

[C.] $\rho_1^2=\rho_2^2$, then (cf. Figures 20 and 21) the optimal partition based on any family of lines is probably a simple two-cell partition as in [1], or a refinement. A definite statement cannot be made, as results were obtained only for four special mean vectors in this case.

To explore the problem of which kind of partition (i.e. a simple two-cell partition or something more complicated) is "better", in the sense that it minimizes the total probability of misclassification, is precisely the main goal of this chapter. A numerical study of the total probability of misclassification as a function of the slope (-d) and of the parameters $\mu$ is presented in Section 4.6.

Meanwhile, it would be useful to know more about the nature of the optimal partitions of the $x$-plane based on lines of slope (-d).
FIGURE IV.17: Contours such that, in the $i^{th}$ population, $i=1,2$, the probability density function $f_i(x|\kappa=0) = 0.15$, when $\sigma_1=\sigma_2=1, \rho_1=-2/3, \rho_2=-1/3,$ and $\delta$ is the mean vector given below.

\[ \delta = (0,0) \]

\[ \delta' = (1,0) \]

\[ \delta' = (0,1) \]

\[ \delta' = (1,1) \]
FIGURE IV.18: Contours such that, in the $i^{th}$ population, $i=1,2$, the probability density function $f_i(x|\delta=0)=0.15$, when $\sigma_1=\sigma_2=1$, $\rho_1=-2/3$, $\rho_2=1/3$, and $\delta$ is the mean vector given below.

\[ \delta = (0) \]

\[ \delta' = (1,0) \]

\[ \delta' = (0,1) \]

\[ \delta' = (1,1) \]
FIGURE IV.19: Contours such that, in the $i^{th}$ population, $i=1,2$, the probability density function $f_i(x|x=0)=0.15$, when $c_1=c_2=1$, $\rho_1=-1/3$, $\rho_2=+2/3$, and $\delta$ is the mean vector given below.

\[ \delta = (a, b) \]

\[ \delta' = (c, d) \]
FIGURE IV.20: Contours such that, in the $i$th population, $i=1,2$, the probability density function $f_i(x|\kappa=0)=0.15$, when $\sigma_1=\sigma_2=1$, $\rho_1=-1/3$, $\rho_2=+1/3$, and $\delta$ is the mean vector given below.

$\delta = (0,0)$

$\delta' = (1,0)$

$\delta' = (0,1)$

$\delta' = (1,1)$
FIGURE IV.21: Contours such that, in the $i$th population, $i=1,2$, the probability density function $f_i(x|\omega=0)=0.15$, when $\sigma_1=\sigma_2=1$, $\rho_1=-2/3$, $\rho_2=+2/3$, and $\delta$ is the mean vector given below.

\[
\begin{array}{c}
\begin{array}{c}
\text{Pop. 1}\\ \text{Pop. 2}
\end{array}
\end{array}
\]

$\delta = (0,0)$

\[
\begin{array}{c}
\begin{array}{c}
\text{Pop. 1}\\ \text{Pop. 2}
\end{array}
\end{array}
\]

$\delta = (1,0)$

\[
\begin{array}{c}
\begin{array}{c}
\text{Pop. 1}\\ \text{Pop. 2}
\end{array}
\end{array}
\]

$\delta = (0,1)$

\[
\begin{array}{c}
\begin{array}{c}
\text{Pop. 1}\\ \text{Pop. 2}
\end{array}
\end{array}
\]

$\delta = (1,1)$
4.5.4 A Numerical Study of U(y).

With limited funds available for purchasing computer time, it was impossible to conduct a thorough exploration of the effects of each parameter in \( \pi \) upon \( U(y) \) and the total probability of misclassification. At a time when some of the preceding theoretical results had not yet been obtained, a small numerical study was designed primarily to give some information about the effects of the various parameters on the misclassification probability. The parameter values chosen for the study were:

\[
\begin{align*}
&k = 0, 1 \\
&\delta = u_1, 1 \\
&\sigma_1 = \sigma_2 = 1/2, 1, 2 \\
&(\rho_1, \rho_2) = (-2/3, -1/3), (-2/3, +1/3), (-1/3, +2/3) \\
&a = 1/2
\end{align*}
\]

It was decided to use only the one value for alpha for the following reasons: 1.) the theory developed in this chapter suggested that the effects of the other parameters would not be as simple as the effect of alpha and that it would be wise to look at at least three relationships between \( \rho_1 \) and \( \rho_2; \) 2.) the results obtained for the untruncated case in Chapter III showed the importance of looking at variances less than, equal to, and greater than unity; 3.) the extensive numerical study of the effect of alpha in Chapter III consistently showed the minimum misclassification probability to decrease as alpha increased. In short, it seemed more essential to study 36 combinations of the other parameters than to eliminate a level of one of the other parameters in order to study two levels of alpha. To have simply added a second level of alpha would have made the study too large.

The two points, \( u_1 \) and \( 1 \), were chosen for the location of \( \delta \), the mean vector of the first population, because they simplified initial computations on an electronic calculator and because distributions centered at any two points having the same \( x_1 \)-coordinate are equally affected by truncation of the type studied here. Unfortunately, points having the same \( x_1 \)-coordinate are necessarily of unequal distance from the origin, which is the mean of the distribution in the second population.

On the following pages are given contour graphs of the two populations in each of the 36 cases of the study. The largest of a set of con-
centric "truncated ellipses" is the set of points $x$ such that the p.d.f.
evaluated at $x$ equals 0.01. The inner curves correspond to density func-
tion values of 0.05 and 0.15. (For details concerning the computation of
these curves, see the Appendix.)

The 36 cases of the study have been divided into six sets of six
cases each. Each set represents one combination of $\delta$ and $\sigma$. Following
the contour graphs for the six cases of a set is a table giving, for
these six cases, the "choose Population 1" region specified by the opti-
mal decision rule based on the linear function $y = dx_1 + x_2$ for specified
d-values and for $y$-values contained in intervals about the origin.

To visualize the optimal partition of the $x$-plane based on a given
linear function $y$, recall that the boundary lines of the decision re-
gions are the lines of slope $t = -d$ having the endpoints of the $y$-inter-
vals as $x_2$-intercepts.

In each column of the tables there appears either a circle about
an entry or an asterisk between entries. These indicate the value of $d$
(to the nearest tenth) which minimizes the misclassification probability
for the given set of parameters $\pi$: if this "optimal" $d$-value appears in
the $d$-column, the corresponding entry in the body of the table is cir-
cled; otherwise, an asterisk is placed between the entries in the table
 corresponding to the $d$-values closest to the "optimal" one, and the "op-
timal" $d$-value can be found in Tables 41-43.
FIGURE IV.22: Contours such that, in the $i^{th}$ population, $i=1,2$, the probability density function $f_i(x|\kappa) = 0.01, 0.05, \text{ and } 0.15$ when $\sigma_1 = \sigma_2 = 0.5$, $\delta^* = (1, 0)$, and $\rho_1$, $\rho_2$, and $\kappa$ are the values given below.
<table>
<thead>
<tr>
<th>(d)</th>
<th>(p_1 = -2/3)</th>
<th>(p_1 = -1/3)</th>
<th>(p_1 = -2/3)</th>
<th>(p_1 = -1/3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-5.0)</td>
<td>(0.23,2.96)</td>
<td>(0.26,2.86)</td>
<td>(-0.34,2.39)</td>
<td>(-0.49,0.63)</td>
</tr>
<tr>
<td>(-4.0)</td>
<td>(0.23,2.59)</td>
<td>(0.29,2.50)</td>
<td>(-0.30,2.03)</td>
<td>(-4.48,0.41)</td>
</tr>
<tr>
<td>(-3.0)</td>
<td>(0.25,2.24)</td>
<td>(0.30,1.78)</td>
<td>(-0.24,1.70)</td>
<td>(-3.46,0.18)</td>
</tr>
<tr>
<td>(-2.0)</td>
<td>(0.26,1.92)</td>
<td>(0.34,1.88)</td>
<td>(-0.12,1.41)</td>
<td>(-2.43,0.07)</td>
</tr>
<tr>
<td>(-1.0)</td>
<td>(0.29,1.65)</td>
<td>(0.37,1.78)</td>
<td>(-0.09,1.59)</td>
<td>(-1.38,0.38)</td>
</tr>
<tr>
<td>(-0.9)</td>
<td>(0.29,1.62)</td>
<td>(0.37,1.79)</td>
<td>(-0.09,1.69)</td>
<td>(-1.27,0.42)</td>
</tr>
<tr>
<td>(-0.8)</td>
<td>(0.29,1.60)</td>
<td>(0.35,1.80)</td>
<td>(-0.08,1.79)</td>
<td>(-1.16,0.46)</td>
</tr>
<tr>
<td>(-0.7)</td>
<td>(0.29,1.59)</td>
<td>(0.36,1.81)</td>
<td>(-0.07,1.87)</td>
<td>(-1.05,0.51)</td>
</tr>
<tr>
<td>(-0.6)</td>
<td>(0.29,1.57)</td>
<td>(0.35,1.81)</td>
<td>(-0.05,1.91)</td>
<td>(-0.94,0.55)</td>
</tr>
<tr>
<td>(-0.5)</td>
<td>(0.29,1.55)</td>
<td>(0.33,1.82)</td>
<td>(-0.02,1.94)</td>
<td>(-0.83,0.60)</td>
</tr>
<tr>
<td>(-0.4)</td>
<td>(0.29,1.54)</td>
<td>(0.32,1.82)</td>
<td>(-0.01,1.94)</td>
<td>(-0.72,0.65)</td>
</tr>
<tr>
<td>(-0.3)</td>
<td>(0.29,1.53)</td>
<td>(0.32,1.82)</td>
<td>(-0.01,1.94)</td>
<td>(-0.72,0.65)</td>
</tr>
<tr>
<td>(-0.2)</td>
<td>(0.28,1.53)</td>
<td>(0.32,1.81)</td>
<td>(-0.01,1.92)</td>
<td>(-0.61,0.71)</td>
</tr>
<tr>
<td>(-0.1)</td>
<td>(0.28,1.52)</td>
<td>(0.25,1.79)</td>
<td>(-0.11,1.90)</td>
<td>(-0.47,0.84)</td>
</tr>
<tr>
<td>(0.0)</td>
<td>(0.27,1.52)</td>
<td>(0.23,1.88)</td>
<td>(-0.15,1.88)</td>
<td>(-0.32,0.91)</td>
</tr>
<tr>
<td>(0.1)</td>
<td>(0.25,1.52)</td>
<td>(0.20,1.76)</td>
<td>(-0.18,1.87)</td>
<td>(-0.23,0.99)</td>
</tr>
<tr>
<td>(0.2)</td>
<td>(0.24,1.52)</td>
<td>(0.18,1.74)</td>
<td>(-0.22,1.85)</td>
<td>(-0.15,1.06)</td>
</tr>
<tr>
<td>(0.3)</td>
<td>(0.15,1.73)</td>
<td>(0.25,1.84)</td>
<td>(-0.07,1.14)</td>
<td>(-0.11,1.19)</td>
</tr>
<tr>
<td>(0.4)</td>
<td>(0.20,1.53)</td>
<td>(0.13,1.71)</td>
<td>(-0.26,1.83)</td>
<td>(-0.01,1.22)</td>
</tr>
<tr>
<td>(0.5)</td>
<td>(0.17,1.53)</td>
<td>(0.10,1.70)</td>
<td>(-0.32,1.82)</td>
<td>(-0.05,1.31)</td>
</tr>
<tr>
<td>(0.6)</td>
<td>(0.15,1.53)</td>
<td>(0.07,1.69)</td>
<td>(-0.35,1.81)</td>
<td>(0.11,1.39)</td>
</tr>
<tr>
<td>(0.7)</td>
<td>(0.12,1.53)</td>
<td>(0.05,1.67)</td>
<td>(-0.36,1.80)</td>
<td>(0.16,1.48)</td>
</tr>
<tr>
<td>(0.8)</td>
<td>(0.09,1.53)</td>
<td>(0.02,1.67)</td>
<td>(-0.41,1.80)</td>
<td>(0.20,1.57)</td>
</tr>
<tr>
<td>(0.9)</td>
<td>(0.07,1.54)</td>
<td>(-0.01,1.66)</td>
<td>(-0.44,1.79)</td>
<td>(0.25,1.66)</td>
</tr>
<tr>
<td>(1.0)</td>
<td>(0.04,1.54)</td>
<td>(-0.04,1.65)</td>
<td>(-0.47,1.78)</td>
<td>(0.29,1.75)</td>
</tr>
<tr>
<td>(2.0)</td>
<td>(-0.27,1.55)</td>
<td>(-0.33,1.61)</td>
<td>(-0.78,1.75)</td>
<td>(0.64,2.74)</td>
</tr>
<tr>
<td>(3.0)</td>
<td>(-0.58,1.55)</td>
<td>(-0.65,1.59)</td>
<td>(-1.08,1.74)</td>
<td>(0.93,3.74)</td>
</tr>
<tr>
<td>(4.0)</td>
<td>(-0.90,1.55)</td>
<td>(-0.97,1.59)</td>
<td>(-1.41,1.73)</td>
<td>(1.18,4.74)</td>
</tr>
<tr>
<td>(5.0)</td>
<td>(-1.22,1.55)</td>
<td>(-1.30,1.58)</td>
<td>(-1.75,1.72)</td>
<td>(1.41,5.74)</td>
</tr>
</tbody>
</table>
Contours such that, in the $i^{th}$ population, $i=1,2$, the probability density function $f_i(x_i) = 0.01, 0.05,$ and $0.15$ when $a_i = 0.5, b_i = (1.1)$, and $c_i = 0.1$, $x_i$, and $y$ are the values given below.

$\beta_i = -\frac{1}{3}$
$\beta_i = +\frac{1}{3}$
TABLE IV.36: The optimal decision rule based on $Y = dX_1 + X_2$:
"Choose Population 1 if $Y$ is in the designated open interval."

<table>
<thead>
<tr>
<th>$k=0$</th>
<th>$k=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1 = -2/3$</td>
<td>$\rho_1 = -1/3$</td>
</tr>
<tr>
<td>$d$</td>
<td>$\rho_2 = -1/3$</td>
</tr>
<tr>
<td>-5.0</td>
<td>(1.33, 4.10)</td>
</tr>
<tr>
<td>-4.0</td>
<td>(1.33, 3.75)</td>
</tr>
<tr>
<td>-3.0</td>
<td>(1.34, 3.42)</td>
</tr>
<tr>
<td>-2.0</td>
<td>(1.34, 3.14)</td>
</tr>
<tr>
<td>-1.0</td>
<td>(1.31, 2.98)</td>
</tr>
<tr>
<td>-0.9</td>
<td>(1.30, 2.98)</td>
</tr>
<tr>
<td>-0.8</td>
<td>(1.29, 2.98)</td>
</tr>
<tr>
<td>-0.7</td>
<td>(1.28, 2.98)</td>
</tr>
<tr>
<td>-0.6</td>
<td>(1.27, 2.99)</td>
</tr>
<tr>
<td>-0.5</td>
<td>(1.26, 3.00)</td>
</tr>
<tr>
<td>-0.4</td>
<td>(1.24, 3.01)</td>
</tr>
<tr>
<td>-0.3</td>
<td>(1.22, 3.02)</td>
</tr>
<tr>
<td>-0.2</td>
<td>(1.20, 3.03)</td>
</tr>
<tr>
<td>-0.1</td>
<td>(1.18, 3.04)</td>
</tr>
<tr>
<td>0.0</td>
<td>(1.16, 3.05)</td>
</tr>
<tr>
<td>0.1</td>
<td>(1.13, 3.06)</td>
</tr>
<tr>
<td>0.2</td>
<td>(1.10, 3.06)</td>
</tr>
<tr>
<td>0.3</td>
<td>(1.07, 3.06)</td>
</tr>
<tr>
<td>0.4</td>
<td>(1.05, 3.06)</td>
</tr>
<tr>
<td>0.5</td>
<td>(1.02, 3.05)</td>
</tr>
<tr>
<td>0.6</td>
<td>(0.99, 3.05)</td>
</tr>
<tr>
<td>0.7</td>
<td>(0.96, 3.05)</td>
</tr>
<tr>
<td>0.8</td>
<td>(0.93, 3.05)</td>
</tr>
<tr>
<td>0.9</td>
<td>(0.90, 3.05)</td>
</tr>
<tr>
<td>1.0</td>
<td>(0.87, 3.05)</td>
</tr>
</tbody>
</table>

* indicates $d < 0$
Figure IV.24: Contours such that, in the $i$th population, $i=1,2$, the probability density function $f_{i}(x(k)) = 0.01, 0.05,$ and $0.15$ when $\alpha_{1} = -2/3, \beta_{1} = (0,0)$, and $\alpha_{2} = -1/3, \beta_{2} = (0,0)$ are the values given below.
TABLE IV: The optimal decision rule based on $Y_0X_1F_1$: "Choose Population 0
tf Y is in the designated open interval or union of open intervals."

<table>
<thead>
<tr>
<th>$k=0$</th>
<th>$p_1 = -2/3$</th>
<th>$p_1 = -1/3$</th>
<th>$p_1 = -1/3$</th>
<th>$p_1 = -2/3$</th>
<th>$p_1 = -1/3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$p_2 = -1/3$</td>
<td>$p_2 = 1/3$</td>
<td>$p_2 = 1/3$</td>
<td>$p_2 = -1/3$</td>
<td>$p_2 = 1/3$</td>
</tr>
<tr>
<td>-5.0</td>
<td>(0.53, 4.92)</td>
<td>(0.74, 4.88)</td>
<td>(0.74, 4.88)</td>
<td>(0.53, 4.92)</td>
<td>(0.74, 4.88)</td>
</tr>
<tr>
<td>-4.0</td>
<td>(0.57, 4.38)</td>
<td>(0.83, 4.78)</td>
<td>(0.83, 4.78)</td>
<td>(0.57, 4.38)</td>
<td>(0.83, 4.78)</td>
</tr>
<tr>
<td>-3.0</td>
<td>(0.62, 3.93)</td>
<td>(0.97, 3.96)</td>
<td>(0.97, 3.96)</td>
<td>(0.62, 3.93)</td>
<td>(0.97, 3.96)</td>
</tr>
<tr>
<td>-2.0</td>
<td>(0.69, 3.79)</td>
<td>(1.10, 3.80)</td>
<td>(1.10, 3.80)</td>
<td>(0.69, 3.79)</td>
<td>(1.10, 3.80)</td>
</tr>
<tr>
<td>-1.0</td>
<td>(0.73, 3.70)</td>
<td>(0.90, 3.73)</td>
<td>(0.90, 3.73)</td>
<td>(0.73, 3.70)</td>
<td>(0.90, 3.73)</td>
</tr>
<tr>
<td>-0.9</td>
<td>(0.72, 3.69)</td>
<td>(0.65, 3.69)</td>
<td>(0.65, 3.69)</td>
<td>(0.72, 3.69)</td>
<td>(0.65, 3.69)</td>
</tr>
<tr>
<td>-0.8</td>
<td>(0.72, 3.69)</td>
<td>(0.60, 3.69)</td>
<td>(0.60, 3.69)</td>
<td>(0.72, 3.69)</td>
<td>(0.60, 3.69)</td>
</tr>
<tr>
<td>-0.7</td>
<td>(0.70, 3.69)</td>
<td>(0.75, 3.69)</td>
<td>(0.75, 3.69)</td>
<td>(0.70, 3.69)</td>
<td>(0.75, 3.69)</td>
</tr>
<tr>
<td>-0.6</td>
<td>(0.69, 3.69)</td>
<td>(0.69, 3.69)</td>
<td>(0.69, 3.69)</td>
<td>(0.69, 3.69)</td>
<td>(0.69, 3.69)</td>
</tr>
<tr>
<td>-0.5</td>
<td>(0.65, 3.69)</td>
<td>(0.63, 3.69)</td>
<td>(0.63, 3.69)</td>
<td>(0.65, 3.69)</td>
<td>(0.63, 3.69)</td>
</tr>
<tr>
<td>-0.4</td>
<td>(0.64, 3.69)</td>
<td>(0.57, 3.69)</td>
<td>(0.57, 3.69)</td>
<td>(0.64, 3.69)</td>
<td>(0.57, 3.69)</td>
</tr>
<tr>
<td>-0.3</td>
<td>(0.60, 3.69)</td>
<td>(0.51, 3.69)</td>
<td>(0.51, 3.69)</td>
<td>(0.60, 3.69)</td>
<td>(0.51, 3.69)</td>
</tr>
<tr>
<td>-0.2</td>
<td>(0.57, 3.69)</td>
<td>(0.44, 3.69)</td>
<td>(0.44, 3.69)</td>
<td>(0.57, 3.69)</td>
<td>(0.44, 3.69)</td>
</tr>
<tr>
<td>-0.1</td>
<td>(0.52, 3.69)</td>
<td>(0.36, 3.69)</td>
<td>(0.36, 3.69)</td>
<td>(0.52, 3.69)</td>
<td>(0.36, 3.69)</td>
</tr>
<tr>
<td>0.0</td>
<td>(0.48, 3.69)</td>
<td>(0.31, 3.69)</td>
<td>(0.31, 3.69)</td>
<td>(0.48, 3.69)</td>
<td>(0.31, 3.69)</td>
</tr>
<tr>
<td>0.1</td>
<td>(0.42, 3.69)</td>
<td>(0.25, 3.69)</td>
<td>(0.25, 3.69)</td>
<td>(0.42, 3.69)</td>
<td>(0.25, 3.69)</td>
</tr>
<tr>
<td>0.2</td>
<td>(0.38, 3.69)</td>
<td>(0.19, 3.69)</td>
<td>(0.19, 3.69)</td>
<td>(0.38, 3.69)</td>
<td>(0.19, 3.69)</td>
</tr>
<tr>
<td>0.3</td>
<td>(0.31, 3.69)</td>
<td>(0.12, 3.69)</td>
<td>(0.12, 3.69)</td>
<td>(0.31, 3.69)</td>
<td>(0.12, 3.69)</td>
</tr>
<tr>
<td>0.4</td>
<td>(0.25, 3.69)</td>
<td>(0.06, 3.69)</td>
<td>(0.06, 3.69)</td>
<td>(0.25, 3.69)</td>
<td>(0.06, 3.69)</td>
</tr>
<tr>
<td>0.5</td>
<td>(0.18, 3.69)</td>
<td>(0.01, 3.69)</td>
<td>(0.01, 3.69)</td>
<td>(0.18, 3.69)</td>
<td>(0.01, 3.69)</td>
</tr>
<tr>
<td>0.6</td>
<td>(0.12, 3.69)</td>
<td>(0.07, 3.69)</td>
<td>(0.07, 3.69)</td>
<td>(0.12, 3.69)</td>
<td>(0.07, 3.69)</td>
</tr>
<tr>
<td>0.7</td>
<td>(0.05, 3.69)</td>
<td>(0.03, 3.69)</td>
<td>(0.03, 3.69)</td>
<td>(0.05, 3.69)</td>
<td>(0.03, 3.69)</td>
</tr>
<tr>
<td>0.8</td>
<td>(0.02, 3.69)</td>
<td>(0.01, 3.69)</td>
<td>(0.01, 3.69)</td>
<td>(0.02, 3.69)</td>
<td>(0.01, 3.69)</td>
</tr>
<tr>
<td>0.9</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
<tr>
<td>1.0</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
<tr>
<td>3.0</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
<tr>
<td>4.0</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
<tr>
<td>5.0</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
<td>(0.00, 3.69)</td>
</tr>
</tbody>
</table>
FIGURE IV.25: Contours such that, in the \( i \)th population, \( i=1,2 \), the probability density function
\[
f_i(x|\kappa) = 0.01, 0.05, \text{ and } 0.15 \quad \text{when } \sigma_1 = \sigma_2 = 1.0, \quad \delta_i = (1,1), \quad \text{and } \rho_1, \rho_2, \text{ and } \kappa \text{ are the}
\text{values given below.}
\]
<table>
<thead>
<tr>
<th>( k=0 )</th>
<th>( \alpha = 0.1 )</th>
<th>( \alpha = 0.05 )</th>
<th>( \alpha = 0.01 )</th>
<th>( \alpha = 0.001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>( \gamma_1 = -1/3 )</td>
<td>( \gamma_2 = 1/3 )</td>
<td>( \gamma_1 = -1/3 )</td>
<td>( \gamma_2 = 1/3 )</td>
</tr>
<tr>
<td>0.1</td>
<td>(1.075,1.11)</td>
<td>(1.075,1.11)</td>
<td>(1.075,1.11)</td>
<td>(1.075,1.11)</td>
</tr>
<tr>
<td>0.2</td>
<td>(1.001,1.25)</td>
<td>(1.001,1.25)</td>
<td>(1.001,1.25)</td>
<td>(1.001,1.25)</td>
</tr>
<tr>
<td>0.3</td>
<td>(0.961,1.13)</td>
<td>(0.961,1.13)</td>
<td>(0.961,1.13)</td>
<td>(0.961,1.13)</td>
</tr>
<tr>
<td>0.4</td>
<td>(0.999,0.60)</td>
<td>(0.999,0.60)</td>
<td>(0.999,0.60)</td>
<td>(0.999,0.60)</td>
</tr>
<tr>
<td>0.5</td>
<td>(0.894,0.99)</td>
<td>(0.894,0.99)</td>
<td>(0.894,0.99)</td>
<td>(0.894,0.99)</td>
</tr>
<tr>
<td>0.6</td>
<td>(0.778,1.15)</td>
<td>(0.778,1.15)</td>
<td>(0.778,1.15)</td>
<td>(0.778,1.15)</td>
</tr>
<tr>
<td>0.7</td>
<td>(0.718,0.92)</td>
<td>(0.718,0.92)</td>
<td>(0.718,0.92)</td>
<td>(0.718,0.92)</td>
</tr>
<tr>
<td>0.8</td>
<td>(0.655,0.99)</td>
<td>(0.655,0.99)</td>
<td>(0.655,0.99)</td>
<td>(0.655,0.99)</td>
</tr>
<tr>
<td>0.9</td>
<td>(0.338,3.53)</td>
<td>(0.338,3.53)</td>
<td>(0.338,3.53)</td>
<td>(0.338,3.53)</td>
</tr>
<tr>
<td>1.0</td>
<td>(0.022,8.97)</td>
<td>(0.022,8.97)</td>
<td>(0.022,8.97)</td>
<td>(0.022,8.97)</td>
</tr>
<tr>
<td>2.0</td>
<td>(0.000,8.27)</td>
<td>(0.000,8.27)</td>
<td>(0.000,8.27)</td>
<td>(0.000,8.27)</td>
</tr>
<tr>
<td>3.0</td>
<td>(0.000,9.56)</td>
<td>(0.000,9.56)</td>
<td>(0.000,9.56)</td>
<td>(0.000,9.56)</td>
</tr>
<tr>
<td>4.0</td>
<td>(0.000,10.55)</td>
<td>(0.000,10.55)</td>
<td>(0.000,10.55)</td>
<td>(0.000,10.55)</td>
</tr>
<tr>
<td>5.0</td>
<td>(0.000,11.54)</td>
<td>(0.000,11.54)</td>
<td>(0.000,11.54)</td>
<td>(0.000,11.54)</td>
</tr>
</tbody>
</table>
FIGURE IV.26: Contours such that, in the ith population, $i=1,2$, the probability density function $f_i(x_i | \theta_i) = 0.01$, 0.05, and 0.15 when $a_i=0, b_i=1, c_i=0$, and $d_i=1/2$. The contours are given below.

\[ \theta_i = \frac{1}{2} \]

\[ \theta_i = -\frac{1}{2} \]
### TABLE IV.39: The optimal decision rule based on $Y = dX_1 + X_2$: "Choose Population 1 iff $Y$ is not in the designated closed interval or half-plane."

<table>
<thead>
<tr>
<th>$k = 0$</th>
<th>$k = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1 = -2/3$</td>
<td>$\rho_1 = -2/3$</td>
</tr>
<tr>
<td>$\rho_2 = -1/3$</td>
<td>$\rho_2 = -2/3$</td>
</tr>
<tr>
<td>$d$</td>
<td>$\rho_2 = +1/3$</td>
</tr>
<tr>
<td>-5.0</td>
<td>$[-1.67, 6.86]$</td>
</tr>
<tr>
<td>-4.0</td>
<td>$[-1.69, 5.56]$</td>
</tr>
<tr>
<td>-3.0</td>
<td>$[-1.71, 4.32]$</td>
</tr>
<tr>
<td>-2.0</td>
<td>$[-1.76, 3.22]$</td>
</tr>
<tr>
<td>-1.0</td>
<td>$[-1.86, 2.24]$</td>
</tr>
<tr>
<td>-0.9</td>
<td>$[-1.87, 2.15]$</td>
</tr>
<tr>
<td>-0.8</td>
<td>$[-1.89, 2.05]$</td>
</tr>
<tr>
<td>-0.7</td>
<td>$[-1.90, 1.95]$</td>
</tr>
<tr>
<td>-0.6</td>
<td>$[-1.92, 1.86]$</td>
</tr>
<tr>
<td>-0.5</td>
<td>$[-1.95, 1.76]$</td>
</tr>
<tr>
<td>-0.4</td>
<td>$[-1.97, 1.66]$</td>
</tr>
<tr>
<td>-0.3</td>
<td>$[-2.00, 1.56]$</td>
</tr>
<tr>
<td>-0.2</td>
<td>$[-2.03, 1.46]$</td>
</tr>
<tr>
<td>-0.1</td>
<td>$[-2.07, 1.36]$</td>
</tr>
<tr>
<td>0.0</td>
<td>$[-2.11, 1.26]$</td>
</tr>
<tr>
<td>0.1</td>
<td>$[-2.15, 1.16]$</td>
</tr>
<tr>
<td>0.2</td>
<td>$[-2.21, 1.06]$</td>
</tr>
<tr>
<td>0.3</td>
<td>$[-2.26, 0.96]$</td>
</tr>
<tr>
<td>0.4</td>
<td>$[-2.32, 0.87]$</td>
</tr>
<tr>
<td>0.5</td>
<td>$[-2.39, 0.78]$</td>
</tr>
<tr>
<td>0.6</td>
<td>$[-2.47, 0.69]$</td>
</tr>
<tr>
<td>0.7</td>
<td>$[-2.54, 0.55]$</td>
</tr>
<tr>
<td>0.8</td>
<td>$[-2.61, 0.54]$</td>
</tr>
<tr>
<td>0.9</td>
<td>$[-2.68, 0.47]$</td>
</tr>
<tr>
<td>1.0</td>
<td>$[-2.75, 0.41]$</td>
</tr>
<tr>
<td>2.0</td>
<td>$[-3.36, 0.15]$</td>
</tr>
<tr>
<td>3.0</td>
<td>$[-4.39, 0.11]$</td>
</tr>
<tr>
<td>4.0</td>
<td>$[-5.61, 0.09]$</td>
</tr>
<tr>
<td>5.0</td>
<td>$[-6.87, 0.09]$</td>
</tr>
</tbody>
</table>
$p_1 = -\frac{2}{3}$
$p_2 = -\frac{1}{3}$

$\kappa = 0$

Pop. 1

$\kappa = 1$

Pop. 2

Contours such that, in the $i^{th}$ population, $f_{i-1,2}(x)$, the probability density function.

$\sigma^2_i = 2.0$, $s_1 = 1.1$, and $p_1, p_2$, and $\kappa$ are the values given below.
TABLE IV: The optimal decision rule based on $Y = AX_1 + X_2$: "Choose Population 1 iff $Y$ is not in the designated interval or half-line."

<table>
<thead>
<tr>
<th>$k=0$</th>
<th>$k=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1 = -2/3$</td>
<td>$\rho_1 = -1/3$</td>
</tr>
<tr>
<td>$\delta'=(1,1)$</td>
<td>$\sigma_1=\sigma_2=2$</td>
</tr>
<tr>
<td>$d$</td>
<td>$\rho_2 = -1/3$</td>
</tr>
<tr>
<td>-3.0</td>
<td>$[-3.19, 4.06]$</td>
</tr>
<tr>
<td>-1.0</td>
<td>$[-3.29, 2.36]$</td>
</tr>
<tr>
<td>-0.9</td>
<td>$[-3.30, 2.27]$</td>
</tr>
<tr>
<td>-0.8</td>
<td>$[-3.31, 2.18]$</td>
</tr>
<tr>
<td>-0.7</td>
<td>$[-3.32, 2.09]$</td>
</tr>
<tr>
<td>-0.6</td>
<td>$[-3.34, 2.00]$</td>
</tr>
<tr>
<td>-0.5</td>
<td>$[-3.36, 1.91]$</td>
</tr>
<tr>
<td>-0.4</td>
<td>$[-3.38, 1.81]$</td>
</tr>
<tr>
<td>-0.3</td>
<td>$[-3.40, 1.72]$</td>
</tr>
<tr>
<td>-0.2</td>
<td>$[-3.43, 1.62]$</td>
</tr>
<tr>
<td>-0.1</td>
<td>$[-3.46, 1.52]$</td>
</tr>
<tr>
<td>0</td>
<td>$(-[3.50, 1.43])$</td>
</tr>
<tr>
<td>0.1</td>
<td>$[-3.55, 1.33]$</td>
</tr>
<tr>
<td>0.2</td>
<td>$[-3.61, 1.23]$</td>
</tr>
<tr>
<td>0.3</td>
<td>$[-3.69, 1.14]$</td>
</tr>
<tr>
<td>0.4</td>
<td>$[-3.79, 1.04]$</td>
</tr>
<tr>
<td>0.5</td>
<td>$[-3.90, 0.95]$</td>
</tr>
<tr>
<td>0.6</td>
<td>$[-4.05, 0.86]$</td>
</tr>
<tr>
<td>0.7</td>
<td>$[-4.22, 0.77]$</td>
</tr>
<tr>
<td>0.8</td>
<td>$[-4.42, 0.68]$</td>
</tr>
<tr>
<td>0.9</td>
<td>$[-4.62, 0.60]$</td>
</tr>
<tr>
<td>1.0</td>
<td>$[-4.79, 0.52]$</td>
</tr>
<tr>
<td>2.0</td>
<td>$[-5.01, 0.04]$</td>
</tr>
<tr>
<td>3.0</td>
<td>$[-5.69, -0.13]$</td>
</tr>
<tr>
<td>4.0</td>
<td>$[-6.74, -0.21]$</td>
</tr>
<tr>
<td>5.0</td>
<td>$[-7.92, -0.25]$</td>
</tr>
</tbody>
</table>
There were two questions raised implicitly in the last section regarding the optimal partition based on a family of lines having slope \( t \). On the basis of this numerical study we can answer them.

1.) Does a nontrivial three-cell partition ever occur? Yes indeed!

Nearly all the partitions given for the first set of cases, i.e. when \( \sigma=1/2 \) and \( \delta=u_1 \), are nontrivial three-cell partitions. So too are many based on families of positive slope whenever \( (\rho_1, \rho_2) = (-1/3, +2/3) \), particularly when \( k=1 \). However, by plotting the regional boundaries on the graphs, one can readily see that many of the three-cell partitions listed in the tables have one cell that appears to have very small probability under both populational distributions; e.g. the case when \( \delta = 1, \sigma=1/2, (\rho_1, \rho_2) = (-1/3, +2/3) \), and \( \theta=1 \).

2.) Do "refinements" of the partitions based on the limiting values of \( U(y) \) ever occur? That is, can \( U(y) \) have more than two zeros? Yes is the answer to this question too. The optimal partition for the family of lines with slope \( +1 \) has five nontrivial cells when \( \sigma=1, \delta=u_1, k=0, \) and \( (\rho_1, \rho_2) = (-1/3, +2/3) \). When \( k=1 \), the optimal partition based on families with slope near \( +1 \) has four cells when \( \sigma=1, \delta=u_1 \), and \( (\rho_1, \rho_2) = (-2/3, \pm 1/3) \).

Whether any of these fancier partitions is much better than a simple two-cell partition is the real question, of course. This small study seems to suggest that this is so whenever one of the populations is more sprawling than the other, particularly if \( \delta_2 \) is definitely smaller than \( \delta_1 \).

4.6 The Probability of Misclassification.

The ultimate purpose of this chapter is to give an acceptable answer to the question: For a given set of known parameters

\[
\pi' = (\delta', \sigma_1, \sigma_2, \rho_1, \rho_2, \alpha, k),
\]

how can we find the linear partition of the \( x \)-plane which has the smallest total probability of misclassification?

4.6.1 A General Procedure for Finding the Optimal Linear Partition.

In the previous sections we have shown that we can find the optimal partition based on the family of lines having slope \( t \), say. The method, not too difficult or expensive to use with the aid of a computer, is to define the linear function \( y = dx_1 + x_2 \), where \( d = -t \), and then
[1.] evaluate the function \( U(y) = Q(y) + G(y) \), defined in equations (5), at a number of points in the interval where \( G(y) \) can be computed;

[2.] find the zeros of \( U(y) \) in this interval;

[3.] identify the intervals in which \( U(y) \) is negative.

The zeros of \( U(y) \) are the \( x_2 \)-intercepts of the lines bounding the decision regions, and the portions of the \( x \)-plane comprising the "choose Population 1" region are those strips or half-planes containing the intervals identified in [3] located on the \( x_2 \)-axis.

Now, among all the possible partitions of the \( x \)-plane based on the family of lines having slope \( t \), the one obtained by the above method has the smallest total probability of misclassification. However, a family of lines having a different slope may produce a partition having a smaller misclassification probability. Consequently, we consider the probability of misclassification corresponding to the optimal partition based on the family of lines of slope \( t \) as a function of \( t \) and evaluate it over a range of \( t \)-values in order to find the slope which produces the smallest value.

[4.] Define

\[
F(t \mid \pi) = a \Pr_1\{U(y) \geq 0\} + (1-a) \Pr_2\{U(y) < 0\},
\]

the probability of misclassification corresponding to the optimal partition based on a family of lines of slope \( t \), where \( y = x_2 - tx_1 \) and \( \Pr_i\{E\} \) is the probability that event \( E \) occurs in Population \( i \), \( i = 1, 2 \). The set of points \( \{y: U(y) < 0\} \) is the union of the intervals identified in [3], and \( \{y: U(y) \geq 0\} \) is its complement, also a union of intervals. From equations (1), (3), and (4) at the beginning of this chapter we have

\[
F(t \mid \pi) = a \int_{\{y: U(y) \geq 0\}} \frac{\exp\left(-\frac{(y-y)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi}\sigma_1} \frac{\phi(a_1 - b_1 y)}{\phi(k)} \, dy +
\]

\[
+ (1-a) \int_{\{y: U(y) < 0\}} \frac{\exp\left(-\frac{y^2}{2\sigma_2^2}\right)}{\sqrt{2\pi}\sigma_2} \frac{\phi(a_2 - b_2 y)}{\phi(k)} \, dy
\]

which can be evaluated on a computer.

A method of finding the linear partition of the \( x \)-plane having the smallest total probability of misclassification, then, is to evaluate \( F(t \mid \pi) \) via steps [1]-[4] over a range of \( t \)-values, say the interval \([-5, +5]\),
and study the results to see where the minimum is likely to be.

4.6.2 Numerical Studies of the "Misclassification Probability" Function.

Two small numerical studies were designed to investigate the effects of each parameter in \( \pi \) upon the function \( F(t|\pi) \). The first of these, which we'll call the General Parameter Study, has already been described in Subsection 4.5.4; the other, hereafter called the Truncation Study, was designed to compare the changes in \( F(t|\pi) \) as the mean or variance increased when \( k=0 \) and when \( k=\infty \), i.e. compare the behavior of \( F(t|\pi) \) under maximum truncation and under no truncation at all.

Recall that there are two figures presented in Chapter III to illustrate how the shape of \( F(d|\pi) \) changes as \( \delta_1=\delta_2=\delta \) increases and as \( \sigma_1=\sigma_2=\sigma \) increases, where \( F(d|\pi) \) is the misclassification probability corresponding to the optimal decision rule based on the function \( Y=X_1+X_2 \) when there is no truncation. Note that in the present chapter the linear function with which we're working is \( Y=dX_1+X_2 \), where \( d = -t \); but for populations with distributions such that \( \delta_1=\delta_2 \) and \( \sigma_1=\sigma_2 \), the "misclassification probability" functions, \( F(d|\pi) \), in the two chapters are completely comparable.

To see, then, how truncation affects the behavior of \( F(d|\pi) \), it was decided to use the same parameter values used to produce Figures 5 and 6 of Chapter III:
a.) to study the effect of \( \delta \):
\[
\begin{align*}
\delta &= 0.001, 0.60, 1.00, 1.50 \\
\sigma &= 2.1 \\
(\rho_1, \rho_2) &= (0.25, 0.75) \\
\alpha &= 0.65
\end{align*}
\]
b.) to study the effect of \( \sigma \):
\[
\begin{align*}
\delta &= 0.4 \\
\sigma &= 0.1, 0.5, 0.9, 1.5, 2.2, 9.0 \\
(\rho_1, \rho_2) &= (0.25, 0.75) \\
\alpha &= 0.65
\end{align*}
\]
The value \( k=0 \) was chosen on the theory that the maximum truncation would have the most damaging effect upon the total probability of misclassification.
4.6.3 Results of the General Parameter Study.

On the following pages are three tables giving the values of $F(d | \pi)$ for selected values of $d$ for the 36 cases of the study. Following each table is a figure containing the graphs of $F(d | \pi)$ corresponding to the six cases in the table with $\delta = \mu_1$. If we let $d^*$ denote the value of $d$ (to the nearest tenth) which minimizes $F(d | \pi)$ for a given set of parameters $\pi$, we can summarize the results of the study as follows:

1.) $F(d^* | \pi)$ is considerably smaller when $\delta = \bar{\mu}_1$ than when $\delta = \mu_1$ for all 18 combinations of the other parameters. This might well be explained by the fact that the means of the distributions of the two populations are closer together—making the populations harder to distinguish—when $\delta = \mu_1$ than when $\delta = \bar{\mu}_1$.

2.) $F(d^* | \pi)$ is decidedly smaller when $k = 0$ than when $k = 1$ for all 18 combinations of the other parameters. This might be explained by the fact that the strip $\{x: 0 < x_1 < 1\}$, which lies between the two population means and which therefore has relatively high probability under both distributions, is not observed when $k = 0$ but is observed when $k = 1$. As this strip undoubtedly contributes heavily to the error rate, it is advantageous to exclude it from the region of observation.

3.) $F(d^* | \pi)$ increases as $\sigma$ increases for all eight combinations having $\rho_1 = -2/3$ and $\rho_2 = 1/3$.

4.) $F(d^* | \pi)$ is larger for $\sigma = 1$ than for $\sigma = 1/2$ or $\sigma = 2$ in three of the four combinations having $(\rho_1, \rho_2) = (-1/3, +2/3)$.

5.) $F(d | \pi)$ does not change very rapidly in an interval about $d^*$: $F(d | \pi)$ evaluated at the multiple of 0.5 nearest $d^*$ differs from $F(d^* | \pi)$ only in the third or fourth decimal place.

6.) When $\sigma = 1/2$ and $\sigma = 1$, the largest $F(d^* | \pi)$ over the three choices of $(\rho_1, \rho_2)$ occurs when $\rho_2 = 2/3$ for all four combinations of $\delta$ and $k$.

7.) When $\sigma = 2$, the largest $F(d^* | \pi)$ over the three choices of $(\rho_1, \rho_2)$ occurs when $(\rho_1, \rho_2) = (-2/3, -1/3)$.

8.) $d^*$ is in the interval $\{[1, 3.1]$ when $\sigma = 1/2\}$ and rounds off to $\{(0, 2.2]$ when $\sigma = 1\}$ $\{(-1, 0]$ when $\sigma = 2\}$
one of the values,

\[
\begin{array}{l}
\{ 1.0, 1.5, 2.0 \text{ when } \sigma=1/2 \} \\
\{ 0.5, 1.0, 1.5 \text{ when } \sigma=1 \} \\
\{ -1.0, -0.5, 0.0 \text{ when } \sigma=2 \} \\
\end{array}
\]

in at least ten of the twelve cases for each variance value.

We shall draw no further conclusions until after seeing the results of the Truncation Study.
<table>
<thead>
<tr>
<th>(d_i)</th>
<th>4/(d_i)</th>
<th>k_0</th>
<th>k_1</th>
<th>4/(d_i)</th>
<th>k_0</th>
<th>k_1</th>
<th>4/(d_i)</th>
<th>k_0</th>
<th>k_1</th>
<th>4/(d_i)</th>
<th>k_0</th>
<th>k_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.1193</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.2</td>
<td>1.1392</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.3</td>
<td>1.1503</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.4</td>
<td>1.1605</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.5</td>
<td>1.1702</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.6</td>
<td>1.1803</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.7</td>
<td>1.1905</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.8</td>
<td>1.2003</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>0.9</td>
<td>1.2102</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
<tr>
<td>1.0</td>
<td>1.2202</td>
<td>0.1193</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
<td>0.0530</td>
<td>0.0346</td>
</tr>
</tbody>
</table>
### Figure IV.28

<table>
<thead>
<tr>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\sigma$</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$\alpha$</th>
<th>$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_2'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
<tr>
<td>$\pi_3'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_4'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
<tr>
<td>$\pi_5'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_6'$</td>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{3}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
</tbody>
</table>

The graph shows various curves with labels $F(d | \pi_6)$, $F(d | \pi_4)$, $F(d | \pi_2)$, $F(d | \pi_5)$, and $F(d | \pi_3)$.
<table>
<thead>
<tr>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\alpha$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pi_1''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$-\frac{\sqrt{3}}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pi_2''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$-\frac{\sqrt{3}}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pi_3''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pi_4''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pi_5''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Pi_6''$</td>
<td>$1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>$1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>( \delta'=(1,0) )</td>
<td>( \sigma_1=0.2, \sigma_2=2 )</td>
<td>( \alpha=1/2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( k=0 )</td>
<td>( k=1 )</td>
<td>( k=0 )</td>
<td>( k=1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \rho_1=-2/3 )</td>
<td>( \rho_1=-2/3 )</td>
<td>( \rho_1=-2/3 )</td>
<td>( \rho_1=-2/3 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5.0</td>
<td>.3592</td>
<td>.3306</td>
<td>.3239</td>
<td>.4068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4.0</td>
<td>.3515</td>
<td>.3163</td>
<td>.3070</td>
<td>.4014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.0</td>
<td>.3386</td>
<td>.2949</td>
<td>.2816</td>
<td>.3932</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2.0</td>
<td>.3217</td>
<td>.2638</td>
<td>.2454</td>
<td>.3792</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.0</td>
<td>.2991</td>
<td>.2296</td>
<td>.2140</td>
<td>.3528</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.9</td>
<td>.2970</td>
<td>.2272</td>
<td>( .2138 )</td>
<td>( .3494 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>.2951</td>
<td>.2253</td>
<td>.2146</td>
<td>.3458</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.7</td>
<td>.2935</td>
<td>.2239</td>
<td>.2165</td>
<td>.3423</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>.2921</td>
<td>.2231</td>
<td>.2195</td>
<td>.3388</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>.2911</td>
<td>( .2230 )</td>
<td>.2236</td>
<td>.3355</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>.2904</td>
<td>.2236</td>
<td>.2289</td>
<td>.3324</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.3</td>
<td>.2894</td>
<td>.2249</td>
<td>.2354</td>
<td>.3298</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>.2890</td>
<td>.2271</td>
<td>.2428</td>
<td>.3278</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.1</td>
<td>.2891</td>
<td>.2300</td>
<td>.2512</td>
<td>.3264</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>.2941</td>
<td>.2339</td>
<td>.2604</td>
<td>.3307</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>.2969</td>
<td>.2386</td>
<td>.2703</td>
<td>.3265</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>.3007</td>
<td>.2442</td>
<td>.2806</td>
<td>.3284</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>.3055</td>
<td>.2507</td>
<td>.2912</td>
<td>.3315</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>.3111</td>
<td>.2581</td>
<td>.3019</td>
<td>.3358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>.3177</td>
<td>.2663</td>
<td>.3125</td>
<td>.3412</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>.3249</td>
<td>.2753</td>
<td>.3228</td>
<td>.3476</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>.3327</td>
<td>.2850</td>
<td>.3326</td>
<td>.3548</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>.3408</td>
<td>.2952</td>
<td>.3419</td>
<td>.3623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>.3499</td>
<td>.3057</td>
<td>.3505</td>
<td>.3698</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.3568</td>
<td>.3162</td>
<td>.3584</td>
<td>.3771</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>.3988</td>
<td>.3999</td>
<td>.4006</td>
<td>.4168</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>.4039</td>
<td>.4265</td>
<td>.4093</td>
<td>.4241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>.4034</td>
<td>.4273</td>
<td>.4102</td>
<td>.4257</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>.4023</td>
<td>.4242</td>
<td>.4093</td>
<td>.4260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6.4 Results of the Truncation Study.

On the following pages are two tables giving the values of $F(d|\pi)$ for selected values of $d$ for the four cases of the $\delta$-study and for the six cases of the $\sigma$-study, respectively. After each table is a figure presenting the graphs of $F(d|\pi)$ corresponding to the cases of the table. Letting $d^*$ denote the value of $d$ minimizing $F(d|\pi)$ for a given set of parameters $\pi$, we can summarize the results found in these tables as follows:

1.) As $\delta$ increases, the general shape of $F(d|\pi)$ does not change, but $F(d^*|\pi)$ decreases and the maximum value of $F(d|\pi)$ increases.

2.) $d^*$ virtually does not change as $\delta$ increases.

3.) The general shape of $F(d|\pi)$ changes rapidly with $\sigma$.

4.) In general, $F(d|\pi)$ increases as $\sigma \uparrow$ and decreases as $1<\sigma^+$. In particular, $F(d^*|\pi)$ increases as $\sigma \uparrow$ and decreases as $1<\sigma^+$.

5.) $d^*$ decreases as $\sigma$ increases. For $\sigma<1$, $d^*$ is positive; for $\sigma>1$, $d^*$ is negative.

6.) If $d'$ denotes the multiple of 0.5 closest to $d^*$, then $F(d'|\pi)$ differs from $F(d^*|\pi)$ only in the third or fourth decimal place.


<table>
<thead>
<tr>
<th>$d$</th>
<th>$\delta = 0.001$</th>
<th>$\delta = 0.60$</th>
<th>$\delta = 1.00$</th>
<th>$\delta = 1.50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.0</td>
<td>.2969</td>
<td>.3058</td>
<td>.3081</td>
<td>.3075</td>
</tr>
<tr>
<td>-4.0</td>
<td>.2878</td>
<td>.2946</td>
<td>.2957</td>
<td>.2936</td>
</tr>
<tr>
<td>-3.0</td>
<td>.2747</td>
<td>.2779</td>
<td>.2769</td>
<td>.2721</td>
</tr>
<tr>
<td>-2.0</td>
<td>.2563</td>
<td>.2536</td>
<td>.2488</td>
<td>.2394</td>
</tr>
<tr>
<td>-1.9</td>
<td>.2543</td>
<td>.2508</td>
<td>.2456</td>
<td>.2356</td>
</tr>
<tr>
<td>-1.8</td>
<td>.2523</td>
<td>.2481</td>
<td>.2423</td>
<td>.2317</td>
</tr>
<tr>
<td>-1.7</td>
<td>.2504</td>
<td>.2454</td>
<td>.2391</td>
<td>.2279</td>
</tr>
<tr>
<td>-1.6</td>
<td>.2487</td>
<td>.2428</td>
<td>.2360</td>
<td>.2242</td>
</tr>
<tr>
<td>-1.5</td>
<td>.2472</td>
<td>.2405</td>
<td>.2331</td>
<td>.2206</td>
</tr>
<tr>
<td>-1.4</td>
<td>.2460</td>
<td>.2384</td>
<td>.2304</td>
<td>.2173</td>
</tr>
<tr>
<td>-1.3</td>
<td>.2453</td>
<td>.2368</td>
<td>.2282</td>
<td>.2144</td>
</tr>
<tr>
<td>-1.2</td>
<td>.2451</td>
<td>.2357</td>
<td>.2266</td>
<td>.2120</td>
</tr>
<tr>
<td>-1.1</td>
<td>.2457</td>
<td>.2355</td>
<td>.2256</td>
<td>.2103</td>
</tr>
<tr>
<td>-1.0</td>
<td>.2472</td>
<td>.2360</td>
<td>.2282</td>
<td>.2093</td>
</tr>
<tr>
<td>-0.9</td>
<td>.2497</td>
<td>.2377</td>
<td>.2267</td>
<td>.2097</td>
</tr>
<tr>
<td>-0.8</td>
<td>.2533</td>
<td>.2405</td>
<td>.2288</td>
<td>.2111</td>
</tr>
<tr>
<td>-0.7</td>
<td>.2580</td>
<td>.2446</td>
<td>.2324</td>
<td>.2138</td>
</tr>
<tr>
<td>-0.6</td>
<td>.2638</td>
<td>.2499</td>
<td>.2372</td>
<td>.2177</td>
</tr>
<tr>
<td>-0.5</td>
<td>.2704</td>
<td>.2564</td>
<td>.2432</td>
<td>.2229</td>
</tr>
<tr>
<td>-0.4</td>
<td>.2776</td>
<td>.2638</td>
<td>.2503</td>
<td>.2292</td>
</tr>
<tr>
<td>-0.3</td>
<td>.2850</td>
<td>.2718</td>
<td>.2582</td>
<td>.2356</td>
</tr>
<tr>
<td>-0.2</td>
<td>.2923</td>
<td>.2802</td>
<td>.2668</td>
<td>.2447</td>
</tr>
<tr>
<td>-0.1</td>
<td>.2991</td>
<td>.2885</td>
<td>.2756</td>
<td>.2534</td>
</tr>
<tr>
<td>0.0</td>
<td>.3051</td>
<td>.2966</td>
<td>.2844</td>
<td>.2624</td>
</tr>
<tr>
<td>0.1</td>
<td>.3103</td>
<td>.3041</td>
<td>.2930</td>
<td>.2715</td>
</tr>
<tr>
<td>0.2</td>
<td>.3145</td>
<td>.3108</td>
<td>.3010</td>
<td>.2805</td>
</tr>
<tr>
<td>0.3</td>
<td>.3178</td>
<td>.3167</td>
<td>.3085</td>
<td>.2892</td>
</tr>
<tr>
<td>0.4</td>
<td>.3204</td>
<td>.3218</td>
<td>.3152</td>
<td>.2975</td>
</tr>
<tr>
<td>0.5</td>
<td>.3223</td>
<td>.3261</td>
<td>.3211</td>
<td>.3052</td>
</tr>
<tr>
<td>0.6</td>
<td>.3237</td>
<td>.3297</td>
<td>.3262</td>
<td>.3124</td>
</tr>
<tr>
<td>0.7</td>
<td>.3246</td>
<td>.3326</td>
<td>.3306</td>
<td>.3188</td>
</tr>
<tr>
<td>0.8</td>
<td>.3253</td>
<td>.3350</td>
<td>.3344</td>
<td>.3246</td>
</tr>
<tr>
<td>0.9</td>
<td>.3257</td>
<td>.3369</td>
<td>.3375</td>
<td>.3297</td>
</tr>
<tr>
<td>1.0</td>
<td>.3259</td>
<td>.3385</td>
<td>.3401</td>
<td>.3341</td>
</tr>
<tr>
<td>2.0</td>
<td>.3253</td>
<td>.3446</td>
<td>.3497</td>
<td>.3500</td>
</tr>
<tr>
<td>3.0</td>
<td>.3250</td>
<td>.3458</td>
<td>.3500</td>
<td>.3500</td>
</tr>
<tr>
<td>4.0</td>
<td>.3256</td>
<td>.3463</td>
<td>.3500</td>
<td>.3500</td>
</tr>
<tr>
<td>5.0</td>
<td>.3266</td>
<td>.3466</td>
<td>.3500</td>
<td>.3500</td>
</tr>
</tbody>
</table>
TABLE IV.45: Values of \( F(d|k) \) for Selected Values of \( d \).

<table>
<thead>
<tr>
<th>( k = 0 )</th>
<th>( k = 0.4 )</th>
<th>( k = 0.25 )</th>
<th>( k = 0.75 )</th>
<th>( k = 0.65 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta = 0.1 )</td>
<td>( \delta = 0.5 )</td>
<td>( \delta = 0.9 )</td>
<td>( \delta = 1.5 )</td>
<td>( \delta = 2.2 )</td>
</tr>
<tr>
<td>0.1</td>
<td>.0430</td>
<td>.2555</td>
<td>.3500</td>
<td>.3470</td>
</tr>
<tr>
<td>0.2</td>
<td>.0473</td>
<td>.2650</td>
<td>.3500</td>
<td>.3434</td>
</tr>
<tr>
<td>0.3</td>
<td>.0557</td>
<td>.2811</td>
<td>.3500</td>
<td>.3351</td>
</tr>
<tr>
<td>0.4</td>
<td>.0749</td>
<td>.3100</td>
<td>.3500</td>
<td>.3162</td>
</tr>
<tr>
<td>0.5</td>
<td>.0779</td>
<td>.3135</td>
<td>.3500</td>
<td>.3136</td>
</tr>
<tr>
<td>0.6</td>
<td>.0811</td>
<td>.3168</td>
<td>.3500</td>
<td>.3108</td>
</tr>
<tr>
<td>0.7</td>
<td>.0845</td>
<td>.3197</td>
<td>.3500</td>
<td>.3080</td>
</tr>
<tr>
<td>0.8</td>
<td>.0881</td>
<td>.3221</td>
<td>.3500</td>
<td>.3051</td>
</tr>
<tr>
<td>0.9</td>
<td>.0918</td>
<td>.3235</td>
<td>.3500</td>
<td>.3023</td>
</tr>
<tr>
<td>1.0</td>
<td>.0954</td>
<td>.3236</td>
<td>.3500</td>
<td>.2997</td>
</tr>
<tr>
<td>1.1</td>
<td>.0988</td>
<td>.3219</td>
<td>.3500</td>
<td>.2973</td>
</tr>
<tr>
<td>1.2</td>
<td>.1016</td>
<td>.3183</td>
<td>.3500</td>
<td>.2953</td>
</tr>
<tr>
<td>1.3</td>
<td>.1034</td>
<td>.3127</td>
<td>.3500</td>
<td>.2939</td>
</tr>
<tr>
<td>1.4</td>
<td>.1038</td>
<td>.3052</td>
<td>.3477</td>
<td>.2934</td>
</tr>
<tr>
<td>1.5</td>
<td>.1025</td>
<td>.2960</td>
<td>.3439</td>
<td>.2938</td>
</tr>
<tr>
<td>1.6</td>
<td>.0994</td>
<td>.2855</td>
<td>.3393</td>
<td>.2953</td>
</tr>
<tr>
<td>1.7</td>
<td>.0947</td>
<td>.2741</td>
<td>.3343</td>
<td>.2981</td>
</tr>
<tr>
<td>1.8</td>
<td>.0888</td>
<td>.2623</td>
<td>.3291</td>
<td>.3021</td>
</tr>
<tr>
<td>1.9</td>
<td>.0825</td>
<td>.2507</td>
<td>.3240</td>
<td>.3073</td>
</tr>
<tr>
<td>2.0</td>
<td>.0555</td>
<td>.2059</td>
<td>.3045</td>
<td>.3409</td>
</tr>
<tr>
<td>2.1</td>
<td>.0494</td>
<td>.1952</td>
<td>.3005</td>
<td>.3495</td>
</tr>
<tr>
<td>2.2</td>
<td>.0429</td>
<td>.1878</td>
<td>.2982</td>
<td>.3500</td>
</tr>
<tr>
<td>2.3</td>
<td>.0386</td>
<td>.1831</td>
<td>.2972</td>
<td>.3500</td>
</tr>
<tr>
<td>2.4</td>
<td>.0368</td>
<td>.1816</td>
<td>.2976</td>
<td>.3500</td>
</tr>
<tr>
<td>2.5</td>
<td>.0353</td>
<td>.1805</td>
<td>.2971</td>
<td>.3500</td>
</tr>
<tr>
<td>2.6</td>
<td>.0339</td>
<td>.1798</td>
<td>.2973</td>
<td>.3500</td>
</tr>
<tr>
<td>2.7</td>
<td>.0326</td>
<td>.1794</td>
<td>.2977</td>
<td>.3500</td>
</tr>
<tr>
<td>2.8</td>
<td>1.1872</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>1.1793</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>1.1795</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>1.1799</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>1.1804</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>.0255</td>
<td>.1838</td>
<td>.3050</td>
<td>.3500</td>
</tr>
<tr>
<td>3.4</td>
<td>.0232</td>
<td>.1911</td>
<td>.3120</td>
<td>.3500</td>
</tr>
<tr>
<td>3.5</td>
<td>.02248</td>
<td>.1966</td>
<td>.3170</td>
<td>.3500</td>
</tr>
<tr>
<td>3.6</td>
<td>.022349</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>.022348</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>.022353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>.02237</td>
<td>.2007</td>
<td>.3205</td>
<td>.3500</td>
</tr>
</tbody>
</table>
In the following figures the graph of \( F(d|k=0) \) is plotted against the graph of \( F(d|k=\infty) \) for each of the ten cases of the Truncation Study. If we let \( d^0 \) and \( d^* \) denote the values of \( d \) minimizing \( F(d|k=0) \) and \( F(d|k=\infty) \) respectively, we can state these results:

7.) \( F(d^0|k=0) < F(d^*|k=\infty) \) in all ten cases! This observation tends to support the theory proposed in the last subsection, i.e. that it is advantageous not to observe the strip between the two means, \( \delta \) and 0.

8.) \( F(d|k=0) \) and \( F(d|k=\infty) \) have the same general shape and behave in the same way as \( \delta \) or \( \sigma \) changes. However, the former spreads out more than the latter.

9.) With one exception, \( d^0 \) is very close to \( d^* \); specifically, \( |d^0-d^*| \leq 0.5 \). In the exceptional case, \( F(d|k=0) \) is so flat that a \( d \)-value can be found which is within 0.5 units of \( d^* \) and whose corresponding value of \( F(d|k=0) \) differs from \( F(d^0|k=0) \) only in the third decimal place.
Figure IV.34

\[\begin{array}{cccccc}
\delta = 0.4 & \xi_1 = 0.1 & \xi_2 = 0.25 & \xi_3 = 0.75 & \xi_4 = 0.65 & \xi_5 = 0.0 \\
\pi_1' &= 0.4 & 0.1 & 0.25 & 0.75 & 0.65 & 0 \\
\pi_2' &= 0.4 & 0.1 & 0.25 & 0.75 & 0.65 & \infty \\
\pi_3' &= 0.4 & 0.9 & 0.25 & 0.75 & 0.65 & 0 \\
\pi_4' &= 0.4 & 0.9 & 0.25 & 0.75 & 0.65 & 0 \\
\pi_5' &= 0.4 & 2.2 & 0.25 & 0.75 & 0.65 & 0 \\
\pi_6' &= 0.4 & 2.2 & 0.25 & 0.75 & 0.65 & \infty \\
\end{array}\]
4.6.5 Conclusions Drawn from the Numerical Studies.

Letting $d^*$ denote the value of $d$ which minimizes the probability of misclassification $F(d|\pi)$ for a given set of parameters $\pi$, we state the following conclusions:

A.) Slope of the optimal linear partition of the $x$-plane.

1.) The general location of $d^*$ depends almost solely on $\sigma$. The "fine tuning" of the location depends on $\delta$, $k$, and the relative values of $\rho_1$ and $\rho_2$.

2.) $d^*$ is probably positive or negative according as $\sigma<1$ or $\sigma>1$. That is, the slope of the optimal linear partition of the $x$-plane is probably negative if $\sigma<1$ and positive if $\sigma>1$.

A possible explanation is that the optimal partition tends to isolate the more compact (i.e. less spread out) population in the smallest possible region and gives the rest of the plane to the more sprawling population. Now, when $\sigma<1$, Population 1 is more compact than Population 2; and since the mean of Population 1 is in Quadrant I while Population 2 is centered at the origin, a descending line passing between the two means will be an effective partition boundary. When $\sigma>1$, Population 2 is more compact, and an effective partition will assign to Population 2 a strip between two ascending parallel lines, leaving the rest of the plane to Population 1.

3.) Since $F(d|\pi)$ seems to be relatively flat in a small neighborhood of $d^*$ in all 46 cases of the studies, it probably is good enough to use any slope within 0.3 units of ($-d^*$), e.g. the multiple of 0.5 nearest ($-d^*$).

4.) As a consequence of (2) and (3), we can design our computer program to evaluate $F(d|\pi)$ at positive multiples of 0.1 up to 3.0, say, when $\sigma<1$ and at negative multiples of 0.1 down to $-3.0$ when $\sigma>1$, in order to find $d^*$, regardless of the values of the other parameters.

B.) The probability of misclassification.

1.) Truncation seems to decrease the probability of misclassification! That is, $\textit{maximum}$ truncation, which removes the region between the two mean vectors, is beneficial. Whether partial
truncation is helpful or harmful is an open question which could not even be guessed at on the basis of the 46 test cases.

2.) There are suggestions that when $|\rho_1| > |\rho_2|$, the minimum probability of misclassification increases as $\sigma$ increases. But when $|\rho_2| > |\rho_1|$, the minimum probability of misclassification decreases as $\sigma$ either increases or decreases from 1. These observations seem to support the explanation offered above in (A)-(2); for when $\rho_1 < \rho_2$ and $|\rho_1| > |\rho_2|$, then Population 1 is distributed in a relatively long, narrow elliptical region with major diameter on a descending line, whereas Population 2 is distributed in a more circular region; conversely, when $|\rho_1| < |\rho_2|$ and $\rho_1 < \rho_2$, then Population 2 must be distributed in a relatively long, narrow elliptical region with major diameter on an ascending line, while Population 1 is distributed in a more circular region.

3.) There is some indication that the size of the minimum probability of misclassification depends heavily on the location of $\hat{\xi}$, not just in its distance from the origin but also in its direction.
CHAPTER V
PLANS FOR FURTHER STUDY

Further study in the area of linear discriminant analysis under truncation will necessarily be of two different types: 1.) completion of projects and theoretical development left unfinished in this dissertation; and 2.) generalization to observations of higher dimension, more complicated forms of truncation, and less restrictive assumptions.

5.1 Unfinished Business.

In Chapter II the major unfinished business is to do a numerical study to compare the performances of the two decision rules (i.e. the Optimal Decision Rule and the decision rule based on Fisher's "best linear" discriminant function) for a variety of choices of parameter sets. Since this study must be done on a computer, it can be designed to give information about the ways in which the linear boundary for each decision rule depends on the various parameters. This information can then be compared to similar information collected for the case of unequal covariance matrices studied in Chapter IV.

To complete the work begun in Chapter III, a numerical investigation of the optimal partition of the $x$-plane (i.e. the regions of the $x$-plane assigned to each population according to the optimal decision rule) should be done. Comparisons could then be made with the optimal two-cell partitions developed by Anderson & Bahadur [2], Clunies-Ross & Riffenburgh [6], etc.

For Chapter IV a much larger numerical study must be done in order to discover the nature of the interaction between the truncation effect and the effect of the relative positions of the two population means and to gain a deeper understanding of the importance of the factors influencing the position of the partition boundaries.

For all three chapters the theoretical results should be obtained for a general $\theta$, no longer restricted to Quadrant I.
5.2 Generalizations.

The most important generalization is to extend the results to include observations of more than two components. For the case of rectangular truncation on a single component, say $x_1$, the extension to higher dimensions should follow in straightforward fashion from the theory developed in Sections 2.1 and 4.1. Because, in the final analysis, the optimal decision rule had to be found and studied by means of numerical evaluation in the case of unequal covariance matrices, we simply expect to investigate the optimal decision rule in higher dimensions by means of a computer. Also, the assumptions, $\delta_i \geq 0$ for all $i = 1, \ldots, n$, and $k = 0$, which served to simplify some of the theoretical calculations, can be omitted in strictly numerical studies.

Finding the optimal linear discriminant function $d'x$ when $x$ has $n$ components involves finding the absolute minimum of a function $F(d)$ of $n-1$ variables. Since the optimal decision rule based on any linear function $d'x$ necessarily partitions $n$-space by means of a set of parallel hyperplanes, the numerical and theoretical results obtained in Chapters II and IV suggest that a valuable first step in the procedure to find the optimal linear discriminant function in higher dimensions is to consider carefully the geometry of the two populational distributions, looking for sets of parallel hyperplanes that contain the more compact population between them and include relatively little of the other population. The coordinates which these parallel hyperplanes have in common will serve as a reasonable starting point in the procedure to find the minimum of the "misclassification probability" function. The program to compute the total probability of misclassification will necessarily calculate the optimal decision rule based on a specified linear function.

The next step in generalizing the theory of linear discriminant analysis under truncation might be to consider rectangular truncation on more than one component, e.g. $x$ cannot be observed unless

$$x \in \{x: x_1 \leq c_1, x_2 \leq c_2, \ldots, x_j \leq c_j\}$$

for some $j$ between 1 and $n$. 

APPENDIX

COMPUTATION OF TRUNCATED ELLIPSES

For a given fixed constant C and any truncation value \( k=0 \),

\[ f_2(x|k) \geq C \text{ iff } \begin{align*}
\text{(a) } & x_1 \leq k \\
\text{(b) } & x_1^2 - 2\rho_2 x_1 x_2 + x_2^2 \leq -2(1-\rho_2^2)\ln[2\pi C\sqrt{1-\rho_2^2} \phi(k)] \\
& \equiv (1-\rho_2^2) \cdot C_1(\rho_2, C, k).
\end{align*} \]

If, in addition, we assume \( \sigma_1=\sigma_2=\sigma > 0 \), then

\[ f_1(x|k) \geq C \text{ iff } \begin{align*}
\text{(a) } & x_1 \leq k \\
\text{(b) } & x_1^2 - 2\rho_1 x_1 x_2 + x_2^2 \leq -2\sigma^2(1-\rho_1^2)\ln[2\pi C\sqrt{1-\rho_1^2} \phi(k/\sigma)] \\
& \equiv (1-\rho_1^2) \cdot C_2(\rho_1, C, \sigma, k).
\end{align*} \]

\[ f_1(x|k) \geq C \text{ iff } \begin{align*}
\text{(a) } & x_1 \leq k \\
\text{(b) } & (x_1-\delta)^2 - 2\rho_1 (x_1-\delta) x_2 + x_2^2 \leq -2\sigma^2(1-\rho_1^2)\ln[2\pi C\sqrt{1-\rho_1^2} \phi(k/\sigma)] \\
& \equiv (1-\rho_1^2) \cdot C_3(\rho_1, C, \sigma, k, \delta).
\end{align*} \]

\[ f_1(x|k) \geq C \text{ iff } \begin{align*}
\text{(a) } & x_1 \leq k \\
\text{(b) } & x_1^2 - 2\rho_1 x_1 (x_2-\delta) + (x_2-\delta)^2 \leq -2\sigma^2(1-\rho_1^2)\ln[2\pi C\sqrt{1-\rho_1^2} \phi(k/\sigma)] \\
& \equiv (1-\rho_1^2) \cdot C_2(\rho_1, C, \sigma, k).
\end{align*} \]

To transform an equation of the form, \( x^2 - 2\rho xy + y^2 = (1-\rho^2) \cdot C \), into graphable form, rotate the coordinate axes \( 45^\circ \). In the new system the equation is \( (1-\rho)u^2 + (1+\rho)v^2 = (1-\rho^2) \cdot C \), i.e.

\[
\frac{u^2}{(1-\rho) \cdot C} + \frac{v^2}{(1+\rho) \cdot C} = 1.
\]

When \( k=0 \) and \( \sigma=1 \), the coefficients, \( C_1-C_3 \), are greatly simplified:

\[
\begin{align*}
C_1(\rho, C) &= -2\ln(\pi C\sqrt{1-\rho^2}) \\
C_2(\rho, C) &= C_1(\rho, C) \\
C_3(\rho, C, \delta) &= C_1(\rho, C) + 2\ln[\phi(0)/\phi(-\delta)].
\end{align*}
\]
BIBLIOGRAPHY


