SOME HYPOTHESIS TESTING PROBLEMS INVOLVING MULTIVARIATE NORMAL DISTRIBUTIONS WITH UNEQUAL AND INTRACLASS STRUCTURED COVARIANCE MATRICES

By

Sheldon James Press

TECHNICAL REPORT NO. 12

June 4, 1964

PREPARED UNDER THE AUSPICES OF

NATIONAL SCIENCE FOUNDATION GRANT GP-214

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
TABLE OF CONTENTS

ACKNOWLEDGMENT ................................................. 111

CHAPTER

0. SUMMARY AND INTRODUCTION ................................. 1

1. SOME PROBLEMS ASSOCIATED WITH THE BEHRENS-FISHER
   PROBLEM .................................................. 3

   1.1 Introduction ............................................ 3

   1.2 Multivariate Behrens-Fisher Problems ................. 5

       1.2.1 Intraclace Structure, General Mean ........... 5

       1.2.2 Intraclace Structure, Restricted Mean ....... 11

       1.2.3 Functionally Dependent Covariances, 
            General Means ................................... 12

       1.2.4 Functionally Dependent Covariances, 
            Restricted Means .................................. 13

   1.3 Optimality of the Multivariate Scheffé 
       Transformation Test Procedure ........................ 15

   1.4 Behrens-Fisher Problems in Multivariate 
       Regression ............................................. 17

   1.5 A Confidence Interval Comparison of Two 
       Test Procedures Proposed for the Univariate 
       Behrens-Fisher Problem ................................ 22

       1.5.1 Criterion for Relative Test 
            Efficiency ......................................... 23

       1.5.2 Evaluation of Expectations .................... 26

       1.5.3 Discussion of Numerical Results .............. 29

       1.5.4 Large Sample Results ............................ 37

2. COMPARATIVE MULTIVARIATE REGRESSION WITH EQUAL 
   COVARIANCE MATRICES ..................................... 39

   2.1 Introduction ............................................ 39

   2.2 Testing with Equal Design Matrices .................. 39

   2.3 Testing with Unequal Design Matrices ............... 41

       2.3.1 Development of the Likelihood 
            Ratio Statistic .................................. 41

       2.3.2 Distribution Theory .............................. 43
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>DICHOTOMOUS CLASSIFICATION INTO NORMAL POPULATIONS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>Classification - General Covariance</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Matrices, $\Sigma_1 &gt; \Sigma_2$</td>
<td>47</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Likelihood Ratio Discriminators</td>
<td>47</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Large Sample Procedures -</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Unknown Parameters</td>
<td></td>
</tr>
<tr>
<td>3.2.3</td>
<td>Distribution Theory -</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Known Parameters</td>
<td></td>
</tr>
<tr>
<td>3.2.4</td>
<td>Determination of the Classification</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Regions</td>
<td></td>
</tr>
<tr>
<td>3.2.5</td>
<td>Approximate Distribution Theory -</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Known Parameters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Method I: Cumulant Matching</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Method II: Edgeworth Series</td>
<td>60</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Approximate Distribution Theory -</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Unknown Parameters</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Classification with Intraclass Structure,</td>
<td>69</td>
</tr>
<tr>
<td>$\Sigma_1 &gt; \Sigma_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3.1</td>
<td>Equal Covariance Matrices -</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Known Parameters</td>
<td></td>
</tr>
<tr>
<td>3.3.2</td>
<td>Equal Covariance Matrices -</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Unknown Parameters</td>
<td></td>
</tr>
<tr>
<td>3.3.3</td>
<td>Unequal Covariance Matrices -</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Known Parameters</td>
<td></td>
</tr>
<tr>
<td>3.3.4</td>
<td>Unequal Covariance Matrices -</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Unknown Parameters</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>Admissible Linear Procedures with Intraclass</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td></td>
</tr>
<tr>
<td>3.4.1</td>
<td>Introduction</td>
<td>82</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Procedures with Intraclass Structure</td>
<td>83</td>
</tr>
<tr>
<td>3.5</td>
<td>Classification with Unrestricted Covariance</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Matrices</td>
<td></td>
</tr>
<tr>
<td>CHAPTER</td>
<td>LINEAR COMBINATIONS OF NON-CENTRAL DISTRIBUTIONS</td>
<td>91</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>4.1 Introduction</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>4.2 Distribution of a Positive Sum of Non-Central Chi-Squared Variates</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>4.3 Distribution of Ratios</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>4.4 Probabilistic Interpretation</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>4.5 Numerical Computation and Approximations for Positive Definite Forms</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>4.6 Indefinite Quadratic Forms in Non-Central Normal Variates</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>4.6.1 Exact Distributions</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>4.6.2 Approximate Percentage Points</td>
<td>114</td>
</tr>
<tr>
<td>5.</td>
<td>DISTRIBUTION OF A SAMPLE COVARIANCE</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>5.1 Introduction</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>5.2 Derivation of the Distribution</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>5.3 Properties of the Distribution</td>
<td>124</td>
</tr>
<tr>
<td>6.</td>
<td>EXTENSIONS OF RESULTS</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>6.1 Introduction</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>6.2.1 Circulant Covariance Matrices</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>6.2.2 Applications</td>
<td>130</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>132</td>
</tr>
</tbody>
</table>
0. SUMMARY AND INTRODUCTION

This report is concerned with tests of hypotheses involving multivariate normal distributions. The problems examined are of three types: (1) tests of equality of population means under various assumptions of the parameter spaces, (2) tests of equality of regression parameters for two regressions on the same phenomenon, (3) tests of hypotheses concerned with classification of observations, under various assumptions of the covariance matrices, and parameters.

The first type of problem (Behrens-Fisher) is discussed in Chapter 1. A transformation developed by Scheffé [27] and extended by Bennett [8] is applied to the data and likelihood ratio procedures are then developed. Also, two multivariate regressions with unequal error covariance matrices are compared using the same type of procedure. Finally, the two test procedures developed by Scheffé and Welch [35] for the univariate Behrens-Fisher problem are compared according to the expected length of their associated confidence intervals and a criterion for application of these procedures is developed.

In the second type of problem treated (Chapter 2), it is assumed that the error covariance matrices of two independent regressions on the same variables are equal. Likelihood ratio procedures are developed for both the equal and unequal design matrices cases.

The classification procedures developed in Chapter 3 are principally likelihood ratio tests. The two populations are assumed
to be multivariate normal, and various assumptions are made concerning the covariance matrices, \( \Sigma_1, \Sigma_2 \). Special cases treated are:

- \( \Sigma_1 - \Sigma_2 \) positive definite; \( \Sigma_1 \) and \( \Sigma_2 \) with intraclass structure but \( \Sigma_1 - \Sigma_2 \) is positive definite (see Section 1.2.1); \( \Sigma_1 = \Sigma_2 \) and each has intraclass structure; \( \Sigma_1 \neq \Sigma_2 \) and non-singular, but otherwise arbitrary. In addition, explicit linear procedures are obtained for the case of intraclass structured covariance matrices using the method of Anderson and Bahadur [3].

The distribution theory associated with the classification statistics is developed in Chapter 4. In that chapter, distribution functions are derived for linear combinations of independent, non-central chi-squared variates. Both positive definite, and indefinite quadratic forms are considered; i.e., the cases in which weighting coefficients of the variates are all positive, and when they are of arbitrary sign, are treated separately. Simple approximations are developed for the tail probabilities since the exact expressions are complicated.

In Chapter 5 the distribution of the sample covariance statistic from a bivariate normal distribution is derived. The moments of the distribution are obtained and some of its properties are presented.

In Chapter 6 directions are given for how the results obtained in previous chapters can be extended to the class of positive definite circulant matrices, and the principal properties of this class are outlined.
CHAPTER 1

SOME PROBLEMS ASSOCIATED WITH THE BEHRENS-FISHER PROBLEM

1.1 Introduction.

Let \( X, Y \) be independent \( p \times m \) and \( p \times n \) random matrices whose columns are independently and identically distributed as \( N(\mu_1, \Sigma_1) \), \( N(\mu_2, \Sigma_2) \), respectively. The problem of testing whether or not \( \mu_1 = \mu_2 \) when \( \Sigma_1 \neq \Sigma_2 \) is called the Behrens-Fisher problem, for the work of Behrens [6] and Fisher [15], when \( p = 1 \). Alternative statistical procedures for \( p = 1 \) were proposed by Scheffé [27], Welch [35], and others. In this chapter, the multivariate analogue of the Scheffé procedure is used to develop likelihood ratio tests for the specially structured Behrens-Fisher problems described below. Also, a numerical comparison of the expected lengths of the confidence intervals is made for the Scheffé and Welch procedures.

In Section 1.2, several problems involve the case in which the covariance matrices have "intraclass structure"; i.e., the case in which all the variances are equal but unknown, and also, all the covariances are equal but unknown. The covariance matrix with intraclass structure has application in psychology, experimental design, and in the spacial or temporal correlation applications of communication theory; e.g., Abramson [1]. Problems involving the intraclass structured matrix have been treated by Wilks [37], Voit [33] and others. The more general Behrens-Fisher problem with arbitrary covariance matrix is also considered for the cases when the covariance matrices are functionally related and when the means are restricted.
In Section 1.3 we establish a previously unproved optimality property of the multivariate version of the Scheffé procedure, which is the analogue of the univariate optimality property established by Scheffé.

Section 1.4 is concerned with the multivariate regression problem. It is assumed that two separate regression analyses have been carried out and that we want to test the equality of the parameter vectors. Moreover, for reasons of cost or physical limitation, the design matrices for the two analyses differ and the experiments are subject to different errors, causing the analyses to have unequal covariance matrices. The optimality property established in Section 1.3 is also applicable to these problems.

Section 1.5 involves the univariate problem. The Scheffé procedure is compared with the Welch procedure using the ratio of the expected lengths of their associated confidence intervals. On the basis of a numerical analysis it is shown that for each procedure, there is an operating region in which the procedure is better than its competitor. Accordingly, a criterion is developed for deciding which of the two procedures is "better" to use in any situation. The varying situations involve relative sample size and variance ratio. Finally, the procedures are compared according to their asymptotic behavior. It is shown that for arbitrarily large sample sizes, the two procedures are equally good according to the criterion of expected confidence interval length.
1.2 Multivariate Behrens-Fisher Problems.

Bennett [8] extended the univariate procedure developed by Scheffé to the multivariate case. Since the multivariate analogue uses a vector version of the transformation, the term "Scheffé transformation" is used for the multivariate case also.

1.2.1 Intraclass Structure, General Mean.

The term "general Behrens-Fisher model" is defined below. In this chapter, only the restrictions to this model are discussed, and considerations of the general model are referred back to this subsection.

The statement "\( \Sigma \) has intraclass structure" will be used frequently in the sequel and it should be interpreted to mean that \( \Sigma \) can be represented in the form \( \Sigma = \sigma^2 [(1 - \rho)I_p + \rho e_p e_p'] \), for some \( \sigma \) and \( \rho \), and where \( I_p \) denotes the identity matrix of order \( p \), and \( e_p \) denotes a \( p \)-dimensional column vector of ones. The subscript "\( p \)" will be deleted whenever the dimension is clear from the context. A matrix \( \Sigma \) with intraclass structure will often be denoted by \( \Sigma(\rho, \sigma^2) \). Also, the notation \( \Sigma > 0 \) should be taken, throughout, to mean that \( \Sigma \) is positive definite (and \( \Sigma_1 > \Sigma_2 \) means that \( \Sigma_1 - \Sigma_2 > 0 \)).

For the "general model", take \( X \) and \( Y \) to be independent \( p \times m \) and \( p \times n \) random matrices \( (m \leq n) \) whose columns are independently and identically distributed as \( N(\mu_1, \Sigma_1) \), \( N(\mu_2, \Sigma_2) \), respectively.
For the problem of this subsection, assume the general model with intraclass structure, and no restrictions on the mean vectors. Define $\Sigma_j = \Sigma(\rho_j, \sigma_j)$. The problem is to test

$$H : (\mu_1 = \mu_2, \Sigma_j > 0, j = 1,2)$$

vs:

$$A : (\mu_1 \neq \mu_2, \Sigma_j > 0, j = 1,2).$$

The problem is simply treated by first diagonalizing the covariance matrices to achieve independence.

Let $\Gamma$ denote any $p \times p$ orthogonal matrix whose first row is $p^{-1/2} e_i$. Let $X^* = \Gamma X$, $Y^* = \Gamma Y$, $\theta_i = \Gamma \mu_i$, $i = 1,2$. The columns of $X^*$ are independently and identically distributed as $N(\theta_1, D_1)$ and the columns of $Y^*$ are independently and identically distributed as $N(\theta_2, D_2)$, where $D_i = \text{diag}(\alpha_i, \beta_i, \ldots, \beta_i)$, and for $i = 1,2$,

\begin{align*}
(1.2.1) & \quad \alpha_i = \sigma_i^2 [1 + \rho_i (p-1)], \\
(1.2.2) & \quad \beta_i = \sigma_i^2 [1 - \rho_i].
\end{align*}

Now the problem is to test $H : (\theta_1 = \theta_2, \alpha_i > 0, \beta_i > 0, i = 1,2)$ vs. $A : (\theta_1 \neq \theta_2, \alpha_i > 0, \beta_i > 0, i = 1,2)$.

Next apply a Scheffé transformation to reduce the two sample problem to a one sample problem. Let $Y^* = (Y_1, Y_2)$, where $Y_1$ is a $p \times m$ submatrix, and define $Z$ by
\begin{align}
Z &= X^* - \frac{1}{n} Y^* e e' \sqrt{\frac{m}{n}} Y_1 + \frac{1}{\sqrt{mn}} Y_1 e e'.
\end{align}

It is a straightforward calculation to show that as a consequence of this transformation the columns of $Z$ are independently and identically distributed as $N(\varphi, D)$, where $\varphi = \theta_1 \theta_2'$, $D = \text{diag}(\alpha, \beta, \ldots, \beta)$, and

\begin{align}
\alpha &= \alpha_1 + \frac{m}{n} \alpha_2, \\
\beta &= \beta_1 + \frac{m}{n} \beta_2.
\end{align}

The reformulated problem is to test $H : \{\varphi = 0, \alpha, \beta > 0\}$ vs. $A : \{\varphi \neq 0, \alpha, \beta > 0\}$. This problem can now be identified as a classical one in multivariate analysis. The likelihood ratio test procedure is developed in terms of statistics related to the sample mean, $w/\sqrt{m}$, and the $p \times p$ sample covariance matrix, $V/(m-1)$, $V = (v_{ij})$, defined by

\begin{align}
V &= Z(I - \frac{e e'}{m})Z', \\
w &= \frac{Ze m}{\sqrt{m}}.
\end{align}

We note that the vector $(v_{11}, \text{tr } V_{11}, w) \equiv (v, r, w)$ which is sufficient for the problem, has independent components. Using the distribution of $(v, r, w)$, given below, the likelihood ratio statistic $\lambda$ is given by
\[
\lambda = \left( \frac{v}{v + w_1^2} \right)^{\frac{m}{2}} \left( \frac{r}{r + \frac{p}{2} \sum_{j=2}^{p} w_j^2} \right)^{\frac{(p-1)m}{2}},
\]

where \( w' = (w_1, \ldots, w_p) \). It is now shown that under \( H \), \( \lambda^m \) is distributed as the product of independent Beta variates.

Let \( \mathcal{L}(X) \) denote the law of a random variable \( X \). Then,

\[
\mathcal{L}(v/\alpha) = \chi^2_{m-1},
\]

\[
\mathcal{L}(w_1^2) = \chi^2_{m\alpha^2},
\]

\[
\mathcal{L}(\frac{r}{\alpha}) = \chi^2_{(p-1)(m-1)},
\]

\[
\mathcal{L}\left( \frac{\sum_{j=2}^{p} w_j^2}{p} \right) = \chi^2_{p-1}\left( \frac{m}{\beta} \sum_j^{p} \phi_j^2 \right),
\]

where \( \chi^2(a,b) \) denotes a non-central chi-squared distribution having a degrees of freedom and non-centrality parameter \( b \). Under \( H \) the non-centrality parameters vanish and all distributions are central.

Since \( v \) and \( w_1^2 \) are independent, and since \( r \) and \( \frac{p}{2} \sum_{j=2}^{p} w_j^2 \) are independent,

\[
\mathcal{L}\left( \frac{v}{v + w_1^2} \right) = \beta\left[ \frac{m-1}{2}, \frac{1}{2} \right],
\]

\[8\]
(1.2.14) \[ z\left( \frac{r}{r + \frac{p}{2} \sum w_j^2} \right) = \beta \left( \frac{(p-1)(m-1)}{2}, \frac{p-1}{2} \right), \]

where \( \beta[a,b] \) denotes a Beta distribution with parameters \( (a,b) \).

Let \( U_{a,b} \) denote a random variable having distribution \( \beta[a,b] \). Then because \( [v(v+w_1^2)^{-1} \text{ and } [r(r+w_j^2)^{-1}] \) are independent, under \( H \),

(1.2.15) \[ z(\lambda_m^{\frac{1}{m}}) = z\left( \frac{U_{m-1, \frac{1}{2}}}{U_{(p-1)(m-1), \frac{p-1}{2}}} \right), \]

where the \( U \)'s are independent Beta variates. Next note the easily verifiable relation that

(1.2.16) \[ z\left( U_{a, b}^n \right) = z\left( \prod_{i=0}^{m-1} U_{a + \frac{i}{m}, b} \right), \quad m = 1, 2, \ldots, \]

where all variates are independent. Hence,

(1.2.17) \[ z\left( \lambda_m^{\frac{2}{m}} \right) = z\left( \frac{U_{m-1, \frac{1}{2}}}{U_{p-1, \frac{p-1}{2}}} \prod_{i=0}^{p-2} v + \frac{i}{p-1}, \frac{p-1}{2} \right), \]

where \( v = \frac{(m-1)(p-1)}{2} \), and the \( U \)'s are all independent.

Since there is no known closed form for this distribution for arbitrary \( p \), the distribution is approximated. We use the method of Box [10]. It is easy to show that for \( h = 0, 1, \ldots, \)
(1.2.18) \[ E(\lambda)^h = K \frac{\Gamma\left[\frac{m}{2}(1+h) - \frac{1}{2}\right] \Gamma\left[\frac{m}{2}(p-1)(1+h) - \frac{p-1}{2}\right]}{\Gamma\left[\frac{m}{2}(1+h)\right] \Gamma\left[\frac{m}{2}(p-1)(1+h)\right] \Gamma\left[\frac{m}{2}(p-1)(m-1)\right]} \]

where

(1.2.19) \[ K = \frac{\Gamma\left(\frac{m}{2}\right) \Gamma\left[\frac{m(p-1)}{2}\right]}{\Gamma\left(\frac{m-1}{2}\right) \Gamma\left[\frac{(p-1)(m-1)}{2}\right]} \]

Using the notation of Anderson [2], p. 203, make the following identifications.

Let \( x_1 = y_1 = \frac{m}{2} \), \( x_2 = y_2 = (p-1) \frac{m}{2} \), \( \xi_1 = -\frac{1}{2} \), \( \xi_2 = -\frac{p-1}{2} \), \( \eta_1 = \eta_2 = 0 \), \( a = b \) and \( a = 2 \). It is seen that all the conditions for application of Box's approximation apply. Under H, the approximation is

(1.2.20) \[ P(-2 \ln \lambda \leq t) = (1-\omega)P\left(\chi^2_p \leq t(1 - \frac{p+4}{2mp})\right) \]

\[ + \omega P\left(\chi^2_{p+4} \leq t(1 - \frac{p+4}{2mp})\right) + O(m^{-3}) \]

where

(1.2.21) \[ \omega = \frac{p(p^3 + 15p^2 - 24p + 16)}{12(p-1)(2mp - p - 4)^2} \]

The power function can be approximated for large \( m \) by considering the asymptotic distribution of \( \lambda \) under the alternative hypothesis, as developed by Wald [34]. Define the \( p \times p \) information
matrix $J$, with elements $J_{ij}$, by

$$ (1.2.22) \quad J_{ij} = -E \frac{\partial^2 \ln p^*}{\partial \varphi_i \partial \varphi_j}, $$

where $p^*$ is the joint density of $(w,v,r)$. Then for large $m$, under $A$, it is easy to find that because the $\varphi_i$ enter $p^*$ linearly, $J$ is diagonal, and approximately,

$$ (1.2.23) \quad z(-2\ln \lambda) \sim \chi^2_p(\varphi'J\varphi) = \chi^2_p(mp'D^{-1}p). $$

The results of this subsection are summarized below.

Let the model be the general model for the multivariate Behrens-Fisher problem with covariance matrices having intraclass structure. If a Scheffé procedure is applied, the likelihood ratio statistic, $\lambda$, has the approximate distribution given by (1.2.20) under $H$, and by (1.2.23), under $A$.

1.2.2 Intraclass Structure, Restricted Mean.

Assume the general Behrens-Fisher model with intraclass structure as defined in Section 1.2.1. In addition, let $\mu_1 = \mu_0^{e_1}p$, $\mu_2 = \mu_0^{e_2}p$; i.e., the means in each population are the same. The problem is to test $H : \{\mu_0^{e_1} = \mu_0^{e_2}, \Sigma(\rho_j, \sigma_j^2) > 0, j = 1,2\}$ vs. $A : \{\mu_0^{e_1} \neq \mu_0^{e_2}, \Sigma(\rho_j, \sigma_j^2) > 0, j = 1,2\}$.

A canonical form for this problem when the mean is general was developed in Section 1.2.1. It is to test $H : \{\varphi = 0; \alpha, \beta > 0\}$ vs. $A : \{\varphi \neq 0; \alpha, \beta > 0\}$, where now,
Clearly, \((v,w_1)\) defined by (1.2.7) and (1.2.9) is a minimal sufficient statistic for the problem. Moreover, since only \(\varphi_1\) is really involved and because of independence, the problem is equivalently univariate. Accordingly, define

\[
(1.2.25) \quad g = \frac{(m-1)w_1^2}{v}.
\]

Then, a UMP unbiased level \(\alpha\) test is to reject \(H\) for \(g > \text{constant}\).

From the relevant distributions above (1.2.11) and (1.2.12), it is clear that

\[
(1.2.26) \quad \chi(g) = F_{1,m-1} \left[ \frac{m}{\alpha} (\mu_{o1} - \mu_{o2})^2 \right],
\]

where \(F_{a,b}(w)\) denotes a non-central \(F\) distribution with degrees of freedom \((a,b)\), and non-centrality parameter \(\omega\).

1.2.3 Functionally Dependent Covariances, General Means.

Assume the general Behrens-Fisher model of Section 1.2.1 with the added condition that

\[
(1.2.27) \quad \Sigma_1 = f(\Sigma_2),
\]

where \(f(\cdot)\) is a \(p \times p\) positive definite matrix. The problem is to test \(H : [\mu_1 = \mu_2, \Sigma_i > 0, \ i = 1,2]\) vs. \(A : [\mu_1 \neq \mu_2, \Sigma_i > 0, \ i = 1,2]\).
This time, apply a Scheffé transformation immediately
(without a prior rotation). Let

\[(1.2.28) \quad Z = X - \frac{1}{n} Y e_n e_m' - \frac{1}{\sqrt{mn}} Y_1^* e_m e_m' + \frac{1}{\sqrt{mn}} Y_2^* e_m e_m',\]

where \(Y = (Y_1^*, Y_2^*)\), and \(Y_1^*\) is \(p \times m\), \(Y_2^*\) is \(p \times (n-m)\). The columns of \(Z\) are independently and identically distributed as \(N(\mu_1 - \mu_2, \Sigma)\), where

\[(1.2.29) \quad \Sigma = \Sigma_1 + \frac{m}{n} \Sigma_2 = f(\Sigma_2) + \frac{m}{n} \Sigma_2.\]

Thus, to test \(\mu_1 = \mu_2\), we are reduced to Hotelling's \(T^2\) test derived by Bennett [8] for the general case of unstructured and non-dependent covariance matrices. This problem has been discussed in order to underscore this insensitivity of the Scheffé procedure to functional dependence of the covariance matrices.

1.2.4 Functionally Dependent Covariances, Restricted Means.

Assume the general Behrens-Fisher model with functionally dependent covariance matrices, i.e., \(\Sigma_1 = f(\Sigma_2)\), where \(f(\cdot)\) defines a positive definite \(p \times p\) matrix. Imposing the further restrictions of equality of component means; i.e., \(\mu_1 = \mu_{01} e_p',\ \mu_2 = \mu_{02} e_p'\). The problem is to test \(H : \{\mu_{01} = \mu_{02}, \ \Sigma_1 > 0, \ i = 1,2\}\) vs. \(A : \{\mu_{01} \neq \mu_{02}, \ \Sigma_1 > 0, \ i = 1,2\}\). The transformations of Section 1.2.1 reduce the problem to its canonical form, in which the columns of the \(p \times m\) matrix \(Z\) are independently and identically distributed as \(N(\varphi, \psi)\), where \(\varphi' = (\mu_{01} - \mu_{02}) (1,0,\ldots,0)\), and \(\psi = f(\Sigma_2) + \frac{m}{n} \Sigma_2\).
The problem is to test \( H : (\varphi_1 = 0, \psi > 0) \) vs. \( A : (\varphi_1 \neq 0, \psi > 0) \).

Let \( V/(m-1), w/\sqrt{m} \) be the sample covariance matrix and mean, respectively, and partition them as

\[
(1.2.30) \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \quad w = \begin{pmatrix} \hat{\omega} \\ \hat{\omega} \end{pmatrix},
\]

where \( V_{11} \) is the (1,1) component of \( V, V_{22} \) is \((p-1) \times (p-1)\), and \( \hat{\omega} \) is the first component of \( w \). It is a straightforward computation using Section 1.2.1 and \( z(w) = \mathbb{N}(\sqrt{m} \varphi, \psi) \) to show that if \( \lambda \) denotes the likelihood ratio statistic for the canonical problem,

\[
(1.2.31) \quad Q = \frac{2}{\lambda^{m-1}} = \frac{1 + \hat{\omega} V^{-1}_{22} \hat{\omega}}{1 + w^t V^{-1} w}.
\]

The distribution of this statistic is given in various contexts by Cochran and Bliss [12], Giri [17], and Olkin and Shrikhande [21].

Under \( H, \)

\[
(1.2.32) \quad z(Q) = \beta\left[ \frac{m-p-1}{2}, \frac{1}{2} \right].
\]

Under \( A, Q \) has density [21] given by

\[
(1.2.33) \quad c \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A(i,j) \frac{\frac{m-p-2}{2} (1-Q)^j - \frac{1}{2}}{B\left(\frac{m-p}{2}, j + \frac{1}{2}\right)},
\]

where \( B(a,b) \) is the usual Beta function,
\[ c = \exp \left[ -\frac{m\Lambda^{-1}}{2} \left( \mu_{01} - \mu_{02} \right)^2 \right] \frac{B \left( \frac{p-1}{2}, \frac{m-p+1}{2} \right)}{B \left( \frac{p-1}{2}, \frac{m-p+1}{2} \right)} , \]

\[ A(i,j) = \frac{1}{i!j!} \left[ \frac{m\Lambda^{-1}}{2} \left( \mu_{01} - \mu_{02} \right)^2 \right]^{i+j} B \left( i + \frac{p-1}{2}, j + \frac{m-p+1}{2} \right) , \]

and \( \Lambda = \psi^{-1} \) is partitioned as in (1.2.30). The above results yield

**Theorem 1.2:** Assume the general Behrens-Fisher model defined in Section 1.2.1. Assume further that \( \Sigma_1 = f(\Sigma_2) \), \( \mu_1 = \mu_{01} e_p \), and \( \mu_2 = \mu_{02} e_p \). If the Scheffé transformation is applied to the problem, the likelihood ratio statistic is given by (1.2.31), with density given in (1.2.33).

1.3 Optimality of the Multivariate Scheffé Transformation Test Procedure.

In the univariate case, Scheffé showed that by suitably selecting the constants of the transformation, the associated confidence interval for the difference of the population means can be minimized. In generalizing the procedure to the multivariate case, Bennett [8] does not justify his choice of constants. In this section, it is shown that for the choice of constants made by Bennett, there is a multivariate optimality property analogous to the one in the univariate case.

Assume the general Behrens-Fisher model as defined in Section 1.2.1. Scheffé, for the case \( p = 1 \), considered a transformation on the data of the form (for \( m \leq n \))
(1.3.1) \[ Z = X - YC , \]

where \( C \) is an \( n \times m \) matrix of constants to be determined.

Demanding that the columns of \( Z \) be independent and identically distributed with mean \( \mu_1 - \mu_2 \) imposes the conditions that

(1.3.2) \[ C \mathbf{e}_m = \mathbf{e}_n , \]

(1.3.3) \[ CC' = \alpha I , \]

where \( \alpha \) is a constant to be determined. Scheffé showed that the smallest value \( \alpha \) can have subject to (1.3.2), (1.3.3) is \( m/n \). Moreover, when we generalize to the multivariate case, the minimum possible value of \( \alpha \) remains the same, and it is this value of \( \alpha \) which Bennett chose for the generalized transformation. It is now shown that by choosing this \( \alpha \), we obtain a confidence ellipsoid for \( \mu_1 - \mu_2 \) which has minimum expected volume.

Regardless of the value assigned to \( \alpha \), under the transformation (1.3.1) the problem is to test \( H : \{ \mu_1 = \mu_2, \Sigma > 0 \} \) vs. \( A : \{ \mu_1 \neq \mu_2, \Sigma > 0 \} \) and the columns of \( Z \) are distributed independently and identically as \( N(\mu_1 - \mu_2, \Sigma = \Sigma_1 + \alpha \Sigma_2) \). But Hotelling's \( T^2 \) test is UMP unbiased and invariant under the linear group. Hence, we know immediately that an associated confidence region for \( \mu_1 - \mu_2 \) is the ellipsoid with expected volume

(1.3.4) \[ \mathbf{E}(v) = \left( \left[ \frac{2\pi T^2}{\alpha} \right]^{\frac{p}{2}} \frac{\Gamma(m/2)}{\Gamma(m-2)} \right) \left| \Sigma \right|^{\frac{1}{2}} = K \left| \Sigma \right|^{\frac{1}{2}} , \]
where $T^2_\alpha$ is the $\alpha$ point of the Hotelling $T^2$ distribution with $p$ degrees of freedom.

\[
E(V) = K|\Sigma_1 + \alpha\Sigma_2|^{\frac{1}{2}} = K|\Sigma_2|^{\frac{1}{2}} |\Sigma_2^{-\frac{1}{2}} \Sigma_1 \Sigma_2^{-\frac{1}{2}} + \alpha I|^{\frac{1}{2}}.
\]

But since $E(V)$ remains invariant under transformations by determinants of orthogonal matrices, it is easily seen that $E(V)$ is a monotone increasing function of $\alpha$, and hence, is minimized if $\alpha = m/n$.

1.4 Behrens-Fisher Problems in Multivariate Regression.

Let $X$ and $Y$ be independent $p \times m$ and $p \times n$ sample matrices. That is, let the columns of $X$ be independently and identically distributed as $N(M\beta, \Sigma_1)$, and the columns of $Y$ be independently and identically distributed as $N(N\gamma, \Sigma_2)$, where $M,N$ are known $p \times k$ matrices of rank $k$, $k \leq p$, and $\beta,\gamma$ are $k$-variate parameter vectors associated with two separate linear regression models; $\Sigma_1, \Sigma_2$ are the error covariance matrices. In this section, the Scheffé transformation is used in the problem of testing $H : \{\beta = \gamma, \Sigma_i > 0, i = 1,2\}$ vs. $A : \{\beta \neq \gamma, \Sigma_i > 0, i = 1,2\}$, under various assumptions of covariance structure. The model described above will be referred to as the "general Behrens-Fisher regression model". For $p = 1$, the problem of testing equality of linear regressions was considered by Barankin [5], who also used the Scheffé approach.

**Theorem 1.4.1:** Let $X$ and $Y$ be sample matrices defined above. Assume $\Sigma_1, \Sigma_2$ each have the structural form

17
\[ \Sigma_i = \begin{pmatrix} \Sigma_{11}^{(i)} & 0 \\ 0 & \Sigma_{22}^{(i)} \end{pmatrix}, \quad i = 1, 2 \text{ with } \Sigma_{11}^{(i)} (k \times k) \text{ and } \Sigma_{22}^{(i)} (p-k \times p-k). \]

Then, there exists a test of \( \beta = \gamma \) vs. \( \beta \neq \gamma \) determined from a Scheffé transformation and based upon Hotelling's \( T^2 \) statistic, which has the optimality property associated with procedures obtained using Scheffé transformations.

**Proof:** Using the well known result (see e.g., Gantmacher [16]) that if \( A \) is any \( p \times k \) matrix, \( p \geq k \), there exists a \( k \times k \) lower triangular matrix \( T = (t_{ij}) \) with \( t_{ii} > 0 \) for all \( i \), and a \( k \times k \) orthogonal matrix \( \Gamma \) such that \( A = \Gamma \left( \begin{array}{c} T_1 \\ 0 \end{array} \right) \), we have that

\[
(1.4.1) \quad M = \Gamma_1 \left( \begin{array}{c} T_1 \\ 0 \end{array} \right), \quad N = \Gamma_2 \left( \begin{array}{c} T_2 \\ 0 \end{array} \right).
\]

Now the columns of \( X, Y \) are \( \mathbb{N}[(\Gamma_1^T \beta, 0)', \Sigma_1], \mathbb{N}[(\Gamma_2^T \gamma, 0)', \Sigma_2] \), respectively. Let \( M_1 = \Gamma_1 T_1, \ N_1 = \Gamma_2 T_2 \). Next partition \( X, Y, \Sigma_1, \Sigma_2 \) as

\[
(1.4.2) \quad X = \begin{pmatrix} \dot{X} \\ X \end{pmatrix}, \quad Y = \begin{pmatrix} \dot{Y} \\ Y \end{pmatrix}, \quad \Sigma_i = \begin{pmatrix} \Sigma_{11}^{(i)} & 0 \\ 0 & \Sigma_{22}^{(i)} \end{pmatrix},
\]

where \( \dot{X}, \dot{Y} \) each have \( k \) rows, and the dimensions of \( \Sigma_{11}^{(i)}, \Sigma_{22}^{(i)} \) are \( k \times k \) and \( (p-k) \times (p-k) \), respectively, \( i = 1, 2 \). Now it is clear that the columns of \( \dot{X} \) are independently and identically
distributed as \( N(M_1 \beta, \Sigma^{(1)}_{11}) \), those of \( \ddot{X} \) as \( N(0, \Sigma^{(1)}_{11}) \), those of \( \ddot{Y} \) as \( N(N_1 \gamma, \Sigma^{(2)}_{11}) \), and those of \( \ddot{Y} \) as \( N(0, \Sigma^{(2)}_{22}) \).

By independence, \( \ddot{X} \) and \( \ddot{Y} \) can only be used to estimate \( \Sigma^{(1)}_{22}, \Sigma^{(2)}_{22} \). Let \( X_1 \) and \( Y_1 \) be independent and defined as in

\[
(1.4.3) \quad X_1 = M^{-1}_1 \ddot{X}, \quad Y_1 = N^{-1}_1 \ddot{Y}.
\]

Then the columns of \( X_1 \) and \( Y_1 \) are independently and identically distributed as \( N(\beta, \psi_1) \), and \( N(\gamma, \psi_2) \), respectively, where

\[
(1.4.4) \quad \psi_1 = M^{-1}_1 \Sigma^{(1)}_{11} (M^{-1}_1)' \quad \text{and} \quad \psi_2 = N^{-1}_1 \Sigma^{(2)}_{11} (N^{-1}_1)'.
\]

The problem is to test \( H : \{ \beta = \gamma, \psi_1 > 0, \psi_2 > 0 \} \) vs.

\( A : \{ \beta \neq \gamma, \psi_1 > 0, \psi_2 > 0 \} \).

Apply a Scheffé transformation to \( X_1 \) and \( Y_1 \) as in (1.2.28).

This gives

\[
(1.4.5) \quad Z = X_1 - \frac{1}{n} Y_1 e_n e_n' - \sqrt{\frac{m}{n}} Y^*_1 + \frac{1}{\sqrt{mn}} Y^*_1 e_n e_n' ,
\]

where \( Y_1 = (Y^*_1 Y^*_2) \), \( Z \) and \( Y^*_1 \) are \((k \times m)\) matrices, \( Y^*_2 \) is a \(k \times (n-m)\) matrix, and we want to test \( H : \{ \beta-\gamma = 0, \psi = \psi_1 + \frac{m}{n} \psi_2 > 0 \} \)

vs. \( A : \{ \beta-\gamma \neq 0, \psi = \psi_1 + \frac{m}{n} \psi_2 > 0 \} \). This problem is treated in the usual way by means of the Hotelling \( T^2 \) test. Define the sufficient statistic \((V, \bar{Z})\), the covariance and mean, by
(1.4.6) \[ V = Z(I - \frac{e e'}{m})Z', \quad \bar{z} = \frac{Ze_m}{m}, \]

Then if \( T^2 = m(m-1)\bar{z}'V^{-1}\bar{z}, \) the test is to reject \( H \) if \( T^2 > C \)
where \( C \) is a constant determined from

(1.4.7) \[ x\left(\frac{T^2}{m-1}, \frac{m-k}{k}\right) = F_{k,m-k}(\delta), \]

where \( \delta = m(\beta - \gamma)'\psi^{-1}(\beta - \gamma). \) The proof that this test has the
optimality property asserted in the theorem was given in Section 1.3.

Theorem 1.4.2: Let \( X, Y \) be independent \( p \times m, p \times n \) matrices
with columns independently and identically distributed as
\( N(M_\beta, \sigma_1^2 I_k), N(M_\gamma, \sigma_2^2 I_k), \) respectively, \( m \leq n. \) Then, using a Scheffé
transformation, the likelihood ratio test of \( \beta = \gamma \) vs. \( \beta \neq \gamma \) is
based upon a variate having the F-distribution. The test has the
usual optimality properties of F-tests in the analysis of variance,
when the class of tests is restricted to the domain of those obtained
by Scheffé transformations.

Proof: From the canonical form developed in the previous proof we
immediately have that the columns of \( X, X \) are independently and
identically distributed as \( N(M_1 \beta, \sigma_1^2 I_k), N(0, \sigma_1^2 I_{p-k}), \)
respectively. Similarly, the columns of \( Y, Y \) are independently and
identically distributed as \( N(M_1 \gamma, \sigma_2^2 I_k), \) and \( N(0, \sigma_2^2 I_{p-k}), \)
respectively. Note that in contrast to the proof of the previous
theorem, in this case the second set of sample data also contributes

20
to the problem since they estimate the population variances. Applying
the Scheffé transformation to both sets of components gives

\[
(1.4.8) \quad \begin{pmatrix} \hat{Z} \\ \hat{Z} \end{pmatrix} = \begin{pmatrix} \hat{X} \\ \hat{X} \end{pmatrix} - \frac{1}{n} \begin{pmatrix} \hat{Y} \\ \hat{Y} \end{pmatrix} e_m e'_m - \sqrt{\frac{m}{n}} \begin{pmatrix} \hat{Y^*} \\ \hat{Y^*} \end{pmatrix} + \frac{1}{\sqrt{mn}} \begin{pmatrix} \hat{Y^*} \\ \hat{Y^*} \end{pmatrix} e_m e'_m ,
\]

where

\[
(1.4.9) \quad \hat{Y} = (\hat{Y^*}, \hat{Y^**}), \quad \hat{Y} = (\hat{Y^*}, \hat{Y^**}),
\]

and \( \hat{Z}, \hat{Y^*} \) are \( k \times m \) matrices, and \( \hat{Y^*} \) is a \( (p-k) \times m \) matrix.

Define

\[
(1.4.10) \quad \varphi = M_1 (\beta - \gamma), \quad \sigma^2 = \sigma_1^2 + \frac{m}{n} \sigma_2^2 .
\]

The reformulated problem is to test \( H : (\varphi = 0, \sigma^2 > 0) \), vs.
\( A : (\varphi \neq 0, \sigma^2 > 0) \) when the columns of \( \hat{Z} \), and those of \( \hat{Z} \) are each
distributed independently and identically as \( N(\varphi, \sigma^2 I_k) \), \( N(0, \sigma^2 I_{p-k}) \),
respectively. In this form it is clear that the problem is just a
special case of the univariate analysis of variance with homoscedasticity,
in which instead of testing for equality of row means, we want to test
whether all row means vanish, given an independent estimate of the
variance.

Define the sample row means and the sample variance by

\[
(1.4.11) \quad z = \frac{Z e_m}{\sqrt{m}},
\]
\[ w = \text{tr}\left[ \hat{Z} \left( I_m - \frac{m \cdot e'e}{m} \right) \right] + \hat{Z} \hat{Z}' , \]

and let

\[ W = \frac{mp-k}{k} \frac{z'z}{w} . \]

Since \((z,w)\) is sufficient, and \(z,w\) are independent with distributions given by

\[ \mathcal{N}(\varphi \sqrt{m}, \sigma^2 I_k) , \]

\[ \chi^2_{mp-k} , \]

it is clear that the appropriate F test (with the usual properties of ANOVA F-tests) is to reject for \( W > \text{constant} \), where

\[ \chi^2(W) = F_{k,mp-k} \left( \frac{mp'\varphi}{\sigma^2} \right) . \]

1.5 A Confidence Interval Comparison of Two Test Procedures

Proposed for the Univariate Behrens-Fisher Problem.

The univariate procedures of Scheffé [27] and Welch [35] are compared according to the expected lengths of the confidence intervals they yield. A criterion is developed for deciding which of the two methods is better in any given situation. An expression for
the expected length of the confidence interval for the Scheffé procedure
was developed by Scheffé. Because of the asymptotic series nature of
the Welch procedure, in this case, the analogous expression is more
involved. It is developed below. All variables are assumed to be
one dimensional.

1.5.1 Criterion for Relative Test Efficiency.

Let \( x_1, \ldots, x_m \) be independently and identically distributed
with \( x(x) = \mathcal{N}(\mu_1, \sigma_1^2) \). Let \( y_1, \ldots, y_n \) be independent of the \( x \)'s
and independently and identically distributed with \( x(y) = \mathcal{N}(\mu_2, \sigma_2^2) \).
Denote the variance ratio by \( \gamma^2 \). That is,

\[
(1.5.1) \quad \gamma = \frac{\sigma_2}{\sigma_1}. 
\]

It is easily seen from (1.2.28) that the confidence interval associated
with the general Scheffé procedure (univariate) is derivable from a
t-test and has length \( L_s \) given by

\[
(1.5.2) \quad L_s = 2C \epsilon \left[ \frac{1}{m(m-1)} \sum_1^m (z_i - \overline{z})^2 \right]^{1/2}, 
\]

and expected length given by

\[
(1.5.3) \quad E(L_s) = \frac{2\sqrt{2} \epsilon \sigma_1 \sqrt{1 + \frac{m}{n} \gamma^2} \Gamma(m/2)}{m \Gamma((m-1)/2)} \epsilon, 
\]

where \( \epsilon \) is the one sided upper 100 \( \epsilon \) percentage point of the
t-distribution with \( m-1 \) degrees of freedom, and the \( z_i \) are the
same as the columns of \( Z \), defined in Section 1.2.1, for \( p = 1 \).
The analogous expression for the Welch procedure is obtained as follows. Welch sought a statistic \( h(s_1^2, s_2^2, \epsilon) \) such that, independently of \( \gamma \),

\[
(1.5.4) \quad \epsilon = P\left( (\bar{x} - \bar{y}) - (\mu_1 - \mu_2) < h(s_1^2, s_2^2, \epsilon) \right),
\]

where

\[
(1.5.5) \quad \bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,
\]

\[
(1.5.6) \quad s_1^2 = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})^2, \quad s_2^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.
\]

His results yielded an asymptotic series solution for \( h \) given by

\[
(1.5.7) \quad \frac{h}{s_t^2} = 1 + \left( \frac{(m-1)s_1^2 + (n-1)s_2^2}{m-1} + \frac{(n-1)s_2^2}{n-1} \right) + \left( \frac{(m-1)s_1^2 + (n-1)s_2^2}{m-1} + \frac{(n-1)s_2^2}{n-1} \right)^2 + \ldots,
\]

where \( \xi \) is defined by \( \xi = \Phi^{-1}(\epsilon) \) for \( \Phi(\cdot) \) the c.d.f. of the standardized normal distribution, and \( s_t^2 \) is defined by

\[
(1.5.8) \quad s_t^2 = \frac{s_1^2}{m} + \frac{s_2^2}{n}.
\]

Using the Welch procedure, the length, \( L_W \), of the confidence
interval for \((\mu_1 - \mu_2)\) has expectation given by

\[
(1.5.9) \quad E(L_W) = 2E(h) = 2 \left\{ \xi E(s) + \xi^2 \left( \frac{s_1^4}{2m^3(m-1)} \right) + \xi \left( \frac{s_2^2}{n^3(n-1)} \right) \right\} + ... \right\}.
\]

The terms not shown explicitly are \(O(m^{-2}), O(n^{-2})\), as \(m, n \to \infty\) and will be omitted in later calculations. Since all functions involved are measurable and integrable, the series for \(h\) can be integrated termwise to yield (1.5.9). The ratio of the expressions in equations (1.5.3) and (1.5.9) provides a measure of relative test efficiency, namely

\[
(1.5.10) \quad R = \frac{E(L_W)}{E(L_S)}.
\]

Thus, small values of \(R\) favor the Welch procedure, and large values, the Scheffe procedure. To assign numerical values to \(R\), it is convenient to evaluate

\[
(1.5.11) \quad U_1 = \frac{1}{m^2(m-1)\sigma_1} E \left( \frac{s_1^4}{s^3} \right),
\]

\[
(1.5.12) \quad U_2 = \frac{1}{n^2(n-1)\sigma_1} E \left( \frac{s_2^4}{s^3} \right).
\]
(1.5.13) \[ U_j = \frac{1}{\sigma_1} E(s). \]

In terms of these quantities, \( R \) is given approximately (after neglecting the higher order terms) by

\[
(1.5.14) \quad R \approx \frac{m \Gamma \left( \frac{m-1}{2} \right)}{\sqrt{2} \Gamma \left( \frac{m}{2} \right) \sqrt{1 + \frac{m}{n} \gamma^2}} \left[ \frac{\xi \left( 1 + \xi^2 \right)}{\xi_4} (U_1 + U_2) + \xi U_3 \right].
\]

1.5.2 Evaluation of Expectations.

The expectations \( U_1, U_2, U_3 \) are all evaluated the same way. Let \( x \) and \( y \) be independent random variables with

\[
(1.5.15) \quad \mathbb{E}(x) = \frac{\chi^2}{\chi_{m-1}}, \quad \mathbb{E}(y) = \frac{\chi^2}{\chi_{n-1}},
\]

and \( h_1, h_2 \) be positive constants. Then, \( E(h_1 x + h_2 y)^{1/2} \) can be evaluated in terms of hypergeometric functions. From (1.5.15),

\[
(1.5.16) \quad E(h_1 x + h_2 y)^{1/2} = g \int_0^\infty \int_0^\infty (h_1 x + h_2 y)^{1/2} x^{m-1} y^{n-1} e^{-xy} \frac{dx}{y^2} \frac{dy}{y^2},
\]

where

\[
(1.5.17) \quad g = 2^{m+n-2} \frac{\Gamma \left( \frac{m-1}{2} \right) \Gamma \left( \frac{n-1}{2} \right)}{\Gamma \left( \frac{m+n-2}{2} \right)}.
\]

Transform \((x,y)\) to \((z,t)\) through \( z = x, t = x(x+y)^{-1}.\)
The Jacobian is \((-z/t^2\)). The integral in \(z\) is just a gamma function, and after some algebraic simplification, the result is

\[
E\left(\frac{h_1 x + h_2 y}{\sqrt{2}}\right)^{1/2} = \frac{\sqrt{2h_1} \Gamma\left(\frac{m+n-1}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right) \Gamma\left(\frac{n-1}{2}\right)} \int_0^1 t^\frac{m-3}{2} (1-t)^\frac{n-3}{2} \left(t + \frac{h_2}{h_1} - \frac{h_2}{h_1} t\right)^{1/2} dt.
\]

Evaluation of the remaining integral is carried out in two cases depending upon the size of \(h_2/h_1\). First recall that the hypergeometric function is defined by

\[
P(a,b;c;k) = 1 + \frac{ab}{c} \frac{k}{1} + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{k^2}{2} + \ldots,
\]

where \(|k| < 1\). Moreover, it is well known (see e.g., Rainville [24], p.47) that if \(\Re(c) > \Re(b) > 0\), and if \(|k| < 1\),

\[
P(a,b;c;k) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tk)^{-a} dt.
\]

Define

\[
\mathbf{E} = \frac{\sqrt{2} \Gamma\left(\frac{m+n-1}{2}\right)}{\Gamma\left(\frac{m+n-2}{2}\right)}
\]

Now by placing (1.5.18) in the form (1.5.20), it easily follows that

\[
E\left(\frac{h_1 x + h_2 y}{\sqrt{2}}\right)^{1/2} = \begin{cases} 
\mathbf{E} \frac{\sqrt{h_1}}{2} F\left(\begin{array}{c} \frac{1}{2}, \frac{n-1}{2}; \frac{m+n-2}{2} \end{array}; \frac{h_1-h_2}{h_1}\right), & h_2 < 2h_1 \\
\mathbf{E} \frac{\sqrt{h_2}}{2} F\left(\begin{array}{c} \frac{1}{2}, \frac{m-1}{2}; \frac{m+n-2}{2} \end{array}; \frac{h_2-h_1}{h_2}\right), & 2h_1 \leq h_2
\end{cases}
\]

27
Using the results of this canonical procedure, it is easy to evaluate $U_3$, as given in (1.5.13).

If $h_1, h_2$ are identified as in

\[(1.5.23) \quad h_1 = \frac{\sigma_1^2}{m(m-1)}, \quad h_2 = \frac{\sigma_2^2}{n(n-1)}, \]

then $\sigma_1 U_3$ is given by (1.5.22). $U_1$ and $U_2$ are evaluated analogously. Since the transformations required are the same as those above, the details are omitted. The final results are summarized below. Define the constants

\[(1.5.24) \quad \eta_1 = \frac{\sqrt{2} \Gamma \left(\frac{m+n-1}{2}\right) \Gamma \left(\frac{m+3}{2}\right)}{m(m-1)^2 \Gamma \left(\frac{m-1}{2}\right) \Gamma \left(\frac{m+n+2}{2}\right)}, \]

and

\[(1.5.25) \quad \eta_2 = \frac{\sqrt{2} \gamma \Gamma \left(\frac{m+n-1}{2}\right) \Gamma \left(\frac{n+3}{2}\right)}{n(n-1)^2 \Gamma \left(\frac{n-1}{2}\right) \Gamma \left(\frac{m+n+2}{2}\right)}. \]

Then $U_1$ and $U_2$ are easily shown to be given by

\[(1.5.26) \quad U_1 = \begin{cases} \eta_1 [m(m-1)]^{1/2} F \left(\frac{3}{2}, \frac{n-1}{2}; \frac{m+2}{2}; \frac{h_1-h_2}{h_1}\right), & h_2 < 2h_1 \\ \eta_1 \left[n(n-1)\right]^{3/2} m(m-1) \gamma^3 F \left(\frac{3}{2}, \frac{m+3}{2}; \frac{m+n+2}{2}; \frac{h_2-h_1}{h_2}\right), & 2h_1 \leq h_2, \end{cases} \]

and

28
\[
U_2 = \begin{cases} 
\eta_2 \left[ \frac{m(m-1)}{n(n-1)} \right]^{3/2} \gamma^3 F\left(\frac{3}{2}, \frac{n+3}{2}, \frac{m+n+2}{2}; \frac{h_1-h_2}{h_1}\right), & h_2 < 2h_1 \\
\eta_2 \left[ n(n-1) \right]^{1/2} F\left(\frac{3}{2}, \frac{m-1}{2}, \frac{m+n+2}{2}; \frac{h_2-h_1}{h_2}\right), & 2h_1 \leq h_2.
\end{cases}
\]

\( R \) is now computed by assigning numerical values to \( m, n, \gamma \), and the level of significance of the tests. This is discussed below.

1.5.3 Discussion of Numerical Results.

A numerical analysis of the asymptotic series solution obtained by Welch was carried out by Aspin [4]. It was shown that inclusion of terms of order \( m^{-2}, n^{-2} \) and higher is required only as \( \epsilon \) departs further from unity. Moreover, at \( \epsilon = .99 \), terms of order \( m^{-2}, n^{-2} \) contributed to the series only in the fourth decimal place. Accordingly, we take \( \epsilon = .995 \) (probability of .99 of being in a two sided symmetric interval), neglect terms of order \( m^{-2}, n^{-2} \), and consider the values so obtained accurate to only two decimal places, to be on the conservative side.

For \( \epsilon = .995 \), \( \Phi^{-1}(\epsilon) \) is 2.58 and \( \zeta_\epsilon \) is the .005 point of the one sided \( t \)-distribution with \( (m-1) \) degrees of freedom. These numbers are fixed throughout. \( m \) and \( n \) were chosen as 5, 10, 15, 20, 30, and \( \gamma \) was chosen as 0.1, 0.5, 0.9, 1.2, 2, 10. With these values, a Burrough's 220 Computer was used with a Stanford University Compiler to evaluate the test efficiency \( R \). The hypergeometric functions were evaluated by series until the series remainder was less than \( 4 \times 10^{-5} \). The results of the computer computations
are plotted in Figures 1 - 5, with data points connected by straight lines.

In Fig. 1, \( m = 5 \), and it is seen that the Welch procedure is best for comparable sample sizes and comparable variances.

In Figures 2, 3, 4 the same trends found in Fig. 1 are repeated and except for a scale factor, give essentially the same results. Note the trend toward use of the Scheffé procedure for extreme variance ratios.

In Fig. 5 the case of equal numbers of observations is examined separately. It is generally evident that for a combination of large number of observations and very large (or very small) variance ratio, Scheffé's result is preferable. In all other cases, Welch's procedure would be preferable.

Owing to the robustness of the \( t \) and \( F \) tests for the equality of means of normal distributions, small differences in population variances can usually be tolerated. Hence, the case which is usually the most difficult to handle is the case in which the Scheffé procedure shows great virtue over that of the Welch procedure.

The recommended test procedure is summarized in chart form below:
## Test Criterion

<table>
<thead>
<tr>
<th>&quot;Equivalent&quot; Variances $\left( \frac{\sigma_2}{\sigma_1} \right) \approx 1$</th>
<th>&quot;Extremely&quot; Large Variance Ratio</th>
<th>&quot;Extremely&quot; Small Variance Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Equivalent&quot; Sample</td>
<td>Welch Method</td>
<td>Scheffé Method</td>
</tr>
<tr>
<td>Sizes $\left( \frac{n}{m} \right) \approx 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Extreme&quot; Sample</td>
<td>Scheffé Method</td>
<td>Welch Method</td>
</tr>
<tr>
<td>Size Ratio $\left( \frac{n}{m} \right)$</td>
<td></td>
<td>Scheffé Method</td>
</tr>
</tbody>
</table>

In this chart, the words equivalent and extreme are to be interpreted liberally. In general, it is seen that the Scheffé procedure is best under a combination of extremes, whereas Welch's procedure is best when the sample numbers and population variances are roughly equivalent.
Figure 23

$E(L_Q) \over E(L_R)$

$m = 15$

$n = 30$

$n = 20$

$n = 15$

Welch Optimality Region

Scheffe Optimality Region
Figure #5
1.5.4 Large Sample Results.

In this section we examine the behavior of the two expected confidence interval lengths for large sample sizes. For simplicity, assume throughout that \( m = n \).

First examine \( E(L_s) \). Apply Stirling's approximation to the Gamma functions in (1.5.3). Direct computation yields

\[
E(L_s) = \frac{[2C \sqrt{1+\gamma^2}]}{\sqrt{n}} \left\{ 1 - \frac{103}{6n} + \ldots \right\}.
\]

Next consider \( E(L_w) \). For large \( n \), it is approximately true that

\[
(1.5.29) \quad \frac{s_1^2}{s_2^2} \approx \frac{\sigma_1^2}{\sigma_2^2}, \quad \frac{s_2^2}{s_2^2} \approx \frac{\sigma_2^2}{\sigma_2^2}, \quad s_2^2 = \frac{s_1^2 + s_2^2}{n},
\]

and

\[
(1.5.30) \quad \frac{s_1^2}{ns_2^2} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \quad \frac{s_2^2}{ns_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.
\]

Since the terms of the series in (1.5.10) decrease in magnitude as \( n \) increases, for asymptotic purposes it suffices to confine attention to the first few terms. Algebraic manipulation of (1.5.9) for large \( n \) gives

\[
(1.5.31) \quad E(L_w) = \frac{2\sqrt{\gamma^2 + 1}}{\sqrt{n}} \left\{ 1 + \frac{(1+\gamma^2)(1+\gamma^4)}{4(1+\gamma^2)^2} \frac{1}{n-1} + \ldots \right\}.
\]
Hence, it is clear that for large $n = m$, both $E(I_s)$ and $E(I_w)$ are $O(n^{-1/2})$. Moreover, asymptotically the relative test efficiency approaches a constant; i.e.,

$$R \to \frac{E}{C} \epsilon.$$  \hfill (1.5.32)

But since the Student t-distribution approaches the normal distribution for a large number of degrees of freedom,

$$R \to 1.$$  \hfill (1.5.33)

Hence, the two procedures are asymptotically equivalent.
CHAPTER 2

COMPARATIVE MULTIVARIATE REGRESSION WITH EQUAL COVARIANCE MATRICES

2.1 Introduction.

In this chapter two separate regressions on the same physical phenomenon are compared. It is assumed throughout that the error covariance matrices are identical. Likelihood ratio procedures are developed both for the case of equal and unequal design matrices. It is shown that regardless of whether the design matrices are equal or unequal, there is a likelihood ratio procedure which is based upon a test statistic having a noncentral $F$-distribution.

2.2 Testing With Equal Design Matrices.

We use the general Behrens-Fisher regression model introduced in Section 1.4. For this problem, we assume in addition that

\[ \Sigma_1 = \Sigma_2 = \sigma^2 I, \]  
and that \( M = N \). The problem of this section then is to test \( H: (\beta = \gamma, \sigma^2 > 0) \), vs. \( A: (\beta \neq \gamma, \sigma^2 > 0) \), when the columns of \( X, Y \) are independently and identically distributed as \( N(M\beta, \sigma^2 I) \), \( N(My, \sigma^2 I) \), respectively, \( M \) has dimension \( p \times k \), \( k \leq p \), and is known, and \( \beta, \gamma \) are \( k \) variate.

Following the transformations of equations (1.4.1) and (1.4.2) we may immediately advance to a canonical form of the model in which the columns of \( \tilde{X}, \tilde{X}, \tilde{Y}, \tilde{Y} \) are independently and identically distributed as \( N(M_{1} \beta, \sigma^2 I_{k}), N(0, \sigma^2 I_{p-k}), N(M_{1} \gamma, \sigma^2 I_{k}), N(0, \sigma^2 I_{p-k}) \), respectively (and mutually independent), where

\[
(2.2.1) \quad M_{1} = \Gamma T, \quad \Gamma T' = I, \quad M = \begin{pmatrix} M_{1} \\ 0 \end{pmatrix},
\]
and $T$ is lower triangular with dimension $k \times k$.

Now formulate the problem in terms of sufficient statistics.

Define

$$\text{(2.2.2)} \quad x = \text{tr} \left[ \frac{\hat{\Psi}^t}{X} + \frac{\hat{\Psi}^t}{Y} + \hat{\Psi} \left( I_k - \frac{e_m e_m^t}{m} \right) X' + \hat{\Psi} \left( I_k - \frac{e_n e_n^t}{n} \right) Y' \right],$$

and

$$\text{(2.2.3)} \quad w = \frac{n \hat{X} e_m - m \hat{Y} e_n}{\sqrt{mn(m+n)}}.$$

It is easily verified that $x$ and $w$ are independently distributed, $(x, w)$ is sufficient for $(\beta, \gamma, \sigma^2)$, and that

$$\text{(2.2.4)} \quad \mathcal{L}(x) = \sigma^2 \chi_{p(m+n)-2k}^2,$$

$$\text{(2.2.5)} \quad \mathcal{L}(w) = N(\varphi, \sigma^2 I_k),$$

where

$$\text{(2.2.6)} \quad \varphi = M_1 \sqrt{\frac{mn}{m+n}} (\beta, \gamma).$$

The reformulated problem is to test $\varphi = 0$ against $\varphi \neq 0$. Let

$$\text{(2.2.7)} \quad F = \frac{\frac{p(m+n)-2k}{k}}{\frac{w'w}{x}}.$$

The problem is easily recognized as a special case of the general linear hypothesis of ANOVA. The result is summarized in

Lemma 2.1.1:

Let $x, w$ be distributed independently as in (2.2.4), (2.2.5).

The "F-test" of $H: \{\varphi = 0, \sigma^2 > 0\}$ vs. $A: \{\varphi \neq 0, \sigma^2 > 0\}$ is to reject $H$ if $\frac{w'w}{x} > \text{constant}$, where
\[ (2.2.8) \quad I \left( \frac{p(m+n)-2k}{k} \right) \cdot \frac{w'w}{x} = F_{k, p(m+n)-2k} \left( \frac{\varphi' \varphi}{\sigma^2} \right). \]

The test is a specialized ANOVA procedure, and is therefore, UMP invariant under the linear transformation group.

2.3 Testing With Unequal Design Matrices.

The problem examined now is the same as that of the previous section except that here, \( M \neq N \). The model is placed into a canonical form and a likelihood ratio test is developed.

2.3.1 Development Of The Likelihood Ratio Statistic

As in the previous problem we can immediately advance to a canonical form of the model in which the columns of \( \hat{X}, \hat{X}, \hat{Y}, \hat{Y} \) are independently and identically distributed as \( N(M_1 \beta, \sigma^2 I_k), \ N(0, \sigma^2 I_{p-k}), \ N(N_1 \gamma, \sigma^2 I_k), \ N(0, \sigma^2 I_{p-k}) \), respectively (and mutually independent), where

\[ (2.3.1) \quad M_1 = \Gamma_1 T_1, \quad N_1 = \Gamma_2 T_2, \quad \Gamma_1 \Gamma_1' = \Gamma_2 \Gamma_2' = I, \]

\[ (2.3.2) \quad M = \begin{pmatrix} M_1 \\ 0 \end{pmatrix}, \quad N = \begin{pmatrix} N_1 \\ 0 \end{pmatrix}, \]

and \( T_1, T_2 \) are lower triangular. The problem is to test \( H: (\beta = \gamma, \sigma^2 > 0) \), vs. \( A: (\beta \neq \gamma, \sigma^2 > 0) \). To reduce to sufficient statistics let \( x \) be defined as in (2.2.2), and define the sample means by

\[ (2.3.3) \quad y = \frac{\hat{X}_m}{\sqrt{m}}, \quad z = \frac{\hat{X}_n}{\sqrt{n}}. \]

It is easy to see that
\begin{equation}
\mathcal{L}(y) = N(\sqrt{m} M_0 \beta, \sigma^2 I_k), \quad \mathcal{L}(z) = N(\sqrt{n} N_1 \gamma, \sigma^2 I_k).
\end{equation}

To simplify the canonical form let
\begin{equation}
\theta = \sqrt{m} M_0 \beta, \quad \varphi = \sqrt{m} M_0 \gamma,
\end{equation}
and
\begin{equation}
L = \sqrt{\frac{n}{m}} N_1 M_0^{-1}.
\end{equation}

Now (2.3.4) may be rewritten as
\begin{equation}
\mathcal{L}(y) = N(\theta, \sigma^2 I_k), \quad \mathcal{L}(z) = N(L\varphi, \sigma^2 I_k).
\end{equation}

Since \( \beta = \gamma \) if and only if \( \theta = \varphi \), in the canonical model we may test \( \theta = \varphi \) vs. \( \theta \neq \varphi \). Note that \( L \) is a known matrix, and that \( (x, y, z) \) is a minimal sufficient statistic with independently distributed components.

Next factor \( L \), since it is convenient in the remainder to have a symmetric matrix in place of the general matrix \( L \). It is always possible to represent any matrix \( L \) (see e.g., Gantmacher [16]) as
\begin{equation}
L = S \Delta
\end{equation}
where \( S \) is symmetric, and \( \Delta \) is orthogonal. Define
\begin{equation}
v = \Delta y, \quad \theta_0 = \Delta \theta, \quad \varphi_0 = \Delta \varphi.
\end{equation}

Then from (2.3.7)-(2.3.9),
\begin{equation}
\mathcal{L}(v) = N(\theta_0, \sigma^2 I), \quad \mathcal{L}(z) = N(S\varphi_0, \sigma^2 I).
\end{equation}

Moreover, the reformulated problem is to test \( \theta_0 = \varphi_0 \). By direct computation, it is found that if
(2.3.11) \[ Q = (Sv - z)'(I + S^2)^{-1}(Sv - z), \]

(2.3.12) \[ \mathcal{L}(\frac{x}{\sigma^2}) = \frac{\chi^2}{p(m+n)-2k}, \]

the likelihood ratio statistic \( \lambda \) is given by

(2.3.13) \[ \frac{2}{\lambda(m+n)p} = \frac{x}{x+Q}. \]

Equivalently, if

(2.3.14) \[ F = \frac{p(m+n)-2k}{k} \frac{Q}{x} \]

the test is to reject for \( F > \text{constant}. \)

2.3.2 Distribution Theory.

From (2.3.10) it is seen that if

(2.3.15) \[ s = Sv - z, \]

(2.3.16) \[ \mathcal{L}(s) = N[S(\theta_0 - \phi_0), \sigma^2(I + S^2)]. \]

It is easily seen that

(2.3.17) \[ \mathcal{L}(\frac{Q}{\sigma^2}) = \frac{\chi^2}{k} \frac{1}{\sigma^2}, \]

where

(2.3.18) \[ \eta \equiv (I + S^2)^{-1/2} S(\theta_0 - \phi_0), \]

and therefore, by independence of \( Q, x, \)

(2.3.19) \[ \mathcal{L}(F) = F_{k, p(m+n)-2k} \left( \frac{\eta'\eta}{\sigma^2} \right). \]
Denote the noncentrality parameter by $\delta$. Then

$$\delta = \frac{\mathbf{n}'\mathbf{n}}{\sigma^2},$$

and after some algebra, we find

$$\delta = \frac{(\mathbf{e}-\mathbf{v})'[\mathbf{I}+(\mathbf{L}'\mathbf{L})^{-1}]^{-1}(\mathbf{e}-\mathbf{v})}{\sigma^2}.$$  

The procedure suggested for this problem may be summarized as follows:

Assume the multivariate regression model in which the parameter vectors of two regressions on the same phenomenon are to be tested for equality. Assume further that the design matrices are unequal, and that the error vectors of both regressions are mutually independent and distributed with law $N(0, \sigma^2\mathbf{I}_p)$. The likelihood ratio procedure is to reject for $F > \text{constant}$, where $F$ is defined in (2.3.14). The distribution of $F$ is that of a noncentral $F$ variate with noncentrality parameter $\delta$ defined in (2.3.21), and $[k, p(m+n)-2k]$ degrees of freedom.
CHAPTER 3

DICHOTOMOUS CLASSIFICATION INTO NORMAL POPULATIONS

3.1 Introduction.

Problems concerned with dichotomous classification arise often in most analytical disciplines. A common situation involves two populations which are assumed to be normal with unknown population parameters. A sample is taken from each of the populations to estimate the parameters. The problem then is to determine to which of the two populations a newly obtained observation belongs. The literature is rich with considerations of this problem in both the univariate and multivariate cases, although usually only for identical covariance matrices.

In this chapter we develop decision procedures for p-variate normal populations with particular emphasis upon the cases of unequal covariance matrices and intraclass structured covariance matrices.

Smith [31] has considered the case of unequal covariance matrices in bivariate normal distributions. However, he did not determine the classification regions. Anderson and Bahadur [3] treat this problem from the aspect of determining the admissible linear procedures, and their approach is applied in this chapter to matrices with intraclass structure. The multivariate problem for equal but general covariance matrices is well summarized by Anderson [2], and the distribution theory for the case of unknown parameters has been summarized by Sitgreaves [29]. Solomon [32] has presented an interesting historical account of the overall subject.
In Section 3.2, likelihood ratio procedures are developed, assuming unequal covariance matrices whose difference is positive definite, both for known and unknown population parameters. The distribution of the test statistic is derived and several criteria are applied to the operating characteristics of the problem (the classification errors). An explicit form for the distribution is given in the case of known population parameters. However, since the form is complicated, approximations are also provided. For the case of unknown population parameters, several asymptotic procedures of varying accuracy are developed. The procedures prescribed should be tractable on a computer.

In Section 3.3 the Classification Problem is considered for the restriction in which the covariance matrices have intraclass structure (see Section 1.2.1). Likelihood ratio procedures are developed both for equal and unequal covariance matrices, and for known and unknown parameters, and it is shown that the associated distribution theory corresponds to that found in the general case, except that now we are restricted to a subspace of dimension two.

In Section 3.4 the method of classification developed by Anderson and Bahadur is applied to covariance matrices with intraclass structure. It is shown that although in the general case the admissible procedures can usually be found only implicitly and evaluated only through trial and error procedures, in the case of intraclass structure, explicit solutions can be obtained.

In Section 3.5 the restriction that the difference of the covariance matrices be positive definite is removed and classification procedures are developed in terms of the resultant indefinite forms.
3.2 Classification - General Covariance Matrices, $\Sigma_1 > \Sigma_2$.

Since the classification problem for general but equal covariance matrices is treated elsewhere (see above), in this section we assume unequal covariance matrices throughout. However, it is assumed, until Section 3.5 that $\Sigma_1 > \Sigma_2$.

3.2.1 Likelihood Ratio Discriminators

Let $z$ be a $p$ variate observation vector coming either from population $\pi_1: N(\mu_1, \Sigma_1)$ or from $\pi_2: N(\mu_2, \Sigma_2)$. The UMP test of $H: z \in \pi_1$ vs. $A: z \in \pi_2$, for the case of all four parameters known, is readily provided by the Neyman-Pearson Lemma. It is to reject $H$ when

$$(3.2.1) \quad T = (z-\mu_2)'\Sigma_2^{-1}(z-\mu_2) - (z-\mu_1)'\Sigma_1^{-1}(z-\mu_1) < c,$$

where $c$ is a constant which will be determined later. Note that $T$ is quadratic in $z$, whereas if $\Sigma_1 = \Sigma_2$, $T$ becomes the well known linear function of $z$.

Now assume that all parameters are unknown and that there is available a $p \times m$ matrix $X = (x_1, \ldots, x_m)$ whose columns are independent vector members of $\pi_1$; also assume there is available a $p \times n$ matrix $Y = (y_1, \ldots, y_n)$ whose columns are independent vector members of $\pi_2$. The hypothesis testing classification problem is now to test

$H: \{X \in \pi_1, z \in \pi_1, Y \in \pi_2\}$ vs. $A: \{X \in \pi_1, Y \in \pi_2, z \in \pi_2\}$. The sample mean and covariance are sufficient and defined as

$$(3.2.2) \quad x = \bar{x} = \frac{X}{m}, \quad y = \bar{y} = \frac{Y}{n},$$
\[ \begin{align*}
V &= X(I - \frac{e_m e_m'}{m})X', \quad W = Y(I - \frac{e_n e_n'}{n})Y'.
\end{align*} \]

The likelihood ratio statistic \( \lambda \) may be found to be \( k_1 \exp(U_{m,n}/2) \), where

\[ U_{m,n} = \ln \left\{ \frac{|W|}{|V|} \cdot \frac{[1 + \frac{n}{m+1}(y-z)'W^{-1}(y-z)]^{n+1}}{[1 + \frac{m}{m+1}(x-z)'V^{-1}(x-z)]^{m+1}} \right\}, \]

and \( k_1 \) = constant. Hence, the procedure is to classify into \( N(\mu_2, \Sigma_2) \) if \( U_{m,n} < k \).

3.2.2 Large Sample Procedures - Unknown Parameters.

In this section an approximate likelihood ratio test is developed for the case of unknown parameters when the sample sizes are large. Also, it is shown that asymptotically, the case of unknown parameters reduces to the case of known parameters.

Assume that \( m,n \) are sufficiently large so that large sample results will be applicable (in the sense that the type I, II errors are negligibly changed by use of asymptotic results in place of actual results). Since \( x, y, V/m, W/n \) are just sample moments, by a multivariate version of Cramér's Theorem ([13], p. 346) on the convergence in probability of sample moments to population moments,

\[ \begin{align*}
p \lim x &= \mu_1, \quad p \lim y = \mu_2, \\
p \lim \frac{V}{m} &= \Sigma_1, \quad p \lim \frac{W}{n} = \Sigma_2,
\end{align*} \]

where "p lim" denotes convergence in probability. Moreover, by the same theorem quoted, the convergence holds for rational functions of
the sample moments. Hence, from (3.2.4),

\[(3.2.7) \quad p\lim U_{m,n} = \ln \left\{ \frac{n^p}{m} \frac{\left| \Sigma_2 \right|}{\left| \Sigma_1 \right|} \left[ 1 + \frac{(z-\mu_2)'\Sigma_2^{-1}(z-\mu_2)}{n+1} \right] \left[ 1 + \frac{(z-\mu_1)'\Sigma_1^{-1}(z-\mu_1)}{m+1} \right] \right\} \]

If we use the Wald result that the maximum likelihood estimates converge almost surely to their expectations, a stronger result is obtained.

Since each of the bracketed expressions in (3.2.7) behaves like an exponential for large \( m,n \), if the approach to infinity takes place in a well behaved way \( (n/m = \delta^p = \text{constant}) \), we must find

\[(3.2.8) \quad p\lim U_{m,n} = U, \]

where

\[(3.2.9) \quad U = \ln \left\{ \frac{\left| \Sigma_2 \right|}{\left| \Sigma_1 \right|} e^{(z-\mu_2)'\Sigma_2^{-1}(z-\mu_2) - (z-\mu_1)'\Sigma_1^{-1}(z-\mu_1)} \right\} \]

Moreover, since convergence in probability implies convergence in law,

\[(3.2.10) \quad \mathcal{L}(U_{m,n}) \to \mathcal{L}(U) . \]

Define

\[(3.2.11) \quad v = \ln \left\{ \delta \frac{\left| \Sigma_2 \right|}{\left| \Sigma_1 \right|} \right\} . \]

Since for unknown parameters the exact test procedure is to reject \( H \) for \( U_{m,n} < k \), for large \( m,n \) by (3.2.10) the test procedure is to reject \( H \) for \( U < \text{constant} \). But this is equivalent to \( T < \text{constant} = c \), the test defined in (3.2.1) for known parameters. Hence, the asymptotic
test for the case of unknown parameters coincides with the test for the case of known parameters. Thus, by determining the constant \( c \) for the known parameters case we simultaneously have an asymptotic test for the unknown parameters case. In the latter case, the parameter values are replaced by their maximum likelihood estimates (the sample moments) to carry out the test.

An approximate non-asymptotic test may also be obtained. Since the exact non-asymptotic test is to reject for \( U_{m,n} < k \), and it is clear that \( k \geq c + \nu \), where \( \exp(\nu) = \frac{|\Sigma_2|}{|\Sigma_1|} \), an approximate test is to reject for

\[
(3.2.12) \quad U_{m,n} < c + \nu .
\]

The method for obtaining \( c \) is discussed in the next section. It involves using the estimates for the parameters. Let

\[
(3.2.13) \quad \nu^* = n \left( \frac{|\nu|}{|\nu|} \right).
\]

Then by the argument above, it follows immediately that

\[
(3.2.14) \quad \text{p lim} \ \nu^* = \nu .
\]

Hence, if \( \nu \) is approximated by \( \nu^* \), and the samples are used to yield some \( c^* \) which estimates \( c \) in the manner to be described, an approximate non-asymptotic test is to reject if

\[
(3.2.15) \quad U_{m,n} < c + \nu^* .
\]

The results of this section are summarized in the following paragraphs:
For the classification problem with unequal covariance matrices and unknown parameters, if \( m, n \to \infty \) so that \( n/m = \text{constant} \), the asymptotic test obtained coincides with that obtained for the known parameter case, eqn. (3.2.1); i.e., for large \((m,n)\) if the population parameters are replaced by their estimated values, eqn. (3.2.1) will be applicable.

For classification with unknown parameters, an approximate test is to classify \( z \) into \( N(\mu_1, \Sigma_1) \) if \( U_{m,n} \geq c^* + v^* \) where \( U_{m,n}, v^* \) are defined in (3.2.7) and (3.2.13), and \( c^* \) is the estimated value of \( c \) obtained by replacing all population parameters by their estimates.

3.2.3 Distribution Theory - Known Parameters.

In this section we find the distribution of \( T \) defined in (3.2.1) for the case when all parameters \((\mu_1, \mu_2, \Sigma_1, \Sigma_2)\) are known. Define

\[
A = \Sigma_2^{-1} - \Sigma_1^{-1}, \quad b' = \mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1}, \quad C = \mu_2' \Sigma_2^{-1} \mu_2 - \mu_1' \Sigma_1^{-1} \mu_1.
\]

Then, from (3.2.1) it follows that

\[
T = z'Az + 2b'z + C,
\]

where \( A, b, C \) are known, \( A \) is a \( p \times p \) matrix, \( b \) is a column vector, and \( C \) is a scalar. Now complete the square in (3.2.17) with the linear transformation

\[
v = z + A^{-1}b,
\]

where it is assumed that \(|A| \neq 0\). If we define
\begin{align}
(3.2.19) \quad & D = C - b^t A^{-1} b, \quad T_0 = T - D, \\
(3.2.20) \quad & T_0 = v' A v.
\end{align}

Since \( T_0 \) is just a translation of \( T \), the latter's distribution will suffice, and the test can also be based upon \( T_0 \). Assume that \( A \) is positive definite. This will be true, of course, if and only if \( \Sigma_1 - \Sigma_2 \) is positive definite. If we let

\begin{align}
(3.2.21) \quad & w = A^{1/2} v, \\
(3.2.22) \quad & T_0 = w' w.
\end{align}

Define

\begin{align}
(3.2.23) \quad & \theta_j = A^{1/2} (\mu_j + A^{-1} b), \quad \psi_j = (A^{1/2}) E_j (A^{1/2})', \quad j = 1, 2.
\end{align}

It is easy to see that since under \( H \), \( \mathcal{L}(z) = N(\mu_1, \Sigma_1) \), and under \( A \), \( \mathcal{L}(z) = N(\mu_2, \Sigma_2) \),

\begin{align}
(3.2.24) \quad & \mathcal{L}(w) = N(\theta_j, \psi_j),
\end{align}

where \( j = 1 \) under \( H \), and \( j = 2 \) under \( A \).

Note from (3.2.22) that \( T_0 \) is invariant under orthogonal transformations of \( w \) so that we may take the \( \psi_j \) to be diagonal. Define the characteristic values and vectors of \( \psi_1, \psi_2 \) by

\begin{align}
(3.2.25) \quad & \psi_j = \Gamma_j^t D_j \Gamma_j, \quad \Gamma_j^t \Gamma_j = I, \quad D_j = \text{diag}(\lambda_1^{(j)}, \ldots, \lambda_p^{(j)}),
\end{align}

for \( j = 1, 2 \). Without loss of generality, take \( \lambda_1^{(j)} > \lambda_2^{(j)} > \ldots > \lambda_p^{(j)} \).

Let

\begin{align}
(3.2.26) \quad & s^{(j)} = \Gamma_j w, \quad \varphi^{(j)} = \Gamma_j \theta_j,
\end{align}

\[ 52 \]
\[(3.2.27) \quad T_0^{(j)} = (s^{(j)})'(s^{(j)}) , \quad j = 1,2 , \]

where \( T_0^{(1)} = T_0 \) under H, and \( T_0^{(2)} = T_0 \) under A, and

\[(3.2.28) \quad \mathcal{L}(s^{(j)}) = N(\phi^{(j)}, D_j) . \]

Denoting the components of \( s^{(j)} \) by \( s_1^j, s_2^j, \ldots, s_p^j \), and those of \( \phi^{(j)} \) by \( \phi_1^j, \phi_2^j, \ldots, \phi_p^j \), by independence we have that

\[(3.2.29) \quad \mathcal{L}(s_k^{(j)}) = N(\phi_k^j, \lambda_k^{(j)}) , \quad k = 1,\ldots,p ; \quad j = 1,2 . \]

Define

\[(3.2.30) \quad \gamma_k^j = \frac{\phi_k^j}{\sqrt{\lambda_k^{(j)}}} , \quad k = 1,\ldots,p ; \quad j = 1,2 . \]

From (3.2.27), (3.2.29), and (3.2.30) it is clear that under H,

\[(3.2.31) \quad \mathcal{L}(T_0) = \mathcal{L} \left\{ \left( \sum_{k=1}^{p} \lambda_k^{(1)} \chi_1^2(\gamma_k^1)^2 \right) \right\} , \]

and under A,

\[(3.2.32) \quad \mathcal{L}(T_0) = \mathcal{L} \left\{ \left( \sum_{k=1}^{p} \lambda_k^{(2)} \chi_1^2(\gamma_k^2)^2 \right) \right\} . \]

Next note from (3.2.23) that since \( \lambda_k^{(j)} \) are the latent roots of \( \psi_j, (1+\lambda_k^{(1)}) \) are the latent roots of \( \Sigma_2^{-1}\Sigma_1 \), and \( (1-\lambda_k^{(2)}) \) are the latent roots of \( \Sigma_1^{-1}\Sigma_2 \). The results of this section are summarized in

**Theorem 3.2.1:** Let \( T = (z-\mu_2)'\Sigma_2^{-1}(z-\mu_2) - (z-\mu_1)'\Sigma_1^{-1}(z-\mu_1) \) where \( (\mu_1, \mu_2, \Sigma_1, \Sigma_2) \) are all known. Assume that under H, \( \mathcal{L}(z) = N(\mu_1, \Sigma_1) \) and under A, \( \mathcal{L}(z) = N(\mu_2, \Sigma_2) \), where \( \Sigma_1^{-1}\Sigma_2 \) is positive definite.
Then under $H \mathcal{L}(T) = \mathcal{L}(D + \sum_{k=1}^{p} \lambda_k^{(1)} \chi^2_1(\gamma_k^{(1)}))$, and under $A$

$\mathcal{L}(T) = \mathcal{L}(D + \sum_{k=1}^{p} \lambda_k^{(2)} \chi^2_1(\gamma_k^{(2)}))$, where $D$ is a known constant defined by (3.2.19), $(1 + \lambda_k^{(1)})$ are the latent roots of $\Sigma_1^{-1} \Sigma_1$, $(1 - \lambda_k^{(2)})$ are the latent roots of $\Sigma_1^{-1} \Sigma_2$, and $\lambda_k^{(1)} = \lambda_k^{(2)}/(1 - \lambda_k^{(2)})$ for $k = 1, \ldots, p$.

Now invoke a result demonstrated in Chapter 4. Let

$X = a[\chi^2_{m_0, d_0} + a_1 \chi^2_{m_1, d_1} + \cdots + a_r \chi^2_{m_r, d_r}]$ where $\chi^2_{m, d}$ denotes a non-central chi-squared variate with $(m)$ degrees of freedom and non-centrality parameter $(d)$, all variates are mutually independent, $a > 0$, and $a_i > 1$, $i = 1, \ldots, r$. Let $F(x)$ denote the c.d.f. of $X$. Then there exist constants $q_i$ such that $F(x) = \sum_{i=0}^{\infty} q_i F_{M+2i}(\frac{x}{a})$, where

$M = m_0 + m_1 + \cdots + m_r$, $q_i > 0$ for each $i$ and $\sum_{i=0}^{\infty} q_i = 1$. The $q_i$ are functions of the $a_i$, and $F_j(x)$ is the c.d.f. of a central chi-squared variate with $(j)$ d.f.

By assumption, $\lambda_p^{(j)}$ is the smallest root, for each $j$.

Define

(3.2.33) $a_{(j)} = \lambda_p^{(j)}$, $a_k^{(j)} = \frac{\lambda_k^{(j)}}{\lambda_p^{(j)}}$, $k = 1, \ldots, p$; $j = 1, 2$.

Now $a_{(j)} > 0$, $a_1^{(j)} > a_2^{(j)} > \cdots > a_p^{(j)} \geq 1$. Rewrite (3.2.31), (3.2.32) as

(3.2.34) $\mathcal{L}(T_0^{(j)}) = \mathcal{L}(a[\chi^2_{p, \lambda_p^{(j)}} + \sum_{k=1}^{p-1} \lambda_k^{(j)} \chi^2_{1, \gamma_k^{(j)}}])$

for $j = 1, 2$. Now (3.2.34) is in the proper form for the last mentioned theorem to be applied. Let $F_j(x)$ denote the c.d.f. of $T_0$ under
H, A; i.e. for \( j = 1, 2 \), respectively. Then obtain

**Theorem 3.2.2:** Let \( T_0 = v' A v \), where \( v = N(\mu_j + A^{-1} b, \Sigma_j) \), \( j = 1 \)
under \( H \) and \( j = 2 \) under \( A \), and \( A > 0 \). Then, there exist constants \( q_i(j) \), defined for each \( j = 1, 2 \) in Chapter 4, which satisfy \( q_i(j) \geq 0 \), \( \sum_0^\infty q_i(j) = 1 \), such that

\[
(3.2.35) \quad F(j)(x) = \sum_{i=0}^\infty q_i(j) F_{p+2i}(x/a(j)),
\]

where \( F_k(x) \) is the c.d.f. of a central chi-squared variate with \( k \) degrees of freedom.

In this section, we have assumed so far that all parameters are known. If in fact, they are unknown, but \( m, n \) are large, \( F^{(1)}(x) \) and \( F^{(2)}(x) \) can be approximated by replacing \( (q_i^{(1)}, a^{(1)}) \) and \( (q_i^{(2)}, a^{(2)}) \) by their estimated values. The estimates can be found as functions of the original maximum likelihood estimates of the population parameters.

Since the \( q_i(j) \) are constants which require considerable effort to compute (particularly without the aid of a machine), approximations to the distribution which can easily be applied with simple computations are discussed in Section 3.2.5.

### 3.2.4 Determination of the Classification Regions

In Section 3.2.3 we have calculated the c.d.f. of \( T_0 \) under both the "H" and "A" hypotheses. It will now be shown how the classification region boundaries depend on these c.d.f.'s.
Define the type I and type II errors for the known parameters case as \( \alpha, \beta \), respectively. Then

\[(3.2.36) \quad \alpha = P(\text{rejecting } H|H) = P(T < c|H)\]

\[(3.2.37) \quad \beta = P(\text{accepting } H|A) = P(T > c|A) .\]

Using (3.2.19), we find

\[(3.2.38) \quad \alpha = P(T_0 + D < c|H) = P(T_0^{(1)} + D < c)\]

\[(3.2.39) \quad \beta = P(T_0 + D > c|A) = P(T_0^{(2)} + D > c) .\]

Equivalently,

\[(3.2.40) \quad \alpha = P(T_0^{(1)} < c-D) = F^{(1)}(c-D)\]

\[(3.2.41) \quad \beta = P(T_0^{(2)} > c-D) = 1 - F^{(2)}(c-D) .\]

Since it is assumed that all parameters are known, \((a_1^{(1)}, q_1^{(1)})\) and \((a_2^{(2)}, q_1^{(2)})\) are known for each \( i \). Hence, \( F^{(1)}(x) \) and \( F^{(2)}(x) \) are known distribution functions.

If there is no prior information available regarding the chances that a given observation will fall into \( \pi_1 \), or \( \pi_2 \), it is reasonable to select a classification scheme based upon the equal error or minimax criterion (with respect to a 0-1 loss function). Equivalently, if there is no reason to favor one type of classification error over the other, we determine the unknown constant "c" subject to the constraint that \( \alpha = \beta \). Then, for the minimax criterion,

\[(3.2.42) \quad F^{(1)}(c-D) + F^{(2)}(c-D) = 1 .\]
Since $P^{(1)}, P^{(2)}$ are both monotone, and $D$ is a known constant, (3.2.42) uniquely determines $c$.

An alternative criterion involving just one type of error is sometimes useful. For example, in a medical diagnosis application, the physician may have a vector of observations on a patient, and may want to classify his malady as one of two diseases. Assume the two diseases are only alike in their early symptoms, but in fact, lack of immediate correct treatment in one causes almost certain death; whereas in the other, the patient will usually recover by himself. In this situation, an equal error criterion would clearly not be indicated, but rather, we would want to weight one type of error, say type I, more than the other. We might then specify a tolerable value of $\alpha$ and determine "c" accordingly from (3.2.40).

In the event there is a priori information available, and perhaps in addition there are computable costs associated with misclassification, the usual type of Bayes solution can be found using the above quantities. This type of solution is carried out in detail in [2] for the case of $\Sigma_1 = \Sigma_2$, and the extension to $\Sigma_1 \neq \Sigma_2$ is immediate.

It is noted that we have merely indicated a method which can be used for solving the classification problem with known parameters. To effect the numerical solution it is undoubtedly most expedient to employ a digital computer. Since the equations involve weighted sums of chi-squared c.d.f.'s it is a formidable but straightforward computational problem. Recursion formulas, useful in numerical computation, for the weight coefficients of the successive $P^{(1)}, P^{(2)}$'s are discussed in Chapter 4. Moreover, since for some purposes, approximate solutions
will suffice, an approximate, easy to apply procedure is provided in the next section.

We next consider the classification region boundaries for the case in which the parameters are unknown. For \( m, n \) large we use Theorem 3.2.3. This is tantamount to using the same boundary discussed above for the case in which the parameters are known; however, in this case, the parameters are replaced by their maximum likelihood estimates to yield \( c^* \), a modified \( c \). When \( m, n \) are not large, we can use the approximate test given in Theorem 3.2.4. Then we use the actual \( U_{m,n} \) instead of \( U \), and use \( (c^* + v^*) \) as the critical constant, where \( v^* \) is defined in (3.2.13). Of course the approximate tests will become exact for large \( m, n \).

3.2.5 Approximate Distribution Theory - Known Parameters.

In Section 3.2.3 it was shown that when the parameters are known and the population covariance matrices are unequal, the c.d.f. of the likelihood ratio test statistic is expressible as a convex combination of c.d.f.'s of central chi-squared variates. Percentage points can be found by direct computation or with the aid of digital computers. However, if the dimensionality of the problem exceeds two or three, the calculations can become quite formidable. For many purposes, it will suffice to have an approximate determination of the fractile values and it will be required to have available a procedure which can be applied easily and quickly. For this purpose several methods of varying precision are available. Two are presented below.
Method I - Cumulant Matching

Define \( U = \sum_{i=1}^{p} b_i X_i^2 \) \(_{m_i, d_i} \) where \( m_i \) denotes the d.f. and \( d_i \) the non-centrality parameter, \( b_i > 0 \) for every \( i \), and the variates are mutually independent. Let \( V \) be a random variable with distribution given by \( f(V) = \rho X_\nu^2 \), where \( \rho, \nu \) are to be determined. The first two cumulants of \( V \) are given, respectively, by \( K_1 = \rho \nu \), \( K_2 = 2\rho^2 \nu \). The first two cumulants of \( U \) are given, respectively, by \( K_1^* = \sum_{i=1}^{p} b_i (m_i + d_i), \)
\( K_2^* = \sum_{i=1}^{p} 2b_i (m_i + 2d_i) \). Equating cumulants \( (K_j = K_j^* ; j = 1, 2) \) and solving for \( \rho \) and \( \nu \) gives

\[
(3.2.43) \quad \rho = \frac{\sum_{i=1}^{p} b_i (m_i + 2d_i)}{\sum_{i=1}^{p} b_i (m_i + d_i)}, \quad \nu = \left[ \frac{\sum_{i=1}^{p} b_i (m_i + d_i)}{\sum_{i=1}^{p} b_i (m_i + 2d_i)} \right]^2.
\]

Using these values of \( \rho, \nu \), we have approximately

\[
(3.2.44) \quad P(U \leq t) \approx P(\rho \chi_\nu^2 \leq t).
\]

Patnaik [23] in examining the characteristics of this type of approximation compared the exact with the approximate values of the associated c.d.f.'s over a wide range of parameter values and percentage points. It was found that in all but one case there was agreement to two significant figures in the c.d.f.

In the one extreme case there was a 20% deviation. This is accounted for in the fact that the density was integrated over an extremely short range so that the variations in the exact and approximate densities could not be smoothed out. Hence, if the c.d.f.'s are used to obtain
ordinates for 95% of the probability mass (or any fractiles: in this vicinity, as long as 5% points are avoided) the method of cumulant matching should be quite adequate for most purposes. However, for situations requiring greater precision, method II is provided.

It should be noted that the approximation introduced in this section involves, in the general case, a chi-squared distribution with a fractional number of degrees of freedom. The associated percentage points can be handled using interpolation. However, it may be more convenient computationally to use an approximation developed by Wilson and Hilferty [38]. Let \( Z = \chi^2_v \). Then approximately, \( z[(Z/v)^{1/3}] = N(1 - 2/9v, 2/9v) \). Thus, for \( U \) defined above, we have approximately

\[
P(U \leq t) \approx \phi \left[ \frac{\left( \frac{t}{\sqrt{v}} \right)^{1/3} - (1 - \frac{2}{9v})}{\sqrt{2/9v}} \right]
\]

where \( \phi(t) \) is the c.d.f. of the standardized normal distribution, and \( \rho, v \) are defined above.

**Method II - Edgeworth Series**

The c.d.f. of \( U \) can be expanded in a series whose leading term is the c.d.f. of \( \rho \chi^2_v \), and from the expansion closer approximations and percentage points will result. To apply the method we compute the cumulants of \( U, V \). For independent variables the \( r^{\text{th}} \) cumulant of a sum is the sum of the individual \( r^{\text{th}} \) cumulants.

\[
K_r(U) = \sum_{i=1}^{P} K_r[b_i \chi^2_{m_i, d_i}] = \sum_{i=1}^{P} b_i^r K_r(\chi^2_{m_i, d_i}).
\]
But \( K_r(\chi^2_{m_i, d_i}) = 2^{r-1}(r-1)! (m_i + r d_i) \). Hence,

\[
K_r(U) = \prod_{i=1}^{p} b_i^r 2^{r-1}(r-1)! (m_i + rd_i).
\]

The cumulants of \( V \) are seen to be given by

\[
K_x(V) = 2^{r-1}(r-1)! \nu^r \frac{2^{r-1}(r-1)!(\sum_{i=1}^{p} b_i (m_i + 2d_i))^{r-1}}{[\sum_{i=1}^{p} b_i (m_i + d_i)]^{r-2}}.
\]

Define \( c_j = K_j(U) - K_j(V) \). Then \( c_1 = c_2 = 0 \), and \( c_j < 0 \) for \( j = 3, 4, \ldots \). Let \( f(y) \) denote the density of a \( \chi^2 \) variate, and \( p(x) \), \( x = \rho y \), denote the density of \( U \). Then, applying the Edgeworth operator (see e.g. Cramér [13]) to \( f(y) \) gives

\[
p(y) = \exp\left[-\frac{c_3}{6\rho^3} \frac{d^3}{dy^3} + \frac{c_4}{24\rho^4} \frac{d^4}{dy^4} + \cdots\right] f(y)
\]

\[
p(y) = [1 + \left(-\frac{c_3}{6\rho^3} D^3 + \frac{c_4}{24\rho^4} D^4 - \cdots\right]
\]

\[
+ \frac{1}{2!} \left[\left(\frac{c_3}{6\rho^3}\right)^2 D^6 + \left(\frac{c_4}{24\rho^4}\right)^2 D^8 - \cdots\right] + \cdots] f(y)
\]

where \( D \) denotes the derivative operator. Hence,

\[
\int_0^y p(y)dy = \int_0^y f(y)dy + \left[\left(-\frac{c_3}{6\rho^3} f''(y) + \frac{c_4}{24\rho^4} f'''(y) + \cdots\right) + \frac{1}{2!} \left(\left(\frac{c_3}{6\rho^3}\right)^2 D^6 + \left(\frac{c_4}{24\rho^4}\right)^2 D^8 - \cdots\right) + \cdots\right] f(y).
\]

Note that since the higher derivatives of \( f(y) \) become smaller in value for a given \( y \), we need only retain the first term in brackets to get an improvement in the approximation over Method I. Then, \( P(U \leq x) = \)

\[
\int_0^x p(x)dx = \int_0^y p(y)dy \approx \int_0^y f(y)dy - \frac{c_3}{6\rho^3} \frac{d^3}{dy^3} \int_0^y f(y)dy,
\]

61
or

\[(3.2.45) \quad P[U \leq x] \approx I(\mu, p) - \frac{c_3}{6p^3(2v)^{3/2}} \frac{d_3^3I(\mu, p)}{d\mu^3},\]

where \(I(\mu, p)\) is the incomplete gamma function

\[q = \frac{\nu}{2} - 1 = \frac{\sum b_i(m_i+d_i)}{2 \sum b_i(m_i+2d_i)}, \quad \text{and} \quad \mu = \frac{\nu}{2\nu} = \frac{x}{\rho\sqrt{2v}} = \sqrt{\frac{\sum b_i(m_i+2d_i)}{2\sum b_i(m_i+2d_i)}}.\]

Now identify the \(\lambda_k^{(j)}\) of Theorem 3.2.3 with the \(a_k\) above

(separately for \(j=1, j=2\)); also the \((\gamma_k^j)^2\) are identified with \(d_k\) and take \(m_i = 1\) for every \(i\). Then approximately

\[(3.2.46) \quad P[T-D \leq t] \approx P \left\{ \chi_v^2 \leq \frac{t}{\sum_{k=1}^{p} \lambda_k^{(j)}[1+(\gamma_k^j)^2]} \right\},\]

where

\[\nu = \frac{\left\{\sum_{k=1}^{p} \lambda_k^{(j)}[1+(\gamma_k^j)^2]\right\}^2}{\sum_{k=1}^{p} \lambda_k^{(j)}[1+2(\gamma_k^j)^2]}.\]

This is the method I approximation. For the method II approximation of \((3.2.45)\) take

\[c_3 = 8 \left\{ \sum_{k=1}^{p} \left(\lambda_k^{(j)} \right)^3 \left[1+(\gamma_k^j)^2\right] - \left[\sum_{k=1}^{p} \lambda_k^{(j)} \left[1+(\gamma_k^j)^2\right]\right] \right\}^2,\]

62
U \equiv T-D, (v, \rho) the same as in (3.2.46).

Note that the above approximations can also be used for the unknown parameters case by substituting estimates for population values.

3.2.6 Approximate Distribution Theory - Unknown Parameters.

In this section we describe a method for approximating the distribution of a function of sample moments and apply the method to approximate the distribution of the classification statistic \( U_{m,n} \) defined in (3.2.4). The approximation is better, though more complicated, than the one obtained in Section 3.2.2 by letting sample moments approach their expected values.

Cramér (see [13], p. 366) has shown that a well behaved function of two sample moments asymptotically has a normal distribution with moments simply derived from the Taylor Series Expansion of the function. This theorem is now generalized to scalar valued functions of vectors and matrices. The proof is an immediate extension of Cramér's proof.

Theorem 3.2.3: Let \( \Omega \equiv (\omega_{ij}) \) denote a \( p \times q \) random matrix whose columns (or subcollections of columns) are sample moments. Let \( f(\Omega) \) be scalar valued, continuous, with continuous derivatives of the first and second order with respect to each of the elements of \( \Omega \), and not an explicit function of the sample size. Then the asymptotic distribution of \( f(\cdot) \) is normal with approximate mean and variance given by

\[
E(f) \cong f(E\Omega),
\]

(3.2.47)
(3.2.48) \( \text{Var}(f) = \sum_{i=1}^{P} \sum_{j=1}^{Q} \sum_{l=1}^{P} \sum_{k=1}^{Q} \frac{\partial f(\Omega)}{\partial \omega_{ij}} \frac{\partial f(\Omega)}{\partial \omega_{kl}} \text{Cov}(\omega_{ij}, \omega_{kl}) \),

where \( \frac{\partial f(\Omega)}{\partial \omega_{ij}} \) denotes a derivative evaluated at \( \Omega = E\Omega \).

If many variables are involved and the associated distributions are arbitrary, the asymptotic variance may involve extensive and laborious computations. However, under certain conditions, considerable simplification can result. For example, consider a \( p \times p \) random matrix \( \Omega \) whose probability density function is given by

\[
p(\Omega) = c |\Omega|^{\frac{n-p-1}{2}} |\Sigma|^{-\frac{n}{2}} \exp \left( -\frac{1}{2} \text{tr} \Sigma^{-1} \Omega \right),
\]

where \( c \) is a normalizing integration constant. Such a matrix is said to have a Wishart distribution with scale matrix \( \Sigma \), dimension \( p \), and \( n \) degrees of freedom. The law is denoted by \( W(\Sigma, p, n) \).

**Corollary 3.2.1:** Let \( \Omega \) be a \( p \times p \) random matrix whose distribution is given by \( p(\Omega) = W(\Sigma/n, p, n) \). If \( f(\Omega) \) is scalar valued, continuous with continuous derivatives of first and second order with respect to each of the elements of \( \Omega \), and not an explicit function of the sample size, the asymptotic variance of \( f \) given in (3.2.48) is representable as \( \left( \frac{\Sigma}{n} \right) \text{tr}(B\Sigma)^2 \), where \( B = \left( \frac{\partial f(\Omega)}{\partial \omega_{ij}} \right) \).

**Proof:** Expand \( \text{tr}(B\Sigma)^2 \) in terms of elements, recalling that if \( \Sigma = (\sigma_{ij}) \), \( \text{Cov}(\omega_{ij}, \omega_{kl}) = \frac{\sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{jk}}{n} \).

Now apply this theorem to \( U_{m,n} \) (see (3.2.4)). For large \( m, n \), almost surely we have
\[(3.2.49) \quad U_{m,n} \sim \mathcal{F} n \left| \frac{W}{V} \right| + n(y-z)'W^{-1}(y-z)-m(x-z)'V^{-1}(x-z). \]

Define the covariance matrices \(S, S^*\) by

\[(3.2.50) \quad S = \frac{W}{n-1}, \quad S^* = \frac{V}{m-1}. \]

Assume \(n, m\) approach infinity in a well behaved way so that \(n/m\) is constant. Then if \(\delta = p \ln(n-1) - p \ln(m-1),\)

\[(3.2.51) \quad U_{m,n} - \delta \sim \mathcal{F} n|S| - \ln|S^*| + (y-z)'S^{-1}(y-z) - (x-z)'(S^*)^{-1}(x-z). \]

Next define the \(\Omega\) matrix of dimension \(p \times (2p+3)\) as

\[(3.2.52) \quad \Omega = (S, S^*, x, y, z), \]

and \(f(\Omega)\) is the right hand side of \((3.2.51)\). The distributions of the five independent variates above are respectively,

\[(3.2.53) \quad \mathcal{F}(S) = W\left(\frac{\Sigma_2}{n-1}, p, n-1\right); \quad \mathcal{F}(S^*) = W\left(\frac{\Sigma_1}{m-1}, p, m-1\right); \]

\[(3.2.54) \quad \mathcal{N}(x) = N(\mu_1, \frac{\Sigma_1}{m}), \quad \mathcal{N}(y) = N(\mu_2, \frac{\Sigma_2}{n}), \quad \mathcal{N}(z) = N(\mu_j, \Sigma_j), \]

where \(j=1\) under \(H\), \(j=2\) under \(A\). Hence, it follows that under \(H\),
(3.2.55) \quad E(\Omega) = (\Sigma_2^2, \Sigma_1^1, \mu_1^1, \mu_2^2, \mu_1^1) ,

and therefore,

(3.2.56) \quad E(f) \approx \ln |\Sigma_2^{-1}-\frac{\ln |\Sigma_1^1|}{(\mu_1^1-\mu_2^2)^T \Sigma_2^{-1}(\mu_1^1-\mu_2^2)} .

Under A,

(3.2.57) \quad E(\Omega) = (\Sigma_2^2, \Sigma_1^1, \mu_1^1, \mu_2^2, \mu_2^2) ,

so that

(3.2.58) \quad E(f) \approx \ln |\Sigma_2^{-1}| - \frac{\ln |\Sigma_1^1| - (\mu_1^1-\mu_2^2)^T \Sigma_1^{-1}(\mu_1^1-\mu_2^2)} .

Because of independence, the asymptotic variance of f can be evaluated separately for each of the components of \( \Omega \). Let \( \text{Var}_A(f) \) denote the contribution to the variance of f due to the component A of \( \Omega \). For S and S* we can immediately apply Corollary 3.2.1 obtaining:

\[
\text{Var}_S(f) = \frac{2}{n-1} \text{tr} \left( B_2 \Sigma_2 \right)^2 , \quad \text{Var}_{S^*}(f) = \frac{2}{m-1} \text{tr} \left( B_1 \Sigma_1 \right)^2 ,
\]

where

\[
B_2 = \left( \frac{\partial f(\Omega)}{\partial s_{ij}} \right) , \quad B_1 = \left( \frac{\partial f(\Omega)}{\partial s^*_{ij}} \right) ,
\]

\[ s = (s_{ij}) , \quad s^* = (s^*_{ij}) .
\]
From the definition of $f$ in (3.2.51) it is readily found that under $H$

$$B_2 = \Sigma_2^{-1} - \Sigma_2^{-1}(\mu_2-\mu_1)(\mu_2-\mu_1)'\Sigma_2^{-1}, \quad B_1 = -\Sigma_1^{-1};$$

and under $A$,

$$B_2 = \Sigma_2^{-1}, \quad B_1 = -\Sigma_1^{-1} + \Sigma_1^{-1}(\mu_2-\mu_1)(\mu_2-\mu_1)'\Sigma_1^{-1}.$$  

Hence, under $H$,

$$\text{Var}_{S}(f) = \frac{2}{n-1} \left\{ p-2(\mu_2-\mu_1)'\Sigma_2^{-1}(\mu_2-\mu_1) + [(\mu_2-\mu_1)'\Sigma_2^{-1}(\mu_2-\mu_1)]^2 \right\},$$

$$\text{Var}_{S*}(f) = \frac{2p}{m-1}.$$  

Analogously, under $A$,

$$\text{Var}_{S}(f) = \frac{2p}{n-1},$$

$$\text{Var}_{S*}(f) = \frac{2}{m-1} \left\{ p-2(\mu_2-\mu_1)'\Sigma_1^{-1}(\mu_2-\mu_1) + [(\mu_2-\mu_1)'\Sigma_1^{-1}(\mu_2-\mu_1)]^2 \right\}.$$  

Next, we compute the contributions to $\text{Var}(f)$ due to $x,y,z$. Define the vectors

$$h_1 = \left( \frac{\partial f(E\Omega)}{\partial x_1} \right), \quad h_2 = \left( \frac{\partial f(E\Omega)}{\partial y_1} \right), \quad h_3 = \left( \frac{\partial f(E\Omega)}{\partial z_1} \right).$$
From (3.2.48) it is seen that

$$\text{Var}_x(f) = \frac{1}{m} h_1^T \Sigma_1 h_1, \quad \text{Var}_y(f) = \frac{1}{n} h_2^T \Sigma_2 h_2,$$

$$\text{Var}_z(f) = h_j^T \Sigma_j h_j,$$

where $j=1$ under $H$, and $j=2$ under $A$. By direct computation we find that under $H$,

$$h_1 = 0, \quad h_2 = 2\Sigma_2^{-1}(\mu_2 - \mu_1), \quad h_3 = 2\Sigma_2^{-1}(\mu_1 - \mu_2),$$

and under $A$,

$$h_1 = 2\Sigma_1^{-1}(\mu_2 - \mu_1), \quad h_2 = 0, \quad h_3 = 2\Sigma_1^{-1}(\mu_1 - \mu_2).$$

Hence, under $H$

$$\text{Var}_x(f) = 0, \quad \text{Var}_y(f) = \frac{4}{n} (\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1),$$

$$\text{Var}_z(f) = 4(\mu_2 - \mu_1)^T \Sigma_1^{-1}(\mu_2 - \mu_1);$$

and under $A$,

$$\text{Var}_x(f) = \frac{4}{m} (\mu_2 - \mu_1)^T \Sigma_1^{-1}(\mu_2 - \mu_1), \quad \text{Var}_y(f) = 0,$$

$$\text{Var}_z(f) = 4(\mu_2 - \mu_1)^T \Sigma_2^{-1}(\mu_2 - \mu_1).$$
The five contributions are now totaled to obtain the final result. It is that under H,

\[(3.2.59) \quad \text{Var}(\hat{r})_H \approx 2p\left(\frac{1}{m-1} + \frac{1}{n-1}\right) + 4(\mu_2 - \mu_1)'\Sigma_1^{-1}(\mu_2 - \mu_1)\]

\[+ 2(\mu_2 - \mu_1)'\Sigma_2^{-1}(\mu_2 - \mu_1)\left[\frac{2}{n} + \frac{(\mu_2 - \mu_1)'\Sigma_2^{-1}(\mu_2 - \mu_1) - 2}{n-1}\right],\]

and under A,

\[(3.2.60) \quad \text{Var}(\hat{r})_A \approx 2p\left(\frac{1}{m-1} + \frac{1}{n-1}\right) + 4(\mu_2 - \mu_1)'\Sigma_2^{-1}(\mu_2 - \mu_1)\]

\[+ 2(\mu_2 - \mu_1)'\Sigma_1^{-1}(\mu_2 - \mu_1)\left[\frac{2}{m} + \frac{(\mu_2 - \mu_1)'\Sigma_1^{-1}(\mu_2 - \mu_1) - 2}{m-1}\right].\]

Hence, for large m, n we have approximately that

\[\mathcal{L}(U_{m,n}) = \mathcal{N}[E(\hat{r}), \text{Var}(\hat{r})], \quad \text{where } E(\hat{r}) \text{ is defined in (3.2.56) and (3.2.58), and Var}(\hat{r}) \text{ is defined in (3.2.59) and (3.2.60). Recall that to evaluate these equations numerically it is necessary to replace parameters by their estimates.}\]

3.3 Classification with Intraclass Structure, \(\Sigma_1 > \Sigma_2\).

In this section we develop likelihood ratio procedures and asymptotic tests for the problem of classification with intraclass structure for both known and unknown population parameters and for both equal and unequal covariance matrices. It is assumed, however, that \(\Sigma_1 > \Sigma_2\).
3.3.1 Equal Covariance Matrices - Known Parameters.

The problem is to test $H : z \in N(\mu_1, \Sigma)$, vs. $A : z \in N(\mu_2, \Sigma)$, where $\Sigma = \sigma^2[(1-\rho)I + \rho e_p e_p^t]$, and $(\mu_1, \mu_2, \sigma^2, \rho)$ are all known. The Neyman-Pearson Lemma immediately yields the UMP level $\alpha$ test: reject $H$ at level $\alpha$ if $w < \text{constant}$, where

\[
(3.3.1) \quad w = \frac{b^t z}{(b^t \Sigma b)^{1/2}} , \quad b = 2\Sigma^{-1}(\mu_1 - \mu_2) ,
\]

and under $H$, $A$, respectively,

\[
(3.3.2) \quad x(w) = N\left(\frac{b^t \mu_1}{(b^t \Sigma b)^{1/2}}, 1\right) , \quad x(w) = N\left(\frac{b^t \mu_2}{(b^t \Sigma b)^{1/2}}, 1\right) .
\]

It is readily verifiable that for the special $\Sigma$ defined above,

\[
(3.3.3) \quad b^t \Sigma b = \frac{\lambda^4}{\sigma^2(1-\rho)} \left\{ (\mu_1 - \mu_2)'(\mu_1 - \mu_2) - \frac{\rho}{1+(p-1)\rho} [(\mu_1 - \mu_2)' e_p]^2 \right\} ,
\]

\[
(3.3.4) \quad b = \frac{2}{\sigma^2(1-\rho)} \left[I - \frac{\rho e e^t}{1+(p-1)\rho}\right] (\mu_1 - \mu_2) .
\]

3.3.2 Equal Covariance Matrices - Unknown Parameters.

Assume, as in Section 3.2.1, that the data matrices $X, Y$ are available. Let $\Gamma$ be any orthogonal matrix whose first row has the form $p^{-1/2} e_p^t$. Define

\[
(3.3.5) \quad x = \bar{x} = \frac{\Gamma X e_m}{m} , \quad y = \bar{y} = \frac{\Gamma Y e_n}{n} ,
\]
(3.3.6) \[ V = \Gamma X \left( I - \frac{e_i e_i'}{m} \right) X' \Gamma', \quad W = \Gamma Y \left( I - \frac{e_j e_j'}{n} \right) Y' \Gamma'. \]

The matrix \( \Gamma \) diagonalizes the intraclass structured covariance matrices. The minimal sufficient statistics are defined in terms of \( V, W \). Let

(3.3.7) \[ v = V_{11} + W_{11}, \quad w = \text{tr}(V + W) - v; \quad V = (V_{ij}), \quad W = (W_{ij}). \]

Clearly \((x, y, v, w)\) is a minimal sufficient statistic with components independently distributed as

(3.3.8) \[ x(x) = N(\Gamma_1, \frac{\psi}{m}), \quad x(y) = N(\Gamma_2, \frac{\psi}{n}), \]

(3.3.9) \[ x(\alpha) = \chi^2_{m+n+2}, \quad x(\beta) = \chi^2_{(p-1)(m+n+2)}, \]

where

(3.3.10) \[ \alpha = \sigma^2 [1+(p-1)\rho], \quad \beta = \sigma^2 (1-\rho), \quad \psi = \text{diag}(\alpha, \beta, \ldots, \beta). \]

Let \( z^* = \Gamma z \), so that \( x(z^*) = N(\Gamma_1, \psi) \), \( \mu = \mu_1 \) under \( H \), and \( \mu = \mu_2 \) under \( A \). Denote the components of \( x, y, z^* \) by \( x_j, y_j, z_j^* \), respectively; \( j = 1, \ldots, p \). The likelihood ratio statistic \( \lambda \) for testing \( H : \{X\xi_1, z\xi_1, Y\xi_2\} \), vs. \( A : \{X\xi_1, Y\xi_2, z\xi_2\} \) is found to be given by
\[ (3.3.11) \quad \frac{2}{\lambda^{m+n+1}} = \left[ \frac{v + \frac{n}{n+1} (z_1^* - y_1^*)^2}{v + \frac{m}{m+1} (z_1^* - x_1^*)^2} \right] \left[ \frac{w + \frac{n}{n+1} \sum_{j=2}^{p} (z_j^* - y_j^*)^2}{w + \frac{m}{m+1} \sum_{j=2}^{p} (z_j^* - x_j^*)^2} \right]^{p-1}. \]

It is now shown that for large sample sizes, the likelihood ratio test produces an easily applied procedure. It is shown that in that case, the test statistic is normally distributed. Define

\[ (3.3.12) \quad M_1 = \frac{(m+n-2)(n)}{n+1} \frac{(z_1^* - y_1^*)^2}{v}; \quad M_2 = \frac{(m+n-2)(m)}{m+1} \frac{(z_1^* - x_1^*)^2}{v}. \]

\[ (3.3.13) \quad M_3 = \frac{(m+n-2)(p-1)n}{n+1} \frac{\sum_{j=2}^{p} (x_j^* - y_j^*)^2}{w}; \quad M_4 = \frac{(m+n-2)(p-1)m}{m+1} \frac{\sum_{j=2}^{p} (z_j^* - x_j^*)^2}{w}. \]

Then simple algebra gives

\[ (3.3.14) \quad \lambda^2 = \left[ 1 + \frac{M_1}{m+n-2} \right]^{m+n+1} \left[ 1 + \frac{M_3}{(m+n-2)(p-1)} \right]^{(m+n+1)(p-1)} \cdot \]

It is seen below that \( M_1, M_2, M_3, M_4 \) remain finite in probability (actually almost surely since they are based on maximum likelihood estimates) as \( m, n \to \infty \).

Hence, it is evident that

\[ (3.3.15) \quad \lim_{m,n \to \infty} (\lambda^2) = \exp[M_1 - M_2 + M_3 - M_4], \text{ a.s.} \]
Asymptotically, therefore, an equivalent test procedure is to reject $H$ if
\begin{equation}
(3.3.16) \quad M_1 - M_2 + M_3 - M_4 < C = \text{constant}.
\end{equation}

Define
\begin{equation}
(3.3.17) \quad \theta = \Gamma_{\mu_1}, \quad \varphi = \Gamma_{\mu_2},
\end{equation}
and let $\theta_j, \varphi_j, \ j = 1, \ldots, p$ denote their components. Also define
\begin{equation}
(3.3.18) \quad W = \frac{(z_1^* - y_1)^2(m+n-2)}{v} - \frac{(z_1^* - x_1)^2(m+n-2)}{v} + \frac{(p-1)(m+n-2)\sum_j (z_j^* - y_j)^2}{2v} - \frac{(p-1)(m+n-2)\sum_j (z_j^* - x_j)^2}{2v}.
\end{equation}

From (3.3.13), (3.3.16) it is clear that an equivalent asymptotic test is to reject $H$ if $W < c$. The distribution of $W$ can be approximated for large $m, n$ by letting all sample moments pass to their expectations, as was done in Section 3.2.2. However, a more precise approximation can be used which is based upon the multivariate extension of Cramer's theorem on functions of sample moments (see Theorem 3.2.3).

Define the six sample moments
\begin{equation}
(3.3.19) \quad \omega = z_1^* - y_1, \quad \nu = z_1^* - x_1, \quad \gamma = \frac{\nu}{m+n-2},
\end{equation}
\[ (3.3.20) \quad \delta = \frac{w}{(p-1)(m+n-2)}, \quad \tau = \sum_{j=2}^{p} (z_j^* - y_j)^2, \quad \xi = \sum_{j=1}^{p} (z_j^* - x_j)^2. \]

In terms of these variables, we can define

\[ (3.3.21) \quad f(\omega, \gamma, \nu, \tau, \xi, \delta) = W = \frac{\omega^2}{\gamma} - \frac{\nu^2}{\gamma} + \frac{\tau}{\delta} - \frac{\xi}{\delta}. \]

In applying Theorem 3.2.3 to approximate the distribution, the first point to note is that by independence, all covariances vanish except for \( \text{Cov}(\omega, \nu) \) and \( \text{Cov}(\tau, \xi) \). Moreover, it is shown below that \( \text{Cov}(\omega, \nu) \) need not be evaluated since its coefficient \( (\partial f/\partial \omega)(\partial f/\partial \nu) \) vanishes under both H and A. Hence, only \( \text{Cov}(\tau, \xi) \) is evaluated. The distributions of the sample moments are tabulated below, for convenience.

<table>
<thead>
<tr>
<th></th>
<th>Under H</th>
<th>Under A</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>( N(\theta_1 - \varphi_1, \frac{n+1}{n} \alpha) )</td>
<td>( N(0, \frac{n+1}{n} \alpha) )</td>
</tr>
<tr>
<td>( \nu )</td>
<td>( N(0, \frac{m+1}{m} \alpha) )</td>
<td>( N(\varphi_1 - \theta_1, \frac{m+1}{m} \alpha) )</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>( \frac{\alpha}{m+n-2} \chi^2_{m+n-2} )</td>
<td>( \frac{\alpha}{m+n-2} \chi^2_{m+n-2} )</td>
</tr>
<tr>
<td>( \delta )</td>
<td>( \beta \frac{(p-1)(m+n-2)}{(p-1)(m+n-2)} \chi^2_{p-1}(m+n-2) )</td>
<td>( \beta \frac{(p-1)(m+n-2)}{(p-1)(m+n-2)} \chi^2_{p-1}(m+n-2) )</td>
</tr>
<tr>
<td>( \tau )</td>
<td>( \beta \frac{(n+1)}{n} \chi^2_{p-1} \left[ \frac{1}{n+1} \sum_{j=1}^{p} (\theta_j - \varphi_j)^2 \right] )</td>
<td>( \beta \frac{(n+1)}{n} \chi^2_{p-1} \left[ \frac{1}{n+1} \sum_{j=1}^{p} (\theta_j - \varphi_j)^2 \right] )</td>
</tr>
<tr>
<td>( \xi )</td>
<td>( \beta \frac{(m+1)}{m} \chi^2_{p-1} \left[ \frac{1}{n+1} \sum_{j=1}^{p} (\theta_j - \varphi_j)^2 \right] )</td>
<td>( \beta \frac{(m+1)}{m} \chi^2_{p-1} \left[ \frac{1}{n+1} \sum_{j=1}^{p} (\theta_j - \varphi_j)^2 \right] )</td>
</tr>
</tbody>
</table>
The quantities needed are computed first under $H$, and then under $A$.

From the above table it is easy to see that since under $H$

$$E(\omega, \gamma, \nu, \tau, \xi, \delta) = \left[ \theta_1 - \phi_1, \alpha, 0, (p-1)\beta + \frac{\beta}{2} (\theta_j - \phi_j)^2, \beta(p-1) + \frac{\beta}{m}(p-1), \beta \right] ,$$

$$E(f)_H = (\theta - \phi)' \psi^{-1}(\theta - \phi) + (p-1)(\frac{1}{n} - \frac{1}{m}) .$$

(3.3.22)

The variance terms are also found from the table to be under $H$

$$(\text{Var } \omega, \text{Var } \gamma, \text{Var } \nu, \text{Var } \tau, \text{Var } \xi, \text{Var } \delta)$$

$$= \left[ \frac{\alpha}{n}, \frac{2\alpha^2}{m+n-2}, \alpha(1 + \frac{1}{m}), \
2\beta^2(p-1) + 4\beta \frac{P}{n} (\theta_j - \phi_j)^2 + \frac{4\beta^2(p-1)}{n} + \frac{4\beta}{n} \sum \frac{P}{2} (\theta_j - \phi_j)^2, \
2\beta^2(p-1) + \frac{4\beta^2(p-1)}{m}, \frac{2\beta^2}{(p-1)(m+n-2)} \right] .$$

All partial derivatives involve simple computations and, therefore, the details have been omitted.

The only significant term remaining is $\text{cov}(\tau, \xi)$.

By definition,

$$\text{Cov}(\tau, \xi) = \sum_{j=2}^{P} \sum_{k=2}^{P} \text{cov} \left[ (z^*_j - y_j)^2, (z^*_k - x_k)^2 \right] .$$

Using the linearity property of covariances, and the mutual independence of the $x's, y's, and z^*'s$, direct computation gives
\[
\text{Cov} (\tau, \xi) = 2\beta^2(p-1).
\]

Combining terms, we obtain under H

\begin{equation}
(3.3.23) \quad \text{Var}(f)_{H} = \frac{4}{n}(\theta - \phi)^{\prime}\psi^{-1}(\theta - \phi) + \frac{4}{n} [(\theta - \phi)^{\prime}\psi^{-1}(\theta - \phi) + p-1] \\
+ \frac{4}{m} \frac{(p-1)}{m+n-2} \left[ \frac{(\theta_1 - \phi_1)^4}{\alpha^2} + \frac{\left\{ \frac{p}{2} \sum_{j=1}^{p} (\theta_j - \phi_j)^2 \right\}^2}{(p-1)\beta^2} \right].
\end{equation}

Under A the evaluation is completely analogous and, therefore, is omitted. The results under A are given below.

\begin{equation}
(3.3.24) \quad \text{E}(f)_{A} = -(\theta - \phi)^{\prime}\psi^{-1}(\theta - \phi) + (p-1)\left( \frac{1}{n} - \frac{1}{m} \right),
\end{equation}

\begin{equation}
(3.3.25) \quad \text{Var}(f)_{A} = \frac{4}{n}(\theta - \phi)^{\prime}\psi^{-1}(\theta - \phi) + \frac{4}{m} [(\theta - \phi)^{\prime}\psi^{-1}(\theta - \phi) + p-1] \\
+ \frac{4}{n} \frac{(p-1)}{m+n-2} \left[ \frac{(\theta_1 - \phi_1)^4}{\alpha^2} + \frac{\left\{ \frac{p}{2} \sum_{j=1}^{p} (\theta_j - \phi_j)^2 \right\}^2}{(p-1)\beta^2} \right].
\end{equation}

All of the above results on the approximate distribution are summarized in

**Theorem 3.3.1:** Let \( z \) be a \( p \)-variate observation to be classified into \( N(\mu_1, \Sigma_1) \) or \( N(\mu_2, \Sigma_2) \). If \( p \)-variate samples of sizes \( m,n \) are taken from the two populations, and \( m,n \) are large, the likelihood ratio procedure is approximately to classify \( z \) into \( N(\mu_2, \Sigma_2) \) if \( W < C \) where \( W \) is defined in (3.3.18), \( C \) is determined from \( \chi(W) \approx N(E(f), \text{Var}(f)) \), and \( E(f), \text{Var}(f) \) are given under H and A.
in (3.3.22)-(3.3.25), respectively. In these equations, the parameters are replaced by their maximum likelihood estimates for numerical computations.

3.3.3 Unequal Covariance Matrices - Known Parameters.

In this section we address ourselves to the same problem as in Section 3.2.1 except that here $\Sigma_1$ and $\Sigma_2$ are assumed to have intraclass structure. The Neyman-Pearson derived test statistic $T$ given in equation (3.2.1) still applies, with $\Sigma_1, \Sigma_2$ restricted to

$$
(3.3.26) \quad \Sigma_i = \sigma_i^2 [(1-\rho_i)I + \rho_i\mathbf{e}_i\mathbf{e}_i'] , \quad i = 1,2 .
$$

For this case, $T$ is defined explicitly in (3.3.40).

We use the same initial transformations as in Section 3.2.3 to find the distribution of $T$. Recall from equation (3.2.19) that $T_0$ is just $T$ translated by a known constant. Let $\Gamma$ be any orthogonal matrix whose first row is $(p)^{-1/2}\mathbf{e}_p'$. Now by making judicious identifications with the quantities defined in Section 3.2.3, it is possible to avoid repetition and find that for this restricted case, the law of $T_0$ is expressible as a sum of two, instead of $p$, chi-squared variates. For both $\Gamma_1, \Gamma_2$ defined in (3.2.25), take the $\Gamma$ just defined. Then the $D_j$ in (3.2.25) become

$$
(3.3.27) \quad D_1 = \text{diag}\left(\frac{\alpha_1-\alpha_2}{\alpha_1}, \frac{\beta_1-\beta_2}{\beta_1}, \ldots, \frac{\beta_1-\beta_2}{\beta_2}\right) ,
$$

$$
D_2 = \text{diag}\left(\frac{\alpha_1-\alpha_2}{\alpha_1}, \frac{\beta_1-\beta_2}{\beta_1}, \ldots, \frac{\beta_1-\beta_2}{\beta_1}\right) ,
$$

77
where \( \alpha_j, \beta_j \) are defined in (1.2.1), (1.2.2). With these identifications, equation (3.2.31) simplifies since \((p-1)\) of the chi-squared variates are independent with the same weight coefficients. Therefore, we obtain under \( H \)

\[
(3.3.28) \quad \chi(T_0) = \chi\left( \frac{\alpha_1 - \alpha_2}{\alpha_2} \chi_1 \left[ (\gamma_1^{(1)})^2 \right] + \frac{\beta_1 - \beta_2}{\beta_2} \chi_p \left[ \sum_{k=2}^{p} (\gamma_k^{(1)})^2 \right] \right),
\]

and under \( A \)

\[
(3.3.29) \quad \chi(T_0) = \chi\left( \frac{\alpha_1 - \alpha_2}{\alpha_1} \chi_1 \left[ (\gamma_1^{(2)})^2 \right] + \frac{\beta_1 - \beta_2}{\beta_1} \chi_p \left[ \sum_{k=2}^{p} (\gamma_k^{(2)})^2 \right] \right).
\]

For this problem, the \( \gamma_k^{(j)} \) are given by

\[
(3.3.30) \quad \gamma_1^{(1)} = \frac{\phi_1^{(1)}}{\sqrt{\frac{\alpha_1 - \alpha_2}{\alpha_2}}}, \quad \gamma_1^{(2)} = \frac{\phi_1^{(2)}}{\sqrt{\frac{\alpha_1 - \alpha_2}{\alpha_1}}},
\]

\[
(3.3.31) \quad \frac{p}{2} (\gamma_k^{(1)})^2 = \left( \frac{\beta_2}{\beta_1 - \beta_2} \right) \frac{p}{2} (\phi_k^{(1)})^2, \quad \frac{p}{2} (\gamma_k^{(2)})^2 = \left( \frac{\beta_1}{\beta_1 - \beta_2} \right) \frac{p}{2} (\phi_k^{(2)})^2,
\]

where the \( \phi_k^{(j)} \) may be found to be

\[
(3.3.32) \quad \phi_1^{(1)} = \sqrt{\frac{\alpha_1}{\alpha_2 (\alpha_1 - \alpha_2)}} (\theta_1 - \phi_1), \quad \phi_1^{(2)} = \sqrt{\frac{\alpha_2}{\alpha_1 (\alpha_1 - \alpha_2)}} (\theta_1 - \phi_1),
\]

\[
(3.3.33) \quad \phi_k^{(1)} = \sqrt{\frac{\beta_1}{\beta_2 (\beta_1 - \beta_2)}} (\theta_k - \phi_k), \quad \phi_k^{(2)} = \sqrt{\frac{\beta_2}{\beta_1 (\beta_1 - \beta_2)}} (\theta_k - \phi_k),
\]

78
for \( k = 2, \ldots, p \). Note that \((\varphi_1, \ldots, \varphi_p)\) are the components of \( \Gamma_{\mu_2} \), whereas \( \varphi^{(j)} \) are the components of \( \varphi^{(j)} \) defined in (3.2.26). Also note that since the original derivation assumed that \( \Sigma_1 = \Sigma_2 > 0 \), in this problem we insist that \( \alpha_1 > \alpha_2 \), \( \beta_1 > \beta_2 \). Assume the parameters are labeled so that \( \alpha_1 / \alpha_2 > \beta_1 / \beta_2 \).

Note that the distribution of \( T_0 \) given in (3.3.28), (3.3.29) can be approximated, when desired, by the methods of Section 3.2.5. To effect the approximation take \( p = 2 \) and identify

\[ b_1, b_2 \] of that section with \( \left( \frac{\alpha_1 - \alpha_2}{\alpha_2} \right), \left( \frac{\beta_1 - \beta_2}{\beta_2} \right) \) under \( H \), respectively, and \( \left( \frac{\alpha_1 - \alpha_2}{\alpha_1} \right), \left( \frac{\beta_1 - \beta_2}{\beta_1} \right) \), respectively, under \( A \).

Take \( m_1 = 1; \ m_2 = p - 1; \ d_1 = \left( \gamma_1^{(j)} \right)^2; \ d_2 = \sum_{k=2}^{p} \gamma_k^{(j)} \).

By analogy with (3.2.33), let

\[ (3.3.34) \quad a^{(1)} = \frac{\beta_1 - \beta_2}{\beta_2}, \quad a^{(2)} = \frac{\beta_1 - \beta_2}{\beta_1}, \]

\[ a^{(1)}_1 = \left( \frac{\beta_2}{\alpha_2} \right) \left( \frac{\alpha_1 - \alpha_2}{\beta_1 - \beta_2} \right), \quad a^{(2)}_1 = \left( \frac{\beta_1}{\alpha_1} \right) \left( \frac{\alpha_1 - \alpha_2}{\beta_1 - \beta_2} \right), \]

so that \( a^{(1)} > 0, \ a^{(2)} > 0, \ a^{(1)}_1 \geq 1, \ a^{(2)}_1 \geq 1 \). Then,

\[ z^{(m_1^{(j)})} = z \left\{ a^{(j)} \left[ \chi^2_{p-1} \left( \frac{\beta_2}{\alpha_2} \right) \left( \gamma_1^{(j)} \right)^2 \right] + a^{(j)}_1 \chi^2_{1} \left( \gamma_1^{(1)} \right)^2 \right\}, \quad j = 1, 2. \]

Analogous with Theorem 3.2.4 is
Theorem 3.3.2: Let $T_0$ be defined as in (3.2.30) for the case of intraclass structure in which $\alpha_1 > \alpha_2$, $\beta_1 > \beta_2$. Then, there exist constants $q_1^{(j)}$, defined for each $j = 1, 2$ in Chapter 4 in terms of $(a(j), a_1(j))$ of (3.3.34) satisfying $q_1^{(j)} \geq 0$, $\sum_0^\infty q_1^{(j)} = 1$, such that

\[(3.3.35) \quad F(j)(x) = \sum_{i=0}^\infty q_1^{(j)} F_{p+2i}(x/a(j)) ,\]

where $F_k(x)$ is the c.d.f. of a central chi-squared variate with $k$ degrees of freedom.

3.3.4 Unequal Covariance Matrices - Unknown Parameters.

As in Section 3.2.1 assume the sample matrices $X,Y$ are available for testing $H: \{X,z \in N(\mu_1, \Sigma_1); \ Y \in N(\mu_2, \Sigma_2)\}$ vs.

$A: \{X \in N(\mu_1, \Sigma_1); \ Y,z \in N(\mu_2, \Sigma_2)\}$. However, $\Sigma_1, \Sigma_2$ are assumed now to have intraclass structure. Let $\Gamma$ be any orthogonal matrix, whose first row is $(p)^{-1/2} e_1^t$. Define $(x, y, V, W)$ as in (3.3.5), (3.3.6). If $V = (V_{ij})$, $W = (W_{ij})$, and we define

\[(3.3.36) \quad v = V_{1l}, \quad w = W_{1l}, \quad r = Tr(V) - v, \quad s = Tr(W) - w ,\]

$(x, y, v, w, r, s)$ is seen to be a minimal sufficient statistic with mutually independent components whose laws are given below.

\[(3.3.37) \quad \mathcal{L}(x) = N(\Gamma \mu_1, \frac{D_1^2}{m}), \quad \mathcal{L}(y) = N(\Gamma \mu_2, \frac{D_2^2}{n})\]
(3.3.38) \( x(v) = \alpha_1 x^2 \), \( x(u) = \alpha_2 x^2 \),

(3.3.39) \( x(r) = \beta_1 x(m-l)(p-l) \), \( x(s) = \beta_2 x(n-l)(p-l) \),

where \( D_j = \text{diag}(\alpha_j, \beta_j, \ldots, \beta_j) \), \( j = 1, 2 \), and \( \alpha_j, \beta_j \) are defined in (1.2.1), (1.2.2)

If we define \( z^* = \Gamma z \), and let

(3.3.40) \( \lambda^* = \frac{\sqrt{m \cdot m(p-l)}}{w \cdot n \cdot n(p-l)} \left[ \frac{w + \frac{n}{n+1}(y_1 - z^*_1)^2}{v + \frac{m}{m+1}(x_1 - z^*_1)^2} \right]^{m+1} \left[ \frac{s + \frac{n}{n+1} \sum_{j=2}^{p} (x_j - z^*_j)^2}{r + \frac{m}{m+1} \sum_{j=2}^{p} (x_j - z^*_j)^2} \right]^{(m+1)(p-l)} \)

where subscripts denotes the components of \( (x, y, z^*) \) the likelihood ratio test may be shown to reduce to rejecting \( H \), for \( \lambda^* < \text{constant} \).

The distribution of \( \lambda^* \) is not known, although it is now shown that using \( \lambda^* \) we can find an asymptotic test whose associated distribution theory can be determined.

Using (3.3.37), (3.3.38), (3.3.39), and Cramér's Theorem as discussed in Section 3.3.2, we have that as \( m, n \to \infty \),

(3.3.41) \( \ p \lim x = \Gamma \mu_1 = \theta \); \( p \lim y = \Gamma \mu_2 = \varphi \);

(3.3.42) \( p \lim \frac{v}{m} = \alpha_1 \), \( p \lim \frac{w}{n} = \alpha_2 \)

(3.3.43) \( p \lim \frac{r}{m(p-l)} = \beta_1 \), \( p \lim \frac{s}{n(p-l)} = \beta_2 \).

81
Note that the bracketed expressions become exponentials. Assume that as \( m,n \to \infty, \) \( m/n \) remains constant. Then these relations immediately yield the asymptotic test, reject \( H \) if \( T < k = \text{constant} \) where

\[
(3.3.44) \quad T = \frac{(z^*_1 - \varphi_1)^2}{\alpha_2} - \frac{(z^*_1 - \theta_1)^2}{\alpha_1} + \frac{1}{\beta_2} \sum_{j=2}^{p} (z^*_j - \varphi_j)^2 - \frac{1}{\beta_1} \sum_{j=2}^{p} (z^*_j - \theta_j)^2.
\]

To apply this test the unknown parameters \( (\alpha_j, \beta_j, \theta, \varphi) \) are replaced by their maximum likelihood estimates. We note that when the \( T \) statistic defined in equation (3.2.1) is simplified by letting \( z^* = \Gamma z, \ \theta = \Gamma \mu_1, \ \varphi = \Gamma \mu_2, \ \Sigma_j = \Gamma D_j \Gamma, \) for \( j = 1,2 \) and \( D_j \) defined in (3.3.27), it reduces to the statistic in (3.3.44). But the distribution of \( T \) under the assumption of intraclass structure was derived in Section 3.3.3.

### 3.4 Admissible Linear Procedures With Intraclass Structure.

In this section the method of classifying proposed by Anderson and Bahadur [3] is specialized to covariance matrices with intraclass structure. All parameters are assumed known.

#### 3.4.1 Introduction.

Let \( b \) be a \( p \)-variate vector and \( c \) a scalar. Then an observation \( z \) will be classified into \( N(\mu_1, \Sigma_1) \) if \( b'z \leq c \) and into \( N(\mu_2, \Sigma_2) \) if \( b'z > c \). It has been shown (see [3], p. 426) that the procedure defined by

\[
b = (t_1 \Sigma_1 + t_2 \Sigma_2)^{-1}(\mu_2 - \mu_1),
\]

\[
c = b'\mu_1 + t_1 b'\Sigma_1 b = b'\mu_2 - t_2 b'\Sigma_2 b \quad \text{for any } t_1, t_2 \text{ such that}
\]

\[
(t_1 \Sigma_1 + t_2 \Sigma_2) \text{ is positive definite is admissible. A procedure}
\]

82
\( \varphi_0 \) is admissible if among all linear procedures, \( \varphi \), there is none better than \( \varphi_0 \) in the sense that each probability of misclassification for \( \varphi_0 \) is no larger than that for any other \( \varphi \), and for at least one of the two errors, the error for \( \varphi_0 \) is strictly less.

3.4.2 Procedures With Intraclass Structure.

Assume \( \Sigma_1, \Sigma_2 \) both have intraclass structure. Then, if \( \Gamma \) is orthogonal with first row of the form \( \frac{e'_p}{\sqrt{\rho}} \), \( \Sigma_j = \Gamma D_j \Gamma' \), \( j = 1, 2 \), where \( D_j = \text{diag}(\alpha_j, \beta_j, \ldots, \beta_j) \), and \( \alpha_j \) and \( \beta_j \) are defined in (1.2.1), (1.2.2). Then the vector \( b \) becomes

\[
b = (t_1 \Gamma D_1 \Gamma' + t_2 \Gamma D_2 \Gamma')^{-1}(\mu_2 - \mu_1).
\]

Equivalently, define the intraclass structured, positive definite matrix

\[
(3.4.1) \quad F = \Gamma D \Gamma', \quad D = \text{diag} \left( \frac{1}{t_1 \alpha_1 + t_2 \alpha_2}, \frac{1}{t_1 \beta_1 + t_2 \beta_2}, \ldots, \frac{1}{t_1 \beta_1 + t_2 \beta_2} \right).
\]

Then, the admissible procedures will use the vector

\[
(3.4.2) \quad b = F(\mu_2 - \mu_1),
\]

for some \( t_1, t_2 \) satisfying optimality conditions.

It has been shown [5] that when a minimax procedure is used (with respect to a 0-1 loss function) the problem is to find \( t (= t_1 = 1-t_2) \) so that

\[83\]
(3.4.3) \[ b'\left[t^2\Sigma_1 - (1-t)^2\Sigma_2\right]b = 0, \]

with \( b \) of the form given in (3.4.2), and \( 0 < t < 1 \). For general covariance matrices, this problem is solvable only by trial and error. However, in this situation a more explicit solution can be found, and it is now described.

(3.4.2) and (3.4.3) are equivalent to

\[
(\mu_2 - \mu_1)'F'[t^2\Sigma_1 - (1-t)^2\Sigma_2]F(\mu_2 - \mu_1) = 0. \quad \text{Since } \Sigma_j = \Gamma D_j \Gamma', \quad j = 1, 2,
\]

and \( F = \Gamma D \Gamma' \), we find \( (\mu_2 - \mu_1)'\Gamma D \Gamma'[t^2\Gamma D_1 \Gamma' - (1-t)^2\Gamma D_2 \Gamma'](\mu_2 - \mu_1) = 0 \), or

\[
(\mu_2 - \mu_1)'\Gamma[t^2D_1D - (1-t)^2D_2D]\Gamma'(\mu_2 - \mu_1) = 0.
\]

Define

(3.4.4) \[ \varphi = \Gamma'(\mu_2 - \mu_1), \]

where, since \( \mu_1, \mu_2, \Gamma \) are all known, \( \varphi \) is a known vector. Then the condition becomes

(3.4.5) \[ \varphi'[t^2D_1D^2 - (1-t)^2D_2D^2]\varphi = 0. \]

In component form, (3.4.5) becomes
\[ (3.4.6) \quad t^2 \left\{ \frac{\phi_1^2 \alpha_1}{[t \alpha_1 + (1-t) \alpha_2]^2} + \frac{\beta_1 \phi_1^2}{[t \beta_1 + (1-t) \beta_2]^2} \right\} \]

\[ = (1-t)^2 \left\{ \frac{\phi_1^2 \alpha_2}{[t \alpha_1 + (1-t) \alpha_2]^2} + \frac{\beta_2 \phi_1^2}{[t \beta_1 + (1-t) \beta_2]^2} \right\}. \]

Algebraic manipulation gives the quartic equation

\[ (3.4.7) \quad \sum_{i=0}^{4} a_i t^i = 0, \]

where:

\[ a_0 = \alpha_2 \phi_1^2 \beta_2^2 + \beta_2 \phi_2^2 \sum_j \phi_j^2, \]

\[ a_1 = 2 \beta_2 \alpha_1 \phi_1^2 + 2 \alpha_2 \phi_1 \phi_2 \sum_j \phi_j^2 - 4 \alpha_2 \phi_1^2 \beta_2 - 4 \alpha_2 \phi_2 \sum_j \phi_j^2, \]

\[ a_2 = \phi_1^2 \alpha_2 \beta_1 + 6 \phi_2 \phi_1 \phi_2 - 6 \beta_2 \alpha_1 \phi_2^2 + \alpha_2 \phi_2 \sum_j \phi_j^2 \]

\[ + 6 \beta_2 \phi_2 \sum_j \phi_j^2 - 6 \alpha_1 \phi_2 \sum_j \phi_j^2 - \beta_2 \phi_1 \alpha_1 - \beta_1 \alpha_2 \sum_j \phi_j^2, \]

\[ a_3 = 6 \beta_1 \phi_2 \phi_1 \phi_1 - 2 \phi_1 \phi_1 \phi_2 - 4 \phi_1 \alpha_1 \beta_2 + 2 \alpha_2 \phi_2 \sum_j \phi_j^2 \]

\[ - 4 \phi_2 \alpha_2 \sum_j \phi_j^2 + 6 \alpha_1 \beta_2 \sum_j \phi_j^2 + 2 \alpha_1 \phi_1 \beta_2 \]

\[ - 2 \beta_1 \phi_1 \phi_1 - 2 \phi_1 \phi_1 \phi_2 - 2 \alpha_1 \beta_1 \sum_j \phi_j^2, \]

85
\[ a_4 = \phi_{12}^2 \alpha_{12}^2 + \phi_{12}^2 \alpha_{12}^2 + 2 \beta_{12}^2 \sum_{j} \frac{p_j}{2} - \phi_{12}^2 \alpha_{12}^2 + \alpha_{12}^2 \sum_{j} \phi_{j}^2 \]

\[-2 \phi_{12}^2 \alpha_{12}^2 - 2 \phi_{12}^2 \alpha_{12}^2 \sum_{j} \frac{p_j}{2} - \phi_{12}^2 \alpha_{12}^2 \]

\[+ 2 \beta_{12}^2 \alpha_{12}^2 \phi_{12}^2 - \beta_{12}^2 \sum_{j} \frac{p_j}{2} - \beta_{12}^2 \sum_{j} \frac{p_j}{2} + 2 \phi_{12}^2 \alpha_{12}^2 \sum_{j} \phi_{j}^2 . \]

To simplify (assuming \( a_4 \neq 0 \)), let

\[(3.4.8) \quad a_3 = \frac{a_3}{a_4}, \quad a_2 = \frac{a_2}{a_4}, \quad a_1 = \frac{a_1}{a_4}, \quad a_0 = \frac{a_0}{a_4} . \]

Then we obtain

\[(3.4.9) \quad t^4 + a_3 t^3 + a_2 t^2 + a_1 t + a_0 = 0 . \]

Of course for \( a_4 = 0 \) we obtain a cubic equation. Now let \( v = t + \frac{a_2}{4} \) in \( (3.4.9) \). Then,

\[(3.4.10) \quad v^4 + qv^3 + rv + s = 0 , \]

where

\[ q = a_2^2 - \frac{3}{8} (a_3^2) \]

\[ r = a_1^2 - a_2^2 a_3^2 + \frac{1}{8} (a_3^2) \]

\[ s = a_0^2 - a_1^2 a_3^2 + \frac{a_2^2 a_3^2}{16} - \frac{3}{256} (a_3^4) \].
For every \( \mu \), (3.4.10) is equivalent to

\[(3.4.11) \quad (v^2 + \frac{\mu}{2})^2 - [(\mu-q)v^2 - rv + (\frac{\mu^2}{4} - s)] = 0.\]

Define \( Q^2 = (\mu-q)v^2 - rv + (\frac{\mu^2}{4} - s) \) for those \( \mu \) such that the discriminant vanishes; i.e.,

\[(3.4.12) \quad 4(\mu-q)(\frac{\mu^2}{4} - s) = r^2.\]

Since \( q, r, s \) are all real it is well known that there will always be at least one real root \( \mu_1 \geq q \) which satisfies (3.4.12). Hence, the first step is to solve (3.4.12) which is a cubic in \( \mu \). This can always be done exactly in terms of radicals. Then if \( \mu_1 \) is substituted into (3.4.11) we obtain

\[(3.4.13) \quad (v^2 + \frac{\mu_1}{2} + Q)(v^2 + \frac{\mu_1}{2} - Q) = 0,\]

where \( Q = Av - B, \ A = \sqrt{\mu_1 - q}, \ B = \frac{r}{2A} \). Hence the roots of (3.4.10) are the roots of the two quadratic factors in (3.4.13). The solution to our problem is the root \( v = v_0 \) such that \( \frac{a_2}{4} < v_0 < 1 + \frac{a_2}{4}; \ i.e., \ 0 < t < 1. \) With this value of \( t \), the admissible linear procedure: classify into \( N(\mu_2, \Sigma_2) \) if \( b^t_2 > c \) is well defined since then \( c \) is determined from \( x(b^t_2) = N(b_1^t \mu_2, b_1^t \Sigma_2 b_1) \).

Now suppose the classification criterion which is most reasonable is not the one of minimax but is to minimize the probability of misclassification when sampling from \( N(\mu_1, \Sigma_1) \), given the probability of misclassification when sampling from \( N(\mu_2, \Sigma_2) \).
Then if the given probability of misclassification, \( \epsilon \), is less than 50\%, it can be shown (see [3]) that the problem is to find \( t_2 = 1 - t_1 \) so that
\[
\epsilon = t_2 (b' \Sigma_2 b)^{1/2}, \quad \text{with} \quad b = [t_1 \Sigma_1 + t_2 \Sigma_2]^{-1} (\mu_2 - \mu_1).
\]
This problem is solved in identical fashion to the one associated with the minimax criterion and again leads to a quartic equation, whose solution can be found explicitly as above.

3.5 Classification with Unrestricted Covariance Matrices.

In this section the restriction, inherent in previous sections, that \( \Sigma_1 > \Sigma_2 \) is lifted. It is found that the percentage points required to determine the classification boundaries can be expressed in terms of tail probabilities of distributions of indefinite quadratic forms. The requisite distribution theory is developed in Chapter 4.

In Section 3.2.3 it was assumed that all population parameters are known and that the test statistic is \( T \) or \( T_0 \) defined in equations (3.2.1), (3.2.19), respectively. The distribution of \( T_0 \) was then found under the assumption that \( \Sigma_1 > \Sigma_2 \). Without this restriction, the problem is to find the distribution of
\[
T_0 = v' Av, \quad A = \Sigma_2^{-1} - \Sigma_1^{-1}, \quad \text{where} \quad \mathcal{L}(v) = N(\eta_j, \Sigma_j),
\]
\[
\eta_j = \mu_j + A^{-1} (\Sigma_1^{-1} \mu_1 - \Sigma_2^{-1} \mu_2), \quad j = 1 \text{ under } H \text{ and } j = 2 \text{ under } A.
\]
Thus, the only structure known about \( A \) is that \( A = A' \).

Since \( \Sigma_j > 0 \), it is well known that there exists a nonsingular matrix \( P \) so that
\[
(3.5.1) \quad P' \Sigma P = I.
\]

88
Specifically, $P = Q\Delta^{-1/2}$, where $G'\Sigma G = \Delta$, $\Delta = \text{diag} (\delta_1, \ldots, \delta_p)$, $GG' = I$. Define

$$w = P'v; \quad B = (P^{-1})A(P^{-1})'; \quad \omega_j = P'\eta_j.$$

Then,

$$T_0 = w'Bw, \quad \mathcal{L}(w) = N(\omega_j, I).$$

Next note that $B = B'$. So there is a $\Gamma$ such that $\Gamma'\Gamma = I$,

$$B = \Gamma'D\Gamma,$$

where $D = \text{diag} (\lambda_1, \ldots, \lambda_p)$. Define

$$x = \Gamma w, \quad \delta_j = \Gamma\omega_j.$$

Then the statistic becomes

$$T_0 = x'Dx, \quad \mathcal{L}(x) = N(\delta_j, I).$$

Let

$$x' = (x_1, \ldots, x_p), \quad \delta_j' = (\delta_j^{(1)}, \ldots, \delta_j^{(p)}).$$

Then

$$T_0 = \sum_{i=1}^{p} \lambda_i x_i^2, \quad \mathcal{L}(x_i^2) = x_i^2[(\delta_j^{(i)})^2],$$

and the $x_i^2$ are mutually independent. Assume without loss of
generality that the $\lambda_i$ are labeled so that the first $r+1$ are positive and the remaining $s+1 = p-r-1$ are negative (assume that $|B| \neq 0$).

Then $T_0$ can be written in the form

\begin{equation}
(3.5.9) \quad T_0 = U - V,
\end{equation}

where

\begin{equation}
(3.5.10) \quad U = \alpha \left[ \chi^2_{1,d_0} + \sum_{i=1}^{r} a_i \chi^2_{1,d_i} \right],
\end{equation}

\begin{equation}
(3.5.11) \quad V = \beta \left[ \chi^2_{1,g_0} + \sum_{i=1}^{s} b_i \chi^2_{1,g_i} \right],
\end{equation}

\[
\alpha = \min(\lambda_j), \quad 1 \leq j \leq r+1
\]

\[
\beta = \min(|\lambda_j|), \quad \alpha > 0, \beta > 0; \quad r+2 \leq j \leq p
\]

\[
\alpha a_i = \lambda_i, \quad 1 \leq i \leq r; \quad \beta b_i = \lambda_i, \quad r+2 \leq i \leq p
\]

\[
d_i = \left( \delta^{(1)}_j \right)^2, \quad 1 \leq i \leq r; \quad g_i = \left( \delta^{(i)}_j \right)^2, \quad r+2 \leq i \leq p.
\]

Note that $U,V$ are independent positive definite quadratic forms in non-central variates. The problem now is to find the percentage points of the $T_0$ distribution. Approximations to these tail probabilities useful for numerical computations are developed in Chapter 4 and are given in Theorem 4.6.4.
CHAPTER 4

LINEAR COMBINATIONS OF NON-CENTRAL DISTRIBUTIONS

4.1 Introduction.

Let \( F_0(x), \ F_1(x), \ldots \) be any sequence of distribution functions. Let \( \alpha_0, \alpha_1, \ldots \) be any sequence of non-negative constants such that \( \sum_j \alpha_j = 1 \). Then \( F(x) = \sum c_j F_j(x) \) is called a "mixture" of distribution functions. Unless otherwise noted, all summations should be taken from 0 to \( \infty \). Situations in which mixtures must be evaluated to find solutions for real problems were discussed in preceding chapters at which points the results developed below were invoked. The mixture problems considered in this chapter extend some of the work of Robbins and Pitman [25] to the non-central case.

Moreover, the distribution theory for the general case in which the \( c_j \) can be of either sign and the variates are non-central, is also developed.

Let \( X = a_0(x_0^2 + a_1 x_1^2 + \ldots + a_r x_r^2) \), where the chi-squared variates are independent and \( a_0, a_1, \ldots, a_r \) are positive constants with \( a_i > 1, \ i = 1, \ldots, r \). Then, Robbins and Pitman found an expression for the c.d.f. of \( X \) as a series expansion in c.d.f.'s of central chi-squared variates. They also carried out a similar calculation for the c.d.f. of the ratio of such weighted sums of independent central chi-squared variates. In this chapter, both of these results will be extended to non-central chi-squared variates.

Shah and Khatri [28] have considered the non-central case giving a power series expansion for \( F(x) \) in powers of \( x \). This
representation is useful for small values of \( x \), although not too practical computationally for large \( x \).

Imhof [19] finds the characteristic function for \( X \), as defined above, and shows how to invert it by evaluating the inversion integral numerically, using the trapezoidal and Simpson's rules. Although this is a brute force procedure giving little insight into the behavior of the distribution, it is possible that in computational work, the results can be obtained quickly and with tolerable precision.

After the work on positive definite forms was carried out it was discovered that Ruben [26] found one of the results earlier, but that his proof, and form for the resulting coefficients bore no point of contact with the current work.

The results obtained on positive definite forms are extended to indefinite forms in non-central variates in Section 4.6.

4.2 Distribution of a Positive Sum of Non-Central Chi-Squared Variates.

Let \( \chi^2_{m,d} \) denote a non-central chi-squared variate with \( m \) degrees of freedom and non-centrality parameter \( d \). Define

\[
(4.2.1) \quad X = a_0 \chi^2_{m_0,d_0} + a_1 \chi^2_{m_1,d_1} + \ldots + a_r \chi^2_{m_r,d_r}
\]

where all chi-squared variates are independent and \( a > 0, \ a_i \geq 1 \) for \( i = 1, 2, \ldots, r \). Let \( F(x) = \Pr[X \leq x] \), and let \( F_j(s) \) denote the c.d.f. of a central chi-squared variate with \( j \) degrees of freedom. Then, \( F(x) \) is a convex combination of the \( F_j(x) \). Specifically,
Theorem 4.2.1:

\[ F(x) = \sum_{i=0}^{\infty} q_i F_{M+2i}(x/a), \]

where \( M = \sum_{i=0}^{r} m_i \), \( q_i > 0 \), \( \sum_{i=0}^{\infty} q_i = 1 \), and the \( q_i \) are given by

\[ q_i = \sum_{\alpha=0}^{i} \frac{e^{-d_0/2} (\frac{d_0}{2})^{i-\alpha}}{(i-\alpha)!} r^{\alpha}, \]

\[ q_0 = e^{-\frac{1}{2} \sum_{i=0}^{r} d_i^2} \prod_{i=1}^{r} a_i^{\frac{m_i}{2}}, \]

where the coefficients \( r^{\alpha} \) are bounded by unity and given by

\[ K_{\alpha} = h_{\alpha}^{(1)}, \quad 2K_{\alpha} = \sum_{i_1=0}^{\alpha} h_{\alpha-i_1}^{(2)} h_{i_1}^{(1)}, \]

\[ 3K_{\alpha} = \sum_{i_1=0}^{\alpha} \sum_{i_2=0}^{i_1} h_{\alpha-i_1}^{(3)} h_{i_1-i_2}^{(2)} h_{i_2}^{(1)}, \text{ etc.,} \]

and the \( h_{\alpha}^{(i)} \) are defined by

\[ h_{\alpha}^{(i)} = \sum_{\beta=0}^{\alpha} \sum_{k=0}^{\beta} \left[ -\frac{d_1^2}{2} \frac{d_1^2}{2} \beta-k \right] \frac{e^{-\frac{1}{2} (\frac{d_1}{2})^{\beta-k}}}{(\beta-k)!} c_{\alpha-\beta}^{(i)} g_{\beta}^{(i,k)} \]

the constants \( c_{\alpha}^{(i)}, g_{\beta}^{(i,k)} \) are defined by the identities
\( (4.2.6) \ \sum_{\alpha=0}^{\infty} c_{\alpha}^{(i)} w^\alpha = a_i^2 \left[ 1 - (1 - \frac{1}{a_i}) w \right]^{\frac{m_i}{2}}, \quad |w| < 1, \)

\( (4.2.7) \ \sum_{\beta=0}^{\infty} g_{\beta}^{(i,k)} w^\beta = a_i^{-k} \left[ 1 - (1 - \frac{1}{a_i}) w \right]^{-k}, \quad |w| < 1. \)

i.e., \( c_{\alpha}^{(i)} = a_i^{\frac{m_i}{2}} \left( 1 - \frac{1}{a_i} \right)^{\alpha} \left( \frac{m_i}{2} \right), \quad g_{\beta}^{(i,k)} = a_i^{-k} \left( 1 - \frac{1}{a_i} \right)^{\beta} \left( \frac{\beta + k - 1}{\beta} \right)(-1)^{\beta}. \)

**Proof:** Let \( \varphi_Y(t) \) denote the characteristic function of \( Y. \)

Let \( \psi_n(t) \) denote the characteristic function of a \( \chi^2_n \) variate; i.e.,

\[ \psi_n(t) = (1 - 2it)^{-\frac{n}{2}}. \]

Then, by independence of variates in \( (4.2.1), \)

\( (4.2.8) \ \varphi_{\chi}(t) = \varphi_{\chi^2}(t)^{\sum_{i=1}^{\infty} \varphi_{\chi^2_{m_i,d_i}}(t)}. \)

Let \( f(U)(x) \) denote the density of \( U, \) where \( z(U) = z(\chi^2_{m_0,d_0}). \)

Then if \( f_j(x) \) denotes the density of a central chi-squared variate with \( j \) degrees of freedom, it is well known that

\( (4.2.9) \ f(U)(x) = \sum_{j=0}^{\infty} e^{-\frac{d_0^2}{2}} \frac{(d_0^2)^j}{j!} f_{m_0+2j}(x) \equiv \sum_{j=0}^{\infty} p_j(d_0) f_{m_0+2j}(x). \)

The characteristic functions are analogously related by

\( (4.2.10) \ \varphi_U(t) = \sum_{j} p_j(d_0) \psi_{m_0+2j}(t). \)
The verification of (4.2.10) is given in [25]. Next, recall that

\[(4.2.11) \quad \varphi_{aU}(t) = E \ e^{i \text{ats}U} = \sum_{j} p_j(d_0) \psi_{m_0 + 2j}^{(at)} = \sum_{j} p_j(d_0) \varphi_{ax^{m_0 + 2j}}(t).\]

Since the characteristic function of a constant times a central chi-squared variate is expressible as (see [25], eqn. 16)

\[(4.2.12) \quad \varphi_{ax^2 n}(t) = \psi_{n(at)} = (1 - 2iat)^{-n/2} = \sum_{k} c_k (1 - 2it)^{-n/2 - k},\]

where

\[(4.2.13) \quad c_k = a^{-n/2} \left(1 - \frac{1}{a} \right)^k \binom{n}{k},\]

we find correspondingly for \(m_0 + 2j\) degrees of freedom as in (4.2.11),

\[(4.2.14) \quad \varphi_{ax^2 m_0 + 2j}(t) = \psi_{m_0 + 2j}^{(at)} = \sum_{k} \tilde{c}_k \psi_{m_0 + 2j + 2k}(t),\]

where the \(\tilde{c}_k\) are defined by

\[(4.2.15) \quad \sum_{k} \tilde{c}_k z^k = a^{-m_0 + 2j/2} \left[1 - \left(1 - \frac{1}{a}\right) z \right]^{-\frac{m_0 + 2j}{2}}.\]

Note in (4.2.8), the characteristic function we desire is expressible as the product of functions each of which takes the form (4.2.11), and each of the functions in the product is in turn expressible as in (4.2.14). These results are now combined. First note that
\( (4.2.16) \quad \psi_2(t) = (1 - 2it)^{-1}. \)

\[
\therefore \varphi_U(t) = \sum_j \frac{m_0^{+2j} \psi_2^2}{\sum_k c_k^{(j)}} \psi_2(t), \quad \text{and}
\]
\[
\varphi_{aU}(t) = \sum_j \frac{m_0^{+2j} \varphi_{aU}^2}{\sum_k c_k^{(j)}} \psi_2(t) \quad \text{Now from (4.2.15) we find}
\]
\[
(4.2.17) \quad \varphi_{aU}(t) = \sum_j \frac{m_0^{+2j} \varphi_{aU}^2}{\sum_k c_k^{(j)}} \psi_2(t) = \frac{m_0^{+2j}}{2} \left[ 1 - \left(1 - \frac{1}{a} \right) \varphi_{aU}(t) \right].
\]

Hence, the function \( \varphi_{a}(t) \) of (4.2.6) may be written as

\[
(4.2.18) \quad \varphi_{a}(t) = \sum_j \frac{m_0^{+2j} \psi_2^2}{\sum_k c_k^{(j)}} \psi_2(t) \quad \left\{ \sum_i \frac{m_i^{+2k} \psi_2^2}{\sum_i c_i^{(j)}} \right\},
\]

\[
= \left[ 1 - \left(1 - \frac{1}{a} \right) \varphi_{a}(t) \right] \frac{m_0^{+2j}}{2} \psi_2(t) \psi_2(t).
\]

Note that

\[
(4.2.19) \quad \prod_{i=1}^{r} \frac{m_i}{\psi_2(t)} = \left[ \psi_2(t) \right] \frac{1}{\sum_{i=1}^{r} m_i}.
\]

Hence, by regrouping the terms, using the definitions of \( c_{a}^{(i)} \), \( g_{p}^{(i,k)} \) given in (4.2.6), (4.2.7), and recalling that \( M = m_0 + \sum_{i=1}^{r} m_i \), we can obtain the compact form.
\[(4.20) \quad q_{x}(t) = \sum_{j} p_{j}(a_{0}) \sum_{i=1}^{M+2j} \frac{t^{i}}{i!} Q_{i}(t) \]

where

\[(4.21) \quad Q_{i}(t) = \sum_{\alpha} \sum_{k} c_{\alpha}(i,k) p_{k}(d_{1}) g_{\alpha}^{(i,k)} \psi_{2}^{\alpha+k}(t). \]

The simplification of (4.20) into a manageable form is carried out in two stages. First, the triply infinite series (4.21) is reduced, using Cauchy products, to a singly infinite series; and second, \[\prod_{i=1}^{\infty} Q_{i} \] is reduced to a singly infinite series. The two series are then combined for the final result.

Since \( p_{k}(d_{1}) \) are just Poisson probabilities, \( 0 \leq p_{k}(d_{1}) \leq 1 \).

Moreover, from (4.27) it is clear that since \( g_{\alpha}^{(i,k)} \geq 0 \) for all \( (i,k) \), and \( w = 1 \) implies \( \sum_{\beta} g_{\alpha}^{(i,k)} = 1 \), we must have \( g_{\alpha}^{(i,k)} \leq 1 \).

Thus, \[ \sum_{k} p_{k}(d_{1}) g_{\alpha}^{(i,k)} \psi_{2}^{k+\beta}(t) \leq \sum_{k} \psi_{2}^{k+\beta}(t) = [1 - \psi_{2}(t)]^{-2}, \]
this double series converges absolutely; hence, rearrangement of terms is permissible. Define

\[(4.22) \quad \gamma_{\alpha}^{(i)} = \frac{\beta}{\sum_{k=0}^{\beta} p_{\beta-k}(d_{1}) g_{k}^{(i,\beta-k)}}. \]

Then the Cauchy product becomes

\[ \sum_{k} \sum_{\beta} p_{k}(d_{1}) g_{\alpha}^{(i,k)} \psi_{2}^{k+\beta}(t) = \sum_{\beta} \gamma_{\alpha}^{(i)} \psi_{2}(t), \]

97
and $Q_1(t)$ reduces to

$$(4.2.23) \quad Q_1(t) = \left[ \sum_{\alpha} c_{\alpha}^{(i)} \psi_{2}^{\alpha}(t) \right] \left[ \sum_{\beta} \gamma_{\beta}^{(i)} \psi_{2}^{\beta}(t) \right].$$

By the identical argument used above for $\varepsilon_{\beta}^{(1,k)}$, we conclude that $c_{\alpha}^{i} \leq 1$. Moreover, since $\sum_{\beta} \gamma_{\beta}^{(i)} \psi_{2}^{\beta}(t)$, $\sum_{\alpha} c_{\alpha}^{(i)} \psi_{2}^{\alpha}(t)$ both converge absolutely, we can further reduce $Q_1$ down to a single series. We find

$$(4.2.24) \quad Q_1 = \sum_{\alpha} h_{\alpha}^{(i)} \psi_{2}^{\alpha}(t),$$

where the product coefficients are given in the usual way by

$$(4.2.25) \quad h_{\alpha}^{(i)} = \sum_{\beta=0}^{\alpha} \gamma_{\beta}^{(i)} c_{\alpha-\beta}^{(i)} = \sum_{\beta=0}^{\alpha} \sum_{k=0}^{\beta} c_{\alpha-\beta}^{(i)} p_{-k} \psi_{2}^{(i,\beta-k)}.$$  

Next examine $\prod_{i=1}^{r} Q_i$. Since $Q_i$ is finite for each $i$, the finite product is finite. Moreover, since $h_{\alpha}^{(i)} \geq 0$ for each $(i,\alpha)$, $Q_i$ converges absolutely for each $i$. Hence, the Cauchy product relations may be applied to the successive terms of the product. Now use the coefficients $K_{\alpha}$ defined in $(4.2.4)$. We see that

$$\sum_{\alpha} h_{\alpha}^{(2)} \psi_{2}^{\alpha}(t) \sum_{\beta} h_{\beta}^{(1)} \psi_{2}^{\beta}(t) = \sum_{\alpha} 2^{K_{\alpha}} \psi_{2}^{\alpha}(t).$$

By direct generalization, it may be found that
\[(4.2.26) \quad \prod_{i=1}^{r} \left[ \sum_{\alpha} h^{(i)}_{\alpha} \psi_{2}(t) \right] = \sum_{\alpha} r^{\alpha} \psi_{2}(t) \].

(Note that \( r^{\alpha} \leq 1 \)). Substitution of (4.2.24), (4.2.26) into (4.2.20) gives

\[(4.2.27) \quad \varphi_{X}(t) = \sum_{i=0}^{M} \sum_{j} p_{j}(d_{0}) r^{\alpha} \psi_{2}^{j+\alpha}(t) \].

Now note that since \( h^{(i)}_{\alpha} \geq 0 \) for each \((i,\alpha)\), \( r^{\alpha} \geq 0 \) for every \( \alpha \).

Hence, the double series in (4.2.27) converges absolutely permitting the final application of the Cauchy product form. Define

\[(4.2.28) \quad q_{i} = \sum_{\alpha=0}^{i} r^{\alpha} p_{i-\alpha}(d_{0}) \].

Then,

\[(4.2.29) \quad \sum_{i} \sum_{\alpha} p_{j}(d_{0}) r^{\alpha} \psi_{2}^{j+\alpha}(t) = \sum_{i} q_{i} \psi_{2}^{i}(t) \].

Note that (4.2.28) is the \( q_{i} \) given in Theorem 4.2.1, equation (4.2.3). Using (4.2.29) in (4.2.27) gives

\[(4.2.30) \quad \varphi_{X}(t) = \sum_{i=1}^{M+2i} q_{i} \psi_{2}^{i}(t) = \sum_{i=1}^{M+2i} q_{i} \psi_{2}^{i}(t) \].

Next, invert the characteristic functions termwise

(the justification is given in [25]). Then,
(4.2.31) \[ P\left\{ \frac{X}{a} \leq y \right\} = \sum_{i} q_{i}F_{M+i}(y), \]

or finally,

(4.2.32) \[ F(x) = \sum_{i} q_{i}F_{M+i}(x/a). \]

An example of the use of Theorem 4.2.1 was given in Chapter 3, Section 3.2.3. It was involved with the problem of multivariate classification.

4.3 Distribution of a Ratio.

We next consider the distribution of a ratio of sums of weighted and independent chi-squared variates, whose numerator variates are non-central and denominator variates are central. Specifically, let

(4.3.1) \[ X = \frac{a\cdot(\chi_{m_0}^2, d_0 + a_1\chi_{m_1}^2, d_1 + \ldots + a_r\chi_{m_r}^2, d_r)}{\chi_{n_0}^2 + b_1\chi_{n_1}^2 + \ldots + b_s\chi_{n_s}^2}, \]

with all chi-squared variates being independent, \((a, a_1, \ldots, a_r, b_1, \ldots, b_s)\) are all positive constants with \(a_i \geq 1, \ b_j \geq 1, \) for \(i = 1, \ldots, r,\)

and \(j = 1, \ldots, s.\) Definite separate numerator and denominator variates as

(4.3.2) \[ U = a(\chi_{m_0}^2, d_0 + a_1\chi_{m_1}^2, d_1 + \ldots + a_r\chi_{m_r}^2, d_r), \]

(4.3.3) \[ V = \chi_{n_0}^2 + b_1\chi_{n_1}^2 + \ldots + b_s\chi_{n_s}^2. \]
From Theorem (4.2.1) we can assert that

\[(4.3.4) \quad \Pr[U \leq x] = \sum_{i=1}^{\infty} q_i F_{M+2i}(x/a),\]

where \((M, q_i)\) are defined in (4.2.2), (4.2.3).

From [25], Theorem 1, it is clear that

\[(4.3.5) \quad \Pr[V \leq x] = \sum_{k=1}^{\infty} \theta_k F_{N+2k}(x),\]

where \(V\) is defined in (4.3.3), \(N = n_0 + \sum_{i=1}^{s} n_i\), and \(\theta_k\) are defined by

\[(4.3.6) \quad \prod_{i=1}^{s} \frac{n_i}{2} \left[ 1 - \left(1 - \frac{1}{b_i}\right)^z \right]^{-\frac{n_i}{2}} = \sum_k \theta_k z^k,\]

for \(|z| < 1\).

From (4.3.4), (4.3.5), it is seen that the c.d.f.'s of \((U)\) and \((V)\) are both mixtures. Since \(U/V\) is a simple Borel function, its c.d.f. can be found just as the c.d.f. of a single non-central F-distribution is found. From (4.3.4), \(X = U/V\). Hence,

\[(4.3.7) \quad \Pr[X \leq x] = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} q_i \theta_k \Pr[\frac{U_i}{V_k} \leq x],\]

where the \((U_i, V_k)\) are random variables having distribution functions \(F_{M+2i}(x/a), F_{N+2k}(x)\), respectively. But \(U_i/V_k\) is just the ratio of two independent chi-squared variates. In this connection, let \(F_{m,n}(x)\)
denote the c.d.f. of the ratio of two independent central chi-squared variates \( \chi_m^2 / \chi_n^2 \). Then,

\[
(4.3.8) \quad F_{m,n}(x) = \frac{1}{\text{B}(m/n, n/2)} \int_0^x t^{m/2 - 1} (1+t)^{-(m+n)/2} \, dt , \quad x \geq 0 ,
\]

and \( F_{m,n}(x) \) is zero otherwise. Note that \( F_{m,n}(x) \) is related to the c.d.f. of a central F-variate. However, in computational work, \( F_{m,n}(x) \) is more convenient since \( F_{m,n}(x) = I_x/(1+x)(m/n, n/2) \) where \( I_x(m,n) \) is the well tabulated incomplete beta function.

Applying (4.3.8) to the ratio in (4.3.) gives

\[
(4.3.9) \quad P(U_k \leq x) = F_{M+2i,N+2k}(x/a) .
\]

We can now determine the distribution function of \( X \) in (4.3.1), which was originally sought.

**Theorem 4.3.1:** Let \( X \) be defined, as in (4.3.1), as the ratio of sums of non-central chi-squares to central chi-squares. Then if \( F_{m,n}(x) \) is defined as in (4.3.8) and \((q_i, \theta_k)\) are defined as in (4.3.2), (4.3.6),

\[
(4.3.10) \quad \Pr(X \leq x) = \sum_i \sum_k q_i \theta_k F_{M+2i,N+2k}(x/a) .
\]

**4.4 Probabilistic Interpretation.**

It is often illuminating and fruitful to examine results of the type expressed in Theorems 4.2.1, 4.3.1 in the light of their
associated probabilistic significance.

Let \( X_i, \ i = 0,1, \ldots \) be a sequence of discrete, integer valued random variables with probabilities \( \Pr(X_i=\alpha) = c_\alpha^{(i)} \), where \( c_\alpha^{(i)} \) are defined in (4.2.7). Similarly define \( Y_i, Z_i \) as integer valued random variables whose joint probability mass function is given by \( \Pr(Y_i=k, Z_i=\beta) = p_k(d_i)g_{\alpha}^{(i),k} \). Assume that for each \( i \), \( X_i \) is independent of \( (Y_i, Z_i) \). Define \( V_i = X_i + Y_i + Z_i \). Then, if \( V = \sum_{i=1}^{r} V_i \), \( \Pr(V_i=\alpha) = h_\alpha^{(i)} \), and the \( h_\alpha^{(i)} \) are defined in (4.2.25). Assume all \( V_i \) are independent.

Let \( W \) be a poisson random variable with parameter \( (\alpha_0^2/2) \). Then the probabilities are \( \Pr(W=j) = p_j(d_0) \), where the \( p_j(d_0) \) are defined in (4.2.9). Define \( R = V+W \), where \( V \) and \( W \) are assumed independent. Then, the coefficients \( (q_i) \) which result in Theorem 4.2.1 are just the probabilities associated with \( R \); to wit, \( \Pr(R=i) = q_i \).

As a consequence of this possible interpretation, the equations used in deriving Theorem 4.2.1 can be examined by thinking of series of the form \( \left( \sum_{\alpha} p_{\alpha} x^\alpha \right) \) as moment generating functions in \( (w) \), and using the well established properties of moment generating functions to develop other, associated relations.

4.5 Numerical Computation and Approximations for Positive Definite Forms.

The mixture representations for the weighted sums of non-central chi-squared and F-variates given above must, in practice, be approximated by finite series for computational work. For such
purposes, it is desirable to have bounds on the error of the approximation.

In the case of chi-squared variates it was shown (Theorem 4.2.1) that for \( q_1 \geq 0, \sum_{i=0}^{\infty} q_i = 1, \ F(x) = \sum_{i=0}^{\infty} q_i F_{M+2i}(x/a). \) For numerical computations we use instead the truncated series

\[
F(x;n) = \sum_{i=0}^{n} q_i F_{M+2i}(x/a).
\]

The error is then

\[
(4.5.1) \quad \epsilon(x) = F(x) - F(x;n).
\]

For any \( p_1, p_2, \)

\[
0 \leq F(x) - \sum_{p_1}^{p_2} q_i F_{M+2i}(x/a) = \sum_{p_1}^{p_1-1} q_i F_{M+2i}(x/a) + \sum_{p_2+1}^{\infty} q_i F_{M+2i}(x/a).
\]

However, since it is easily shown that for any \( L \) and \( y, \)

\[
F_L(y) \geq F_{L+2}(y),
\]

\[
0 \leq F(x) - \sum_{p_1}^{p_2} q_i F_{M+2i}(x/a)
\leq F_M(x/a) \sum_{i=p_1}^{p_1-1} q_i + F_{M+2p_2+2}(x/a) \sum_{i=p_2+1}^{\infty} q_i
\leq 1 - \sum_{i=p_1}^{p_2} q_i.
\]

Hence, if \( p_1 = 0, \ p_2 = n, \) we obtain the bounds
\[(4.5.2)\quad 0 \leq \epsilon(x) \leq 1 - \sum_{i=0}^{n} q_i ,\]

and the tighter bounds

\[(4.5.3)\quad 0 \leq \epsilon(x) \leq F_{M+2n+2}(x/a) \left[ 1 - \sum_{i=0}^{n} q_i \right].\]

Thus, in computing \( F(x) \) numerically, the upper bound on the error term can be computed with each successive term added to the series.

For some purposes, bounds will be needed for the coefficients themselves. Since \( q_i = \sum_{\alpha=0}^{i} e^{-d_0^2/2} \frac{d_0^2}{2} \alpha^i \frac{1}{\alpha!} r^\alpha \), and since \( r^\alpha \leq 1 \), it is clear that a crude estimate is

\[(4.5.4)\quad 0 \leq q_i \leq \sum_{\alpha=0}^{i} e^{-d_0^2/2} \frac{d_0^2}{2} \alpha^i \frac{1}{\alpha!} (x_0)^\alpha .\]

Thus, it is seen that the \( q_i \) are bounded by the c.d.f. of a Poisson random variable.

It should be recalled that for many purposes it is sufficient to use an approximation to the c.d.f. of the weighted sum of non-central chi-squared variates. Several approximations were developed in Section 3.2.5. The simplest approximation, when applied to the variate \( X \) of equation (4.2.1), gives

\[(4.5.5)\quad F(x) \approx P(pX^2 \leq x) ,\]

where
\[
\rho = \frac{m_0+2d_0 + \sum_{i=1}^{r} a_i (m_{i1}+2d_{i1})}{m_0+d_0 + \sum_{i=1}^{r} a_i (m_{i1}+d_{i1})}, \quad \nu = \frac{a[m_0+d_0 + \sum_{i=1}^{r} a_i (m_{i1}+d_{i1})]^2}{m_0+2d_0 + \sum_{i=1}^{r} a_i (m_{i1}+2d_{i1})}.
\]

4.6 **Indefinite Quadratic Forms In Non-Central Normal Variates.**

In this section the distribution theory is developed for linear combinations of independent non-central chi-squared variates without regard to sign of the coefficients. The results are applicable to the general classification problem with unequal covariance matrices considered in Chapter 3, in which the property that \(\Sigma_1 > \Sigma_2\) is not satisfied. Gurland [18], in considering the problem for central variates found an infinite series expansion in terms of Laguerre polynomials for the case in which the number of positive (or negative) coefficients is even.

4.6.1 **Exact Distributions.**

The first problem considered is that of the weighted difference of two independent central chi-squared variates with differing degrees of freedom. Define

\[
(4.6.1) \quad R = X - \gamma Y,
\]

where \(\chi^2(X) = \chi^2_m, \ \chi^2(Y) = \chi^2_n, \ \gamma > 0,\) and \(X\) and \(Y\) are assumed to be independent. Pachares [22] considered this problem for \(m = n\) and \(\gamma = 1,\) and found the distribution of \(R\) expressible in terms of modified Bessel functions of the second kind. It will be seen
that the form of the solution developed below will be more suited to percentage point calculations.

Let \( p_0(t) \) denote the density of \( R \). Then, if \( f_x(t), h_y(t) \)
denote the densities of \( X \) and \( Y \), respectively,

\[
(4.6.2) \quad p_0(t) = \int_{-\infty}^{\infty} f_x(t+\gamma x) h_y(x) \, dx.
\]

The problem is now to evaluate the integral in (4.6.2), given that \( X, Y \) both have chi-squared densities. The evaluation is carried out in two stages, depending upon whether \( t \) is positive or negative. First, however, we introduce a function which will be used frequently throughout the remainder.

Define

\[
\psi(a,b;x) = \frac{1}{\Gamma(a)} \int_{0}^{\infty} e^{-xt} t^{a-1} (1+t)^{b-a-1} \, dt,
\]

for \( a > 0, \ x > 0 \). This function satisfies the confluent hyper-

geometric differential equation of Kummer: \( xd^2y/dx^2 + (b-x)dy/dx - ay = 0 \)
(see Erdélyi [14] and is identical with the function \( U(a,b;x) \)
discussed by Slater [30]. The \( \psi \) function can also be expressed in terms of the usual confluent hypergeometric function as

\[
(4.6.3) \quad \psi(a,b;x) = \frac{\Gamma(1-b)}{\Gamma(1+a-b)} {}_1F_1(a,b;x) + \frac{\Gamma(b-1)}{\Gamma(a)} x^{1-b} {}_1F_1(1+a-b,2-b;x).
\]

107
The main advantage of the $\psi$ function over the $\phantom_{\phantom{1}}{1}\phantom{\phantom{1}}{F}\phantom{\phantom{1}}{1}$ function is that $\psi$ always remains finite, regardless of the value assumed by $x$.

Assume $t \geq 0$. From the laws in (4.6.1), (4.6.3) becomes

\[
(4.6.4) \quad p_0(t) = \int_0^{\infty} \frac{n^2 - 1}{x^n} \frac{m^2 - 1}{(t+\gamma x)^{n/2}} \frac{1}{\gamma^{n/2}} e^{-\frac{1}{2}(t+\gamma x)} \ dx.
\]

After letting $\gamma x = ty$, straightforward algebra yields

\[
(4.6.5) \quad p_0(t) = \frac{m+n-2}{t^{n/2}} \frac{m+n}{\gamma^{n/2}} \psi\left(\frac{n}{2}, \frac{m+n}{2}; \frac{1+\gamma}{2\gamma} t\right), \quad t \geq 0.
\]

Now assume $t \leq 0$. Let $t = -\theta$, $\theta > 0$. Then, by symmetry, it is easy to find that

\[
(4.6.6) \quad p_0(-\theta) = \frac{m+n-2}{t^{n/2}} \frac{m+n}{\gamma^{n/2}} \psi\left(\frac{n}{2}, \frac{m+n}{2}; \frac{1+\gamma}{2\gamma} \theta\right).
\]

These results are now combined in a form which is more useful for the sequel.

**Theorem 4.6.1**: Let $Z = \alpha X - \beta Y$, $\alpha > 0$, $\beta > 0$ where $X, Y$ are independent chi-squared variates with $m, n$ degrees of freedom respectively. The probability density function of $Z$ is given by
\[
\begin{align*}
\frac{c}{\Gamma\left(\frac{m}{2}\right)} t^{\frac{m+n-2}{2}} e^{-\frac{t}{2\alpha}} \psi\left(\frac{m}{2}, \frac{m+n}{2}; \frac{\alpha+\beta}{2\alpha\beta} t\right), & \quad t \geq 0 \\
\frac{c}{\Gamma\left(\frac{n}{2}\right)} (-t)^{\frac{m+n-2}{2}} e^{\frac{t}{2\beta}} \psi\left(\frac{n}{2}, \frac{m+n}{2}; -\frac{\alpha+\beta}{2\alpha\beta} t\right), & \quad t \leq 0
\end{align*}
\]

(4.6.7) \( P_z(t) = \)

where \( \psi \) is the finite version of the confluent hypergeometric function defined in (4.6.3), and \( c = \left(\frac{\frac{m+n}{2}}{2 \frac{m}{\alpha^2} \frac{n}{\beta^2}}\right)^{-1} \).

This theorem will be needed in the development of the density function of an indefinite form in non-central variates. First we establish a required modification of Theorem 4.2.1.

In equation (4.2.2) we had \( F(x) = \sum_{i=0}^{\infty} q_i F_{M+2i}(x/a) \) where \( F_{M+2i}(x) \) is the c.d.f. of a central chi-squared variate with \( M+2i \) degrees of freedom, and the \( q_i \) are constants defined in Theorem 4.2.1. Hence, all the \( F \)'s are differentiable with \( F'_k(\cdot) = f_k(\cdot) \). Consider the series \( \sum_{i=0}^{\infty} q_i f_{M+2i}(x) \) = \( \sum_{i=0}^{\infty} q_i \left(\frac{M+2i}{2}\right)^{-1} \frac{x^2}{2} e^{-x/2} \). It is clear that the series converges uniformly on every finite interval of \( x \). Hence, the series converges to some function \( f(x) \) giving

**Theorem 4.6.2:**

Let \( X \) be defined as a linear combination of non-central
chi-squared variates as in (4.2.1), and let \( f(x) \) denote the probability
density function of \( X \). Let \( f_k(x) \) denote the density of a central
chi-squared variate with \( k \) degrees of freedom. There exist constants
\( q_i > 0, \sum_0^\infty q_i = l \), defined in Theorem 4.2.1, so that

\[
(4.6.8) \quad f(x) = \sum_{i=0}^\infty \frac{q_i}{a_i} f_{M+2i}(x/a).
\]

Next we derive the density of an indefinite form in
non-central variates. Define

\[
(4.6.9) \quad U = \alpha \left[ X^2_{n_0,d_0} + \sum_{i=1}^r a_i X^2_{n_i,d_i} \right],
\]

\[
(4.6.10) \quad V = \beta \left[ X^2_{n_0,g_0} + \sum_{i=1}^s b_i X^2_{n_i,g_i} \right],
\]

where \( \alpha > 0, \beta > 0, a_i \geq l, b_i \geq l \), and \( U, V \) are independent.
Also define

\[
(4.6.11) \quad T = U - V,
\]

and, in this derivation only, denote the densities of \( T, U, V \) by
\( h(t), f(t), g(t) \), respectively. From Theorem 4.6.2, we know there
exist constants \( q_i \geq 0, q^*_i \geq 0, \sum_0^\infty q_i = l, \sum_0^\infty q^*_i = l \) so that

\[
(4.6.12) \quad f(t) = \sum_{i=0}^\infty \frac{q_i}{\alpha} f_{M+2i}(t/\alpha), \quad t \geq 0,
\]
\[ g(t) = \sum_{j=0}^{\infty} \frac{q^x_j}{\beta} f_{N+2j}(t/\beta), \quad t \geq 0, \]

and the \( q^x_j \) are defined in Theorem 4.2.1. Moreover, \( f, g, h \) are related through
\[ h(t) = \int_{-\infty}^{\infty} f(t+x)g(x) \, dx. \]

First assume \( t \geq 0 \). Then (4.6.12), (4.6.13) give
\[ h(t) = \int_{0}^{\infty} \sum_{i=0}^{\infty} \frac{q_i}{\alpha} f_{M+2i}(\frac{t+x}{\alpha}) \sum_{j=0}^{\infty} \frac{q^x_j}{\beta} f_{N+2j}(x/\beta) \, dx, \]
and since the series' converge uniformly,
\[ h(t) = \sum_{i} \sum_{j} \frac{q_i q^x_j}{\alpha \beta} \int_{0}^{\infty} f_{M+2i}(\frac{t+x}{\alpha}) f_{N+2j}(x/\beta) \, dx. \]

Let
\[ I_{ij}(t) = \int_{0}^{\infty} f_{M+2i}(\frac{t+x}{\alpha}) f_{N+2j}(x/\beta) \, dx, \]
so that
\[ h(t) = \sum_{i} \sum_{j} \frac{q_i q^x_j}{\alpha \beta} I_{ij}(t). \]

It is seen that with a change of variables,
\[ I_{ij}(at) = \beta \int_0^\infty f_{M+2i}(t+y) f_{N+2j}(y) \, dy, \]

where \( \gamma = \beta / \alpha \). But if \( m = M+2i \), \( n = N+2j \), and \( Z = \alpha X - \beta Y \) with \( x(x) = x_m^2, y(y) = x_n^2, x, y \) independent, \( I_{ij}(t) = \alpha \beta p_z(t) \), with \( p_z(t) \) the density of \( Z \), for \( t \geq 0 \).

Now assume \( t \leq 0 \), and let \( t = -\theta, \theta \geq 0 \). The transformations are completely analogous and are omitted. The results are combined and summarized by using Theorem 4.6.2 for \( p_z(t) \), and we obtain

**Theorem 4.6.3:** Let

\[ T = \alpha \left[ \chi^2_{m_0, 0} c_0 + \sum_{i=1}^r a_i \chi^2_{\hat{m}_i, \hat{d}_i} \right] - \beta \left[ \chi^2_{n_0, 0} b_0 + \sum_{i=1}^s b_i \chi^2_{\hat{n}_i, \hat{g}_i} \right], \]

where \( \alpha > 0, \beta > 0, a_i \geq 1, b_i \geq 1 \), all chi-squared variates are independent, and let \( h(t) \) denote the probability density function of \( T \). Then

\[ h(t) = \sum_{i=0}^\infty \sum_{j=0}^\infty q_i^* q_j^* p_z(t), \]

where \( p_z(t) \) is defined in equation (4.6.7) with the identifications \( m = M+2i \), \( n = N+2j \), and \( q_i, q_j^* \) are defined in Theorem 4.2.1 for \( U, V \), respectively, as defined in (4.6.9), (4.6.10).

Now consider a general, indefinite quadratic form

\[ T = x'Ax \]

112
where \( \mathbf{z}(\mathbf{x}) = N(\mu, I) \), \( \mathbf{x} \) is a \( p \)-variate column vector, and \( A \) is an arbitrary symmetric matrix. Let \( A = \Gamma' \Lambda \Gamma \), where \( \Gamma' = I \), and \( D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \). Then (4.6.17) is equivalent to

\[
(4.6.18) \quad T = w'Dw,
\]

where \( \mathbf{z}(w) = N(\Gamma\mu, I) \); that is,

\[
(4.6.19) \quad T = \sum_{1}^{p} \lambda_i w_i^2,
\]

where the \( \lambda_i \) are real, and the \( w_i^2 \) are non-central chi-squared variates. Assume without loss of generality that the latent roots are labeled so that \( \lambda_1 > \lambda_2 > \ldots > \lambda_{r+1} > 0 \), and \( \lambda_{r+2} < \lambda_{r+3} < \ldots < \lambda_p < 0 \). By identifying \( \lambda_{r+1} \) with \( \alpha \), and \( \lambda_p \) with \( \beta \), it is clear that \( T \) of (4.6.17) can be put into the form used in Theorem 4.6.3.

The density expression given in equation (4.6.16), although exact, is too complicated for numerical computations of the percentage points of the distribution. However, it is not really necessary, for most purposes, to evaluate the c.d.f. of the distribution for all values of the argument, but only for those values which yield the tail probabilities and since asymptotic expansions are available, this is possible. In this spirit, we now obtain approximations to the tail probabilities which can be used for numerical work such as in the classification problems discussed in Chapter 3.
4.6.2 Approximate Percentage Points.

Define $T$ as in Theorem 4.6.3 and let $H(t)$ denote its c.d.f. The density, $h(t)$, is defined in (4.6.16). Then

\[ H(t) = \int_{-\infty}^{t} h(x) \, dx = \int_{-\infty}^{t} \sum_{i} \sum_{j} q_i^* q_j^* p_z(x) \, dx. \]

But the series converges uniformly so that

\[ H(t) = \sum_{i} \sum_{j} q_i^* q_j^* \int_{-\infty}^{t} p_z(x) \, dx. \]

Assume $t \leq 0$.

Substitution of $p_z(t)$ from (4.6.7) gives

\[ H(t) = \sum_{i} \sum_{j} \frac{q_i^* q_j^* c_{i,j}}{1 / 2 \pi} J_{i,j}(t), \quad t \leq 0 \]

where

\[ J_{i,j}(t) = \int_{-\infty}^{t} (-x)^{m+n-2} e^{2\beta x} \psi\left(\frac{m}{2}, \frac{m+n}{2}; -\frac{\alpha+\beta}{2\alpha\beta} x\right) \, dx \]

for $t \leq 0$, and $m = M+2i$, $n = N+2j$. Letting $-\frac{\alpha+\beta}{2\alpha\beta} x = y$, $t = -\theta$, $\theta \geq 0$,

\[ J_{i,j}(-\theta) = \int_{(\frac{\alpha+\beta}{2\alpha\beta})^\theta}^{\infty} \frac{m+n}{2} e^{2\beta y} \psi\left(\frac{m}{2}, \frac{m+n}{2}; y\right) \, dy. \]

Since in this integral, $y \geq (\frac{\alpha+\beta}{2\alpha\beta})^\theta$, and for large $y \psi(a,b; y) \approx y^{-a}$ (see Slater [30]), asymptotically,
(4.6.22) \quad J_{ij}(-\theta) \approx \left(\frac{\alpha \beta}{\alpha + \beta}\right)^{\frac{m+n}{2}} \int_{-\infty}^{\infty} \frac{n-2}{y^2} \frac{\alpha}{\alpha + \beta} y \, dy.

Consistent with previous notation, let \( F_k(x) \) denote the c.d.f. of a central chi-squared distribution with \( k \) degrees of freedom. Then simple algebra yields

(4.6.23) \quad J_{ij}(-\theta) \approx \frac{\frac{m+n}{2}}{\frac{\alpha^2}{\alpha + \beta}} \frac{\frac{m+n}{2}}{(\frac{\beta^2}{\alpha + \beta})} \left[ 1 - F_{n+b}^{(-\theta)} \right].

Substitution of (4.6.23) into (4.6.20) gives for large negative \( t \),

(4.6.24) \quad H(t) \approx \sum_{i} \sum_{j} q_i \overline{q}_j (\frac{\beta}{\alpha + \beta})^2 \left[ 1 - F_{N+2j}^{(-\frac{t}{\beta})} \right].

Next assume \( t \geq 0 \). Then analogously,

(4.6.25) \quad 1 - H(t) = \int_{t}^{\infty} \frac{h(x)}{x} \, dx = \sum_{i} \sum_{j} \frac{q_i \overline{q}_j c}{\Gamma(\frac{M+2i}{2})} J_{ij}^*(t), \quad t \geq 0,

where

(4.6.26) \quad J_{ij}^*(t) = \int_{t}^{\infty} x^\frac{m+n-2}{2} e^{-\frac{x}{2\alpha}} \psi\left(\frac{\alpha \beta}{\alpha + \beta}, \frac{m+n}{2}, \frac{\alpha + \beta}{\alpha} x\right) \, dx, \quad t \geq 0.

It is clear by comparison with (4.6.21) that \( J_{ij}^*(t, \alpha, \beta, m, n) = J_{ij}(-t, \beta, \alpha, n, m) \). Hence,

(4.6.27) \quad J_{ij}^*(t) \approx \frac{\frac{m+n}{2} \frac{n}{\beta^2} \frac{m+n}{\alpha^2} \Gamma\left(\frac{m}{2}\right)}{(\alpha + \beta)^{n/2}} \left[ 1 - F_{m}^{(-\frac{t}{m \alpha})} \right].
and substitution gives, for large positive $t$,

$$
(4.6.28) \quad 1 - H(t) \approx \sum_1^N \sum_j q_i^* q_j^* \frac{N+2j}{(\alpha + \beta)^2} \left[ 1 - F_{M+2J} \left( \frac{t}{\alpha} \right) \right].
$$

In summary, the tail probabilities of the indefinite quadratic form defined in (4.6.17) or in Theorem 4.6.3 are given asymptotically by (4.6.24) and (4.6.28).

It is possible to further simplify the formulas for the tail probabilities by invoking approximations introduced previously. Consider first the lower tail of the distribution. Equation (4.6.24) can be written for $\theta \geq 0$ as

$$
(4.6.29) \quad H(-\theta) \approx \left\{ \sum_1^N q_i^* \frac{M+2i}{(\alpha + \beta)^2} \right\} \left\{ 1 - \sum_j q_j^*_{M+2j} \left( \frac{\theta}{\beta} \right) \right\}.
$$

Now associate $\sum_j q_j^*_{M+2j} \left( \frac{\theta}{\beta} \right)$ with the c.d.f. of the positive definite quadratic form $V$ defined in (4.6.10). We use the "cumulant matching method" of Section 3.2.5 and conclude from equation (3.2.44) and Theorem 4.2.1,

$$
(4.6.30) \quad \sum_j q_j^*_{M+2j} \left( \frac{\theta}{\beta} \right) = P(V \leq \theta) \approx \rho X_v^2 \leq \theta,
$$

where

$$
\rho = \frac{\sum_{i=0}^s c_i \left( n_1 + 2g_1 \right)}{\sum_{i=0}^s c_i \left( n_1 + g_1 \right)} \quad \text{and} \quad \theta = \frac{\sum_{i=0}^s c_i \left( n_1 + g_1 \right)}{\sum_{i=0}^s c_i \left( n_1 + g_1 \right)}.
$$

116
and $c_0 = \beta$, $c_i = \beta b_i$ for $i \geq 1$. Now (4.6.24) becomes (for $t \leq 0$),

\[
(4.6.31) \quad H(t) \approx \left\{ \sum_{i=0}^{\infty} q_i \left( \frac{\beta}{\alpha+\beta} \right)^{\frac{M}{2}} + 1 \right\} \left\{ 1 - F_v\left( -\frac{t}{\nu} \right) \right\}.
\]

Several cases can be treated separately.

**Case (1):** Suppose $\alpha \ll \beta$. Then $\beta(\alpha+\beta)^{-1} \approx 1$ and we have approximately

\[
(4.6.32a) \quad H(t) \approx \left( \frac{\beta}{\alpha+\beta} \right)^{\frac{M}{2}} \left[ 1 - F_v\left( -\frac{t}{\nu} \right) \right], \quad t \leq 0.
\]

**Case (2):** $\beta \ll \alpha$. Then $\beta(\alpha+\beta)^{-1} \approx 0$ and therefore forces the series to converge extremely rapidly. We take the first term as a reasonable approximation and obtain

\[
(4.6.32b) \quad H(t) \approx \left( \frac{\beta}{\alpha+\beta} \right)^{\frac{M}{2}} q_0 \left[ 1 - F_v\left( -\frac{t}{\nu} \right) \right], \quad t \leq 0,
\]

with $q_0$ defined by Theorem 4.2.1 for the positive definite form $U$ of (4.6.14); i.e.,

\[
q_0 = \exp\left( -\frac{1}{2} \left( \sum_{i=0}^{\infty} d_i^2 \right) \right) \cdot \prod_{i=1}^{\infty} a_i^{-\frac{m_i}{2}}.
\]

**Case (3):** $\alpha \approx \beta$. In this situation the series does not converge as rapidly as in Case (2) so that it may be necessary to take several terms of the series. Then
\[(4.6.32c) \quad H(t) \approx \left[ \sum_{i=0}^{k} q_i \left( \frac{B}{\alpha + \beta} \right)^{\frac{M}{2} + 1} \right] \left[ 1 - F_{\nu}(\frac{t}{\rho}) \right], \quad t \leq 0, \]

where \( k \) is chosen in accordance with the size of the ratio \( \alpha/\beta \) so that the neglected terms are insignificant.

For the upper tail probabilities analogous approximations can be made. Equation (4.6.28) rewritten becomes

\[(4.6.33) \quad 1 - H(t) \approx \left\{ 1 - \sum_{i=1}^{\nu} q_i F_{M+2i} \left( \frac{t}{\alpha} \right) \right\} \left\{ \sum_{j=1}^{N+2j} q_j \left( \frac{\alpha}{\alpha + \beta} \right)^2 \right\}. \]

Define

\[(4.6.34) \quad \rho^* = \frac{\sum_{i=0}^{r} c_i^*(m_i + 2d_i)}{\sum_{i=0}^{r} c_i^*(m_i + d_i)}, \quad \nu^* = \frac{\sum_{i=0}^{r} c_i^*(m_i + d_i)^2}{\sum_{i=0}^{r} c_i^*(m_i + 2d_i)}, \]

where the \( m_i, d_i \) refer to the representations in (4.6.9), (4.6.10), and \( c_0^* = \alpha, \quad c_i^* = \alpha a_i \) for \( i \geq 1 \). Then (4.6.33) becomes

\[(4.6.35) \quad 1 - H(t) \approx \left[ \sum_{j=1}^{\nu^*} q_j \left( \frac{\alpha}{\alpha + \beta} \right)^2 \right] \left[ 1 - F_{\nu^*}(\frac{t}{\rho^*}) \right], \quad t \geq 0. \]

By analogy with (4.6.32) obtain for \( \alpha \ll \beta, \)

\[(4.6.36a) \quad 1 - H(t) \approx \left( \frac{\alpha}{\alpha + \beta} \right)^{\frac{N}{2}} q_0^* \left[ 1 - F_{\nu^*}(\frac{t}{\rho^*}) \right], \quad t \geq 0, \]

\( q_0^* = \exp\left\{ -\frac{1}{2} \sum_{i=0}^{s} s_i^2 \right\} \cdot \prod_{i=1}^{n_i} \frac{s_i}{b_i} \cdot \frac{n_i}{2} \).
For \( \beta \ll \alpha \),

\[
(4.6.36B) \quad 1 - H(t) \approx \left( \frac{\alpha}{\alpha + \beta} \right)^{\frac{N}{2}} [1 - F_{\nu^*}(\frac{t}{\rho^*})], \quad t \geq 0,
\]

For \( \alpha \approx \beta \),

\[
(4.6.36C) \quad 1 - H(t) \approx \sum_{j=0}^{k^*} g_j^* \left( \frac{\alpha}{\alpha + \beta} \right)^{\frac{N+2j}{2}} [1 - F_{\nu^*}(\frac{t}{\rho^*})], \quad t \geq 0,
\]

and \( k^* \) is the number of terms required for the neglected terms to be insignificant.

The results of this section are summarized in

**Theorem 4.6.4:**

Let \( T \) be an indefinite quadratic form in non-central variates defined either by Theorem 4.6.3 or by equations (4.6.9)-(4.6.11). The lower and upper tail probabilities of the distribution of \( T \) are given asymptotically by equations (4.6.24), (4.6.28), respectively, and approximately for several different cases by (4.6.32), (4.6.36), respectively.
CHAPTER 5
DISTRIBUTION OF A SAMPLE COVARIANCE

5.1 Introduction

In this chapter we derive the distribution of the sample covariance statistic under the assumption that the sample is obtained from a non-singular bivariate normal distribution. The distribution obtained is absolutely continuous with all intervals on the real line having positive probability. It will be shown how the sample covariance can be expressed as the difference of two gamma distributed variates, and the moments and several properties of the distribution will be developed.

As an illustration of how the need for this statistic can arise, let \( \mathcal{L}(X) = N(0, \Sigma) \) where \( X \) is a bivariate vector. Let \( V/n \) denote the \( 2 \times 2 \) sample covariance matrix based upon a sample of size \( n \). Then it is well known that \( \mathcal{L}(V) = W(\Sigma, 2, n) \) where the notation is the same as that in Section 3.2.6. Let \( \Sigma^{-1} = \begin{pmatrix} a & \tau \\ \tau & b \end{pmatrix} \), where \( a \) and \( b \) are known constants. Then it is easy to check that if \( V = (v_{11}, v_{12}) \), the sample covariance, is sufficient for the parameter \( \tau \). Hence, the distribution of \( v_{12} \) will be needed for problems involving estimation and tests of hypotheses concerning \( \tau \).

5.2 Derivation of the Distribution.

The main result of this section is given in

**Theorem 5.2:** Let \( \mathcal{L}(V) = W(\Sigma, 2, 2m) \), where \( m = 1, 2, \ldots \) and \( W \) denotes the Wishart distribution. Then if \( V = (v_{ij}) \), and \( \Sigma = (\sigma_{ij}) \),

120
\[ |\Sigma| \neq 0, \text{ the probability density function of } v_{12} \text{ is given by} \]

\[ f(x) = [b_0 + b_1|x| + \ldots + b_{m-1}|x|^{m-1}] e^{-\frac{|x|}{\omega(x)}} \]

where

\[ \omega(x) = \sqrt{\sigma_{11}\sigma_{22}} + \sigma_{12} \operatorname{sgn}(x), \]

\[ b_j = \binom{m-1}{j} \frac{\Gamma(2m-j-1)}{\Gamma(2m)|\Sigma|^m} \left( \frac{|\Sigma|}{2\sqrt{\sigma_{11}\sigma_{22}}} \right)^{2m-j-1}. \]

**Proof:** The theorem is proved with the aid of several technical lemmas.

**Lemma 5.2.1:** If \( z(v) = W(\Sigma, 2, 2m) \), \( m = 1, 2, \ldots \) where \( V \) is a \( 2 \times 2 \) random matrix \( (v_{ij}) \), the characteristic function of \( v_{12} \) is given by

\[ \varphi(t) = (1 - 2i\sigma_{12}t + |\Sigma|t^2)^{-m}. \]

**Proof:** In the joint characteristic function of the elements of \( V \), let arguments corresponding to \( v_{11} \) and \( v_{22} \) vanish.

The distribution of \( v_{12} \) is now found from the standard inversion formula. Since it is clear that \( \varphi(t) \) is absolutely integrable over the doubly infinite range, a density must exist, and it is given by

\[ f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(-itx)}{(1 - 2i\sigma_{12}t + |\Sigma|t^2)^m} \, dt, \]

wherever the integral is defined. Note that \( |\Sigma| > 0 \). The integral is evaluated using the theory of residues. The pertinent results of complex variable theory are recalled in the following lemmas.

121
Lemma 5.2.2: Let \( Q(z) \) be a function of the complex variable \( z \). If \( Q(z) \) has a pole of order \( m \) at the point \( z=z_0 \), the residues of \( Q(z) \) at \( z_0 \) is given by

\[
(5.2.6) \quad r_{-1} = \frac{1}{\Gamma(m)} \frac{d^{m-1}}{dz^{m-1}} [(z-z_0)^m Q(z)] \bigg|_{z=z_0}.
\]

Proof: Since the singularity of \( Q(z) \) at \( z_0 \) is non-essential, the principal part of \( Q(z) \) is a finite series and \( Q(z) \) can be put in the form

\[
Q(z) = \sum_{j=0}^{\infty} r_j (z-z_0)^j + \frac{r_{-1}}{z-z_0} + \frac{r_{-2}}{(z-z_0)^2} + \cdots + \frac{r_{-m}}{(z-z_0)^m}.
\]

Multiplication through by \( (z-z_0)^m \) gives

\[
Q^*(z) \equiv (z-z_0)^m Q(z) = r_{-m} + r_{-m+1}(z-z_0) + \cdots \quad + r_{-1}(z-z_0)^{m-1} + r_0(z-z_0)^m + \cdots,
\]

where \( r_{-m} \neq 0 \). But this is just a Taylor series expansion for \( Q^*(z) \) about \( z_0 \). Hence, the coefficient of the \( (m-1) \) power of \( (z-z_0) \) is given by the result in Lemma 5.2.2.

Lemma 5.2.3: Assume \( Q(z) \) is analytic when \( \text{Im}(z) \geq 0 \), except at a finite number of poles (\( Q(z) \) is meromorphic); \( Q(z) \) has no poles on the real axis; as \( |z| \to \infty \), \( zQ(z) \to 0 \) uniformly for all values of arg \( z \) such that \( 0 \leq \text{arg } z \leq \pi \); when \( t \) is real, \( tQ(t) \to 0 \) as \( t \to \pm \infty \) in such a way that \( \int_{0}^{\infty} Q(t)dt \) and \( \int_{-\infty}^{0} Q(t)dt \) both converge. Then \( \int_{-\infty}^{\infty} Q(t)dt = 2\pi i \sum R_k \), where \( R_k \) denotes the residual of \( Q(z) \) at its \( k^{th} \) pole above the real axis.

122
Proof: See Whittaker and Watson [36], p. 113.

Now the integral in (5.2.5) can be evaluated simply. Assume first that $x < 0$. Define

$$Q(z) = \frac{1}{2\pi} \frac{\exp(-izx)}{(1 - 2iz\alpha_2 + |\Sigma|^m)^m},$$

and

$$\begin{cases}
\beta_1 = i/(\sqrt{\sigma_{11} \sigma_{22}} - \sigma_{12}) \\
\beta_2 = i/(\sqrt{\sigma_{11} \sigma_{22}} + \sigma_{12})
\end{cases}$$

and note that $\sqrt{\sigma_{11} \sigma_{22}} \pm \sigma_{12} > 0$, by positive definiteness. Now $Q(z)$ can be written in the factored form

$$Q(z) = \frac{\exp(-izx)}{2\pi |\Sigma|^m(z - \beta_1)^m(z - \beta_2)^m}.$$

It is readily verifiable that the meromorphic function $Q(z)$ with two $m$th order poles at $\beta_1, \beta_2$, satisfies the conditions of Lemma 5.2.3. Hence,

$$f(x) = \int_{-\infty}^{\infty} Q(t) \, dt = 2\pi i R_1, \quad x < 0$$

where $R_1$ is the residue of $Q(z)$ at $z = \beta_1$ (since $\beta_1$ lies in the upper half plane, and $\beta_2$ in the lower half plane).

Next note that since the expression for the joint density

$$p(v_{11}, v_{22}, v_{12}) = c|V|^{-\frac{n-p-1}{2}} |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \text{tr } \Sigma^{-1} V\right)$$

shows that if $v_{12}$ is transformed into $-v_{12}$, $|V|$ does not change and the density remains invariant if $\sigma_{12}$ is transformed into $-\sigma_{12}$, it is unnecessary
to compute \( f(x) \) for \( x > 0 \) separately. That is, \( f(x) \) for \( x > 0 \)
can be obtained from \( f(x) \) for \( x < 0 \) by transforming \( \sigma_{12} \) into
\(-\sigma_{12}\) (and of course \( x \) to \(-x\)).

\( R_1 \) is evaluated from Lemma 5.2.2 and is given by

\[
R_1 = \frac{1}{\Gamma(m)} \frac{d^{m-1}}{dz^{m-1}} \left[ \frac{\exp(-izx)}{2\pi |\Sigma|^{m}(z-\beta_2)^m} \right]_{z=\beta_1}.
\]

Substitution into (5.2.10) gives, for \( x < 0 \),

\[
f(x) = \frac{i}{|\Sigma|^{m} \Gamma(m)} \frac{d^{m-1}}{dz^{m-1}} \left[ \frac{\exp(-izx)}{(z-\beta_2)^m} \right]_{z=\beta_1}.
\]

Define \( g(z) = \frac{\exp(-izx)}{(z-\beta)^m} \). It may be verified by successive stages of
straightforward differentiation and the induction argument that

\[
\frac{d^{m-1}}{dz^{m-1}} [g(z)] = \exp(-izx) \sum_{k=0}^{m-1} \frac{(-ix)^k (-1)^{m-k-1}(2m-k-2)!}{\Gamma(m)(z-\beta)^{2m-k-1}} \binom{m-1}{k}.
\]

When this expression is evaluated at \( z = \beta_1 \), the required result is obtained. ||

5.3 Properties of the Distribution

In the following theorem, the distribution of \( V_{12} \) is
characterized in terms of gamma variates for an arbitrary (not just
an even number) of degrees of freedom. For this purpose we introduce
the notation \( \mathcal{U}(X) = \text{gamma}(\beta, \lambda) \) to denote the fact that \( X \) is a
random variable with density

\[
p(x; \beta, \lambda) = \lambda^{-\beta} \Gamma^{-1}(\beta) \exp(-x/\lambda) , \quad x > 0
\]

and zero otherwise. Then obtain

124
Theorem 5.3.1: Let \( z(V) = W(\Sigma, 2, n) \) where \( V = (v_{ij}) \), \( \Sigma = (\sigma_{ij}) \) are non-singular matrices of order 2. If \( X \) and \( Y \) are independent random variables such that \( z(X) = \text{gamma} \left( \frac{n}{2}, \frac{\sqrt{\sigma_{11} \sigma_{22}} + \sigma_{12}}{2} \right) \), and \( z(Y) = \text{gamma} \left( \frac{n}{2}, \frac{\sqrt{\sigma_{11} \sigma_{22}} - \sigma_{12}}{2} \right) \), then \( z(v_{12}) = z(X - Y) \).

Proof: Define \( \varphi_j(t) = (1 - it/k_j^2)^{-q}, \) for \( j = 1,2, \) where \( k_1^{-2} = \sqrt{\sigma_{11} \sigma_{22}} + \sigma_{12} \) and \( k_2^{-2} = \sqrt{\sigma_{11} \sigma_{22}} - \sigma_{12} \), and \( q = \frac{n}{2} \). Note that \( \varphi_1(t), \varphi_2(t) \) are the characteristic functions of random variables \( X, -Y, \) where \( z(X) = \text{gamma}(q, k_1^{-2}), z(Y) = \text{gamma}(q, k_2^{-2}), \) respectively. Moreover, since \( k_1^{-2}k_2^{-2} |\Sigma| = 1, \varphi(t) = \varphi_1(t)\varphi_2(t). \) Hence, \( X \) and \( (-Y) \) are independent. By the P. Lévy uniqueness theorem if \( Z = X - Y, \) the distribution of \( Z \) is the convolution of the \( X \) and \( (-Y) \) distributions.

Note that when \( q \) is an integer this theorem provides a second proof of Theorem 5.2; and when \( q \) is not an integer (as is the case when \( n \) is odd), the distribution of \( v_{12} \) can be found by this technique to be a function of confluent hypergeometric functions.

Theorem 5.3.2: Let \( z(V) = W(\Sigma, 2, 2m), \) and \( V = (v_{ij}), m = 1,2, \ldots. \) Then the \( k^{th} \) moment of \( v_{12} \) is given by

\[
\mu_k = \sum_{j=0}^{m-1} b_j \Gamma(j+k+1)[(\sqrt{\sigma_{11} \sigma_{22}} + \sigma_{12})^{j+k+1} + (-1)^j (\sqrt{\sigma_{11} \sigma_{22}} - \sigma_{12})^{j+k+1}],
\]

(5.2.13)

where the \( b_j \) are defined in (5.2.3).
Proof: \( \mu_k = \mathbf{E} V_{12}^k = \int_{-\infty}^{\infty} x^k f(x) \, dx \), where \( f(x) \) is defined in Theorem 5.2. Then,

\[
\mu_k = \int_{-\infty}^{0} x^k \sum_{j=0}^{m-1} (-1)^j b_j x^j e^{\frac{x}{a}} \, dx + \int_{0}^{\infty} x^k \sum_{j=0}^{m-1} b_j x^j e^{-\frac{x}{b}} \, dx,
\]

where \( a = \sqrt{\sigma_{11} \sigma_{22} - \sigma_{12}^2} \), \( b = \sqrt{\sigma_{11} \sigma_{22} + \sigma_{12}^2} \). Simplifying,

\[
\mu_k = \sum_{j=0}^{m-1} b_j \left[ (-1)^j I(a) + I(b) \right], \quad \text{where} \quad I(g) = \int_{0}^{\infty} x^{k+j} e^{-\frac{x}{g}} \, dx, \quad g > 0.
\]

This integral is easily seen to be \( I(g) = \Gamma(k+j+1) g^{k+j+1} \). Substitution yields the result.

**Theorem 5.3.3:** If \( \mathcal{X}(V_1) = W(\Sigma, 2, 2r) \), and \( \mathcal{X}(V_2) = W(\Sigma, 2, 2s) \), and \( V_1 \) and \( V_2 \) are independent, then the density of \( \left( v_{12}^{(1)} + v_{12}^{(2)} \right) \) is given by (5.2.1) with \( m = r+s \), where \( V_1 \equiv (v_{ij}^{(1)}) \), \( V_2 \equiv (v_{ij}^{(2)}) \).

**Proof:** This result follows directly from the well known analogous result on the addition of independent variates having Wishart distributions (see Anderson [2], p. 162). It also follows immediately from Theorem 5.3.1.
CHAPTER 6
EXTENSIONS OF RESULTS

6.1 Introduction.

In this chapter it is seen that the methods applied previously to specialized Behrens-Fisher and Classification problems can be extended to include a wider class of problems. The problems considered in Chapters 1 and 3 involved parameter spaces which were either completely unrestricted, or the space of covariance matrices was restricted to those matrices with intraclass structure (see Section 1.2.1). Although a parameter space restriction of arbitrary nature might be extremely intractable, the fact that the intraclass structured problem has the property that the covariance matrices can be diagonalized by a matrix whose elements do not depend upon the parameters permits it to be reduced to an equivalent problem in which the observation vectors have independent components, and the covariance matrices have only two distinct characteristic roots. The transformed Behrens-Fisher or Classification problems are then not too difficult to solve. A related question is whether or not this technique can be applied to a larger class of problems than those with covariance matrix having intraclass structure. An affirmative answer is found in the class of circulant matrices discussed below.

6.2.1 Circulant Covariance Matrices.

A circulant matrix of order \( n \) is defined (see e.g., Kowalewski [20]) as
(6.2.1) \[ A(a_1, \ldots, a_n) = \begin{pmatrix}
  a_1 & a_2 & a_3 & \cdots & a_n \\
  a_n & a_1 & a_2 & \cdots & a_{n-1} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  a_2 & a_3 & a_4 & \cdots & a_1
\end{pmatrix}. \]

Its properties are well known and are summarized below.

Let \( r_1, r_2, \ldots, r_n \) denote the \( n \) roots of unity and choose

(6.2.2) \[ r_k = \exp \left[ \frac{2\pi i}{n} (k-1) \right]. \]

Define

(6.2.3) \[ \lambda_k = \sum_{j=1}^{n} a_j r_k^{j-1}. \]

Since \( r_k^n = 1 \), for every \( 1 \leq k \leq n \), by multiplying (6.2.3) successively by increasing powers of \( r_k \), it is not hard to see that \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are the eigenvalues of \( A \), and

(6.2.4) \[ v'_k = \frac{1}{\sqrt{n}} (1, r_k, r_k^2, \ldots, r_k^{n-1}), \quad 1 \leq k \leq n \]

are their associated normalized eigenvectors. Note that the \( v_k \) do not depend upon the elements \( a_1, \ldots, a_n \).

Now restrict the class of circulants to be considered to those for which \( A = A' \). Then since \( a_{n-j+1+k} = a_{n-k+1+j} \), for \( j = 1 \) take

(6.2.5) \[ a_{n+1} = a_k = a_{n-k+2}. \]

Since \( r_k^{n-p+1} = r_k^{p-1} \), for the symmetric circulants

(6.2.6) \[ \lambda_k = \lambda_{n-k+2}, \quad k = 2, 3, \ldots, n, \]

128
and of course all \( \lambda_k \) will be real valued. They are given explicitly by

\[
(6.2.7) \quad \lambda_k = \sum_{p=1}^{n} a_p \cos \frac{2\pi}{n} (k-1)(n-p+1), \quad k = 1, \ldots, n.
\]

The following geometrical argument can be used to establish (6.2.7). From (6.2.2), (6.2.3),

\[
(6.2.8) \quad \lambda_k = \sum_{j=1}^{n} a_j e^{\frac{2\pi i}{n} (k-1)(j-1)}.
\]

Let \( \theta_j = \frac{2\pi}{n} (k-1)(j-1) \). The vectors \( e^{i\theta_j} \) for \( 1 \leq j \leq n \) can be plotted in the complex plane as \( n \) points, \( P_1, P_2, \ldots, P_n \), lying on the circumference of a circle around the origin, of unit radius. All \( P_j \) located on the horizontal axis have the property that \( \sin \theta_j = 0 \). All \( P_j \) off the horizontal axis are paired with points \( P_j^* \) (corresponding to some \( \theta_j^* \)) with the property that \( \sin \theta_j^* = -\sin \theta_j \).

Moreover, reference to (6.2.5) shows such paired points have the same coefficients. Hence, all \( \sin \theta_j \) terms either vanish or cancel and the imaginary part of (6.2.8) disappears, leaving (6.2.7).

Let \( v_{ks} \) denote the \( s^{th} \) component of the eigenvector associated with \( \lambda_k \). Direct computation shows that for symmetric circulants,

\[
(6.2.9) \quad v_{ks} = \frac{1}{\sqrt{n}} \left[ \cos \frac{2\pi}{n} (k-1)(s-1) + \sin \frac{2\pi}{n} (k-1)(s-1) \right].
\]

Note as in (6.2.4) that \( v_{ks} \) possesses the property of being independent of the elements of \( A \).
Finally, restrict $A$ so that not only $A(a_1, \ldots, a_n) = A'(a_1, \ldots, a_n)$ but also, the $n$ independent elements of $A$ have the property that the determinants of all principal submatrices of $A$ are strictly positive; that is, $A$ is positive definite. This restriction in no way alters the eigenvalues except that now $\lambda_k > 0$, $k = 1, \ldots, n$. This class is the class of circulant covariance matrices.

6.2.2 Applications.

The circulant covariance matrices arise naturally in many problems in which there is some type of lattice array in temporal or spacial coordinates. Thus, for example, there may be a spacial array of identical seismometers with neglectable internal noise, but high level noise generated between instruments and correlated (see Abramson [1]). Similar arrays of radar stations have been analyzed (see Capon [11]) using circulants, and Berlin and Kac [9] have constructed a spherical model of a ferromagnet based on the supposition that there is a spin at each site of a regular lattice of $n$ sites, and the associated partition functions involve quadratic forms in circulants.

A particular circulant covariance matrix which has application in many problems has the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ \vdots & \vdots & \vdots \\ 0 & \rho & 1 \end{pmatrix}$$

(6.2.10)

in which all elements in the lattice have the same variance but only adjacent elements are correlated. Such situations might occur in arrays of biological or chemical gas detecting instruments in which
the interstices had large separation compared with the distance required for signal attenuation.

The covariance matrix with intraclass structure is the special case of the positive definite circulants in which

\[ a_2 = a_3 = \ldots = a_n \neq a_1. \]

Procedures for handling the Behrens-Fisher problems treated in Chapter 1 for intraclass structured covariance matrices can be extended to the positive definite circulants in a direct manner. The development of likelihood ratio tests will require maximization of the corresponding density functions over an increased number of parameters although conceptually the problems are identical and the same reduction to canonical form transformations will be applicable.

The Classification procedures for intraclass structured covariance matrices, developed in Section 3.3, will also extend directly to positive definite circulants. Moreover it is clear at the outset that the associated distribution theory required for computing the classification region boundaries corresponds to that found for general covariance matrices except that in this case there is a restriction to a subspace of dimension \( r \), where \( r \) is the number of distinct eigenvalues of the positive definite circulant (in the general case, the distribution of the test statistic depends on the weighted sum of \( p \) independent chi-squared variates, and for intraclass structure, it depends on two such variates).
ACKNOWLEDGMENTS

My principal advisor, Dr. Ingram Olkin, was a continuous source of encouragement throughout this investigation. His helpful recommendations made our many discussions stimulating and fruitful.

I wish to thank my other advisors, Dr. Norman Abramson for pointing out some of the problems to which the methods of this study can be applied, and Dr. Gerald Lieberman for his patience in reading the manuscript.

I am thankful, also, to Dr. Herbert Solomon for his support and encouragement; and to the National Science Foundation for their support.

Finally, I am grateful to Mrs. Evelyn Getz for typing this manuscript and taking a personal interest in its final preparation.
REFERENCES


