VANCE, Lonnie Charles, 1941-
DISTRIBUTION OF DISCRIMINANT FUNCTIONS WITH UNEQUAL COVARIANCE MATRICES UNDER INTRACLASS CORRELATION MODELS.

Iowa State University, Ph.D., 1973
Statistics

Distribution of discriminant functions with unequal
covariance matrices under intraclass
correlation models

by

Lonnie Charles Vance

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1973

TABLE OF CONTENTS

I. INTRODUCTION AND REVIEW OF THE LITERATURE

A. Introduction

The problem of classifying an observation $\underline{X}$ into one of two (or more) populations is commonplace in statistical literature. Let the two populations be denoted by π_1 and π_2. It is usually assumed that the populations are multivariate normal with mean vectors $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$, respectively, and with common covariance matrix $\underline{\Sigma}$ where $\underline{\Sigma}$ is positive-definite. Let the random vector $\underline{X}$ have a k-dimensional normal distribution, then $\underline{X}$ has probability density

$$f(\underline{X}) = (2\pi)^{-k/2} \, |\underline{\Sigma}|^{-1/2} \exp\{-\tfrac{1}{2}(\underline{X}-\underline{\mu})'\underline{\Sigma}^{-1}(\underline{X}-\underline{\mu})\}.$$

The distribution of $\underline{X}$ will be denoted by $N_k(\underline{\mu},\underline{\Sigma})$. (The subscript k will not be written when k=1.) If $\underline{X}\varepsilon\pi_i$, then the distribution of $\underline{X}$ is $N_p(\underline{\mu}^{(i)},\underline{\Sigma})$, which we write as $\underline{X} \sim N_p(\underline{\mu}^{(i)},\underline{\Sigma})$, i=1,2. Now the observation $\underline{X}$ is assigned to π_1 if the likelihood ratio $\lambda = \frac{f_1(\underline{X})}{f_2(\underline{X})} \geq c$ and to π_2 if $\lambda<c$ where c is a constant and $f_i(\underline{X})$ is the density of the random vector $\underline{X}$ in population π_i, i=1,2. Using the likelihood ratio procedure when $\underline{\mu}^{(1)} \neq \underline{\mu}^{(2)}$ and $\underline{\Sigma}$ is the common covariance matrix, the discriminant function obtained is linear in $\underline{X}$, often referred to as Fisher's linear discriminant function.

In many situations the assumption that the covariance matrix $\underline{\Sigma}$ is the same in the two populations seems unlikely. Further, if both populations have a common covariance matrix, then no discrimination is possible when $\underline{\mu}^{(1)} = \underline{\mu}^{(2)}$.

Bartlett and Please (1963) obtained the discriminant function for assigning an observation to one of two multivariate normal populations when $\underline{\mu}^{(1)} = \underline{\mu}^{(2)}$ (zero mean difference) and $\underline{\Sigma}^{(1)} \neq \underline{\Sigma}^{(2)}$ where $\underline{\Sigma}^{(i)}$ is the covariance matrix in population π_i, i=1,2. In particular, Bartlett and Please obtained the discriminant function when $\underline{\mu}^{(1)} = \underline{\mu}^{(2)}$ and the covariance matrices $\underline{\Sigma}^{(i)}$ are of the form

$$\underline{\Sigma}^{(i)} = \begin{bmatrix} \sigma_i^2 & \sigma_i^2\rho_i & \cdots & \sigma_i^2\rho_i \\ \sigma_i^2\rho_i & \sigma_i^2 & \cdots & \sigma_i^2\rho_i \\ \vdots & \vdots & & \vdots \\ \sigma_i^2\rho_i & \sigma_i^2\rho_i & \cdots & \sigma_i^2 \end{bmatrix} = \sigma_i^2[(1-\rho_i)\underline{I}+\rho_i\underline{J}], \quad i=1,2, \qquad (1.1)$$

where $\sigma_1^2 \neq \sigma_2^2$, $\underline{I}$ is the (pxp) identity matrix, and $\underline{J}$ is a (pxp) matrix of ones. The matrices, $\underline{\Sigma}^{(i)}$, represented in (1.1) are said to have the intraclass correlation matrix pattern. When $\underline{\Sigma}^{(1)} \neq \underline{\Sigma}^{(2)}$, the discriminant function obtained is a quadratic discriminant function.

In this thesis, we shall consider intraclass correlation models with unequal covariance matrices. The cases of equal or unequal mean vectors will be treated. Further, if covariates are available, they may be used in the discriminant function. In view of these, three situations appear to be of importance both from a theoretical and a practical viewpoint.

Suppose there are two normal populations

$$\pi_i: N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(i)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}^{(i)}\right), \ i=1,2,$$

where $\underline{\mu}^{(i)}$ is a (px1) vector and $\underline{\mu}$ is a (qx1) vector. Let $\underline{\Sigma}^{(i)}$ be of the form

$$\underline{\Sigma}^{(i)} = \begin{bmatrix} \sigma_i^2\,[(1-\rho_i)\underline{I}+\rho_i\underline{J}] & \sigma_i\sigma\rho_i'\,\underline{J} \\ \sigma_i\sigma\rho_i'\,\underline{J} & \sigma^2[(1-\rho)\underline{I}+\rho\underline{J}] \end{bmatrix}$$

$$= \begin{bmatrix} \underline{\Sigma}_{11}^{(i)} & \underline{\Sigma}_{12}^{(i)} \\ \underline{\Sigma}_{21}^{(i)} & \underline{\Sigma}_{22} \end{bmatrix}, \tag{1.2}$$

for i=1,2, where $\sigma_1^2 \neq \sigma_2^2$. For i=1,2, the matrix $\underline{\Sigma}_{11}^{(i)} = \sigma_i^2\,[(1-\rho_i)\underline{I} + \rho_i\underline{J}]$ is a (pxp) positive-definite matrix, $\underline{\Sigma}_{12}^{(i)} = \sigma_i\sigma\rho_i'\,\underline{J}$ is a (pxq) matrix, $\underline{\Sigma}_{21}^{(i)}$ is the transpose of $\underline{\Sigma}_{12}^{(i)}$, written $\underline{\Sigma}_{21}^{(i)} = [\Sigma_{12}^{(i)}]'$, and $\underline{\Sigma}_{22} = \sigma^2[(1-\rho)\underline{I} + \rho\underline{J}]$ is a (qxq) positive-definite matrix. The matrices $\underline{\Sigma}_{11}^{(1)}$, $\underline{\Sigma}_{11}^{(2)}$, and $\underline{\Sigma}_{22}$ have intraclass correlation structure. If $\underline{X} = \begin{pmatrix}\underline{X}_1\\ \underline{X}_2\end{pmatrix} \epsilon \pi_i$, then

$$\underline{X} \sim N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(i)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}^{(i)}\right), \quad i=1,2,$$

where $\underline{X}_1$ is a (px1) vector and $\underline{X}_2$ is a (qx1) vector. Notice that $\underline{X}_2$ has the same mean and the same covariance matrix in the two populations. Hence $\underline{X}_2$ alone does not have discriminating power, and it may be treated as q covariates.

In the first situation, we consider only the variates of $\underline{X}_1$ as discriminators ignoring $\underline{X}_2$ completely. Therefore, $\underline{X}_1 \sim N_p(\underline{\mu}^{(i)}, \underline{\Sigma}_{11}^{(i)})$ when the observation comes from π_i, i=1,2. In Chapter II, under the assumption that $\underline{\Sigma}_{11}^{(1)}-\underline{\Sigma}_{11}^{(2)}$ is positive-definite, the discriminant function is given and its distribution is derived when all parameters are known. The distribution is also obtained when the means $\underline{\mu}^{(i)}$, equal or unequal, are unknown, but the covariance matrices are known. When the parameters are unknown, an asymptotic expansion for the distribution of the discriminant function is derived.

The second situation would be to view the p variates of $\underline{X}_1$ as discriminators and the q variates of $\underline{X}_2$ as covariates. Cochran and Bliss (1948) and Cochran (1964a) have considered this situation when the two populations have the same covariance matrix.

Suppose $\underline{B}^{(i)} = \underline{\Sigma}_{12}^{(i)}\underline{\Sigma}_{22}^{-1}$ is the matrix of regression coefficients of the discriminators $\underline{X}_1$ on the covariates $\underline{X}_2$ in population π_i, i=1,2. Ordinarily, one does not know to which population π_i the observation belongs. Accordingly, a natural procedure is to form the variates

$$\underline{Z}_i = \underline{X}_1 - \underline{B}^{(i)}\underline{X}_2 = \underline{X}_1 - \underline{\Sigma}_{12}^{(i)}\Sigma_{22}^{-1}\underline{X}_2 \tag{1.3}$$

and then form the weighted variates

$$\underline{Z} = W_1\underline{Z}_1 + W_2\underline{Z}_2 = \underline{X}_1 - (W_1\underline{B}^{(1)} + W_2\underline{B}^{(2)})\underline{X}_2, \tag{1.4}$$

where the W_i are weights. (The choice of weights W_i will be discussed in Chapter V.) The discriminant function could then be calculated in the usual way from $\underline{Z}$. Let $\underline{V}^{(i)}$ denote the covariance matrix of $\underline{Z}$ when $\underline{X}\varepsilon\pi_i$, i=1,2. In Chapter III, using constant weights W_1 and W_2 such that $W_1 + W_2 = 1$ and under the assumption that $\underline{V}^{(1)}-\underline{V}^{(2)}$ is positive-definite, the distribution of the discriminant function is obtained when all parameters are known. Also the distribution is derived when the means $\underline{\mu}^{(i)}$ are unknown, but the covariance matrices are known. When all parameters are unknown, the limiting distribution for the discriminant function is found.

In Chapter IV, we consider the third situation where the q covariates, $\underline{X}_2$, are included in the discriminant function as discriminators. Therefore, all of the p+q variates of $\underline{X}$ are treated as discriminators. If the observation comes from π_i, then

$$\underline{X} \sim N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(i)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}^{(i)}\right), \quad i=1,2,$$

where $\underline{\Sigma}^{(i)}$ is given in (1.2). The discriminant function is calculated using the p+q variates of $\underline{X} = \begin{pmatrix}\underline{X}_1\\ \underline{X}_2\end{pmatrix}$. Its distribution is obtained when all parameters are known and also for the case when the means $\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}$ are unknown, but the covariance matrices are known. When all parameters are unknown, asymptotic results are obtained.

Based on the above setting, we might wish to determine which one of these three classification procedures would be preferable. One criterion used for comparing methods of classification is the minimization of some function of the probabilities of misclassification P(2/1) and P(1/2) where P(i/j) is the probability of classifying the observation $\underline{X}$ as belonging to π_i when $\underline{X}$ comes from π_j. In Chapter V, a comparison of the probabilities of misclassification for these three classification procedures is made when π_i are bivariate (p=1, q=1) normal populations or when the populations π_i are trivariate (p=1, q=2) normal populations.

B. Review of the Literature

1. Linear discriminant function

Fisher (1936) developed the linear discriminant function as a classificatory measure. However, Pearson (1926) proposed a coefficient of racial likeness which would express the measure of resemblance (or divergence) between two groups. Pearson's coefficient was essentially the ratio of the difference between the group means to a pooled standard deviation of the group means with the assumption that the mean differences were independent. Mahalanobis (1930, 1936) proposed a measure of distance using the within group standard deviation in the denominator, which was later termed the generalized distance or Mahalanobis' distance. Mahalanobis' distance provided a measure of the magnitude of separation of the two groups.

Suppose a random sample of size N_i is taken from a p-variate population π_i, i=1,2. Let $Y = \sum_{j=1}^{p} L_j U_j$ denote a linear function of characters U_j, j=1,2,...,p. Fisher (1936) defined the discriminant function between the two populations as that linear function of the characters for which the ratio

$$[\sum_{j=1}^{p} L_j(\bar{U}_j^{(1)} - \bar{U}_j^{(2)})]^2 / \sum_{j=1}^{p} \sum_{k=1}^{p} L_j L_k S_{jk} \qquad (1.5)$$

is maximized where $\bar{U}_j^{(i)}$ is the mean of the jth character in population π_i, i=1,2, and S_{jk} is the pooled estimate of the covariance between the jth and kth characters. Fisher found that the coefficients L_j, j=1,2,...,p, which maximize the above ratio are

$$L_j = \sum_{k=1}^{p} S^{jk}(\bar{U}_k^{(1)} - \bar{U}_k^{(2)}), \qquad (1.6)$$

where

$$S^{jk} = (S_{jk})^{-1}.$$

Welch (1939) derived a discriminant function using the likelihood ratio procedure of Neyman and Pearson. He supposed that a priori probabilities q_i of drawing an individual from π_i with density $p_i(\underline{X})$, i=1,2, were known. If there were two multivariate normal populations π_1 and π_2 with common covariance matrix $\underline{\Sigma}$ and mean vectors $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$, respectively, and if all parameters were known, then Welch classified an individual $\underline{X}$ into π_1 if $\lambda=p_1(\underline{X})/p_2(\underline{X}) \geq c$ and into π_2 if $\lambda<c$ where $c=q_2/q_1$. Taking the logarithm of $p_1(\underline{X})/p_2(\underline{X})$, Welch obtained as the discriminant function

$$U = \underline{X}'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) - \tfrac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) \quad . \quad (1.7)$$

The first term in (1.7) is Fisher's linear discriminant function when all parameters are known. The distribution of U is easily obtained.

Wald (1944) introduced the cost function into the constant c. Anderson (1958) presented a thorough discussion of classification including decision theory considerations. If π_i: $N_p(\underline{\mu}^{(i)},\underline{\Sigma})$ with probability density $p_i(\underline{X})$ where $\underline{\mu}^{(i)}$ and $\underline{\Sigma}$ are known, if q_i are known a priori probabilities, $i=1,2$, and if $C(i/j)$ is the cost of misclassifying an individual from π_j as from π_i, then Anderson (1958) proved that the "best" regions of classification are given by

$$R_1: \underline{X}'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) - \tfrac{1}{2}(\underline{\mu}^{(1)}+\underline{\mu}^{(2)})'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) \geq \log k,$$

$$R_2: \underline{X}'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) - \tfrac{1}{2}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})'\underline{\Sigma}^{-1}(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}) < \log k, \quad (1.8)$$

where $k = q_2C(1/2)/q_1C(2/1)$ and R_i denotes the region of classification into π_i. The "best" regions refer to the regions that minimize the expected loss $q_1C(2/1)P(2/1) + q_2C(1/2)P(1/2)$.

In the case when all parameters are unknown, a random sample $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \ldots, \underline{X}_{N_1}^{(1)}$ is taken from $\pi_1: N_p(\underline{\mu}^{(1)}, \underline{\Sigma})$ and an independent random sample $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \ldots, \underline{X}_{N_2}^{(2)}$ is taken from $\pi_2: N_p(\underline{\mu}^{(2)}, \underline{\Sigma})$. On the basis of this information we wish to classify an observation $\underline{X}$ as coming from π_1 or π_2. Our estimate of $\underline{\mu}^{(i)}$ is the sample mean $\underline{\overline{X}}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \underline{X}_\alpha^{(i)}$ and of $\underline{\Sigma}$ is $\underline{S}$ where

$$\underline{S} = \frac{1}{N_1+N_2-2} \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (\underline{X}_\alpha^{(i)} - \underline{\overline{X}}^{(i)})(\underline{X}_\alpha^{(i)} - \underline{\overline{X}}^{(i)})' \quad . \qquad (1.9)$$

Substituting these estimates for the unknown parameters in (1.7) we obtain

$$W = \underline{X}'\underline{S}^{-1}(\underline{\overline{X}}^{(1)}-\underline{\overline{X}}^{(2)}) - \tfrac{1}{2}(\underline{\overline{X}}^{(1)}+\underline{\overline{X}}^{(2)})'\underline{S}^{-1}(\underline{\overline{X}}^{(1)}-\underline{\overline{X}}^{(2)}) . \qquad (1.10)$$

This is often called Anderson's statistic W. The distribution of W is very complicated and has been considered by Anderson (1951), Wald (1944), Sitgreaves (1952), and others.

Okamoto (1963) has given an asymptotic expansion for the distribution of the discriminant function W for the case $\underline{\mu}^{(1)} \neq \underline{\mu}^{(2)}$ and $\underline{\mu}^{(1)}, \underline{\mu}^{(2)}$ unknown when the covariance matrix is the same in both populations but is general and unknown.

The statistic Z (see Kudo, 1959 and John, 1960) where

$$Z = \frac{N_1}{N_1+1} (\underline{X}-\overline{\underline{X}}^{(1)})'\underline{S}^{-1}(\underline{X}-\overline{\underline{X}}^{(1)}) - \frac{N_2}{N_2+1}(\underline{X}-\overline{\underline{X}}^{(2)})'\underline{S}^{-1}(\underline{X}-\overline{\underline{X}}^{(2)}) \tag{1.11}$$

is a criterion which has been proposed as a competitor to the Anderson statistic W where $\overline{\underline{X}}^{(i)}$ and $\underline{S}$ are defined above. Memon (1968) and Memon and Okamoto (1971) have obtained an asymptotic expansion of the distribution of Z.

2. Discriminant analysis with covariance

The idea of combining discriminant analysis with the analysis of covariance was proposed by Cochran and Bliss (1948). This case occurred when in addition to the discriminators there were measurements (covariates) whose means were known to be the same in both populations. Both populations were assumed to have the same covariance matrix. Although such variates have no discriminating power by themselves, they may still be utilized in the discriminant function. If these covariates are correlated with the discriminators, they may serve in some way to "improve" the discriminant: e.g. to increase the power of Hotelling's T^2-test, or to reduce the number of errors in classification. The problem is analogous to that which is solved by the analysis of covariance. In covariance, as applied for instance in a controlled experiment, variates that are unaffected by the experimental treatments can be used to provide more accurate estimates of the effects of the treatments or to increase the power of the

F-test of the differences among treatment means.

The procedure suggested by Cochran and Bliss was to obtain the multiple regression of each discriminator on all of the covariates. These regressions were calculated from the within sample sums of squares and products. Then they replaced each discriminator by its deviations from multiple regression and calculated the discriminant function using these deviations. An attempt was also made to obtain a quantity that will measure what has been gained by the use of covariance.

Cochran (1964a) was concerned with the question of what happens if the covariates are simply included in the discriminant function in exactly the same way as the discriminators. If the covariates are treated in the same manner as the discriminators, one avoids the necessity of learning the Cochran-Bliss technique for handling covariates. Cochran found that for tests of significance the covariance technique is more powerful. For classifying observations into one of two populations, it appears that the gain from the covariance technique is trivial provided that the discriminant function is constructed from a reasonably large sample.

Rao (1949, 1966) was concerned with a number of problems arising out of the discrimination between two populations when it is known that they do not differ with respect to a given subset of characters (covariates). Rao stated that in practice if the correlations between the discriminators and covariates are unknown and have to be estimated from the data, there may, indeed be loss of information, unless the correlations are high. Therefore, some caution is needed in the choice and use of covariates.

Rao (1966) was primarily interested in various tests of hypotheses. He considered the case when one has a (p+q)-dimensional random variable $(\underline{Y},\underline{Z})$ where $\underline{Y}$ is p-dimensional and $\underline{Z}$ is q-dimensional. The variable $\underline{Y}$ was the main variable (discriminator) and $\underline{Z}$ was the covariate. Let E_1 and E_2 denote expectations with respect to two (p+q)-variate normal populations with the same covariance matrix and possible different mean vectors. Some hypotheses of interest were:

(a) $E_1(\underline{Y}) = E_2(\underline{Y})$, $E_1(\underline{Z}) = E_2(\underline{Z})$,

(b) $E_1(\underline{Y}|\underline{Z}) = E_2(\underline{Y}|\underline{Z})$,

(c) $E_1(\underline{Y}|\underline{Z}) = E_2(\underline{Y}|\underline{Z})$ given $E_1(\underline{Z}) = E_2(\underline{Z})$,

(d) $E_1(\underline{Y}) = E_2(\underline{Y})$ given $E_1(\underline{Z}) = E_2(\underline{Z})$,

(e) $E_1(\underline{Y}) = E_2(\underline{Y})$. (1.12)

The interpretation of these hypotheses and their tests based on independent samples of size N_1 and N_2 from the two distributions were examined.

Subrahmaniam (1970) developed a unified approach to the theory of discriminant analysis in the presence of covariates by partitioning the D^2-statistic. She demonstrated the equivalence of the methods of Rao (1949) and Cochran and Bliss (1948). A study was made of the null and non-null distributions of the test statistic, which is a function of the D^2-statistic, for the five hypotheses given in (1.12). Although Cochran and Bliss were essentially interested in this same problem, namely, adjusting the distance function of the main variables (discriminators) for the effect of the co-variates, their development parallels that of a multiple regression technique. Cochran and Bliss made the further assumption that the covariates have the same population means. Imposing this restriction, the hypothesis under test becomes

$$E_1(\underline{Y}|\underline{Z}) = E_2(\underline{Y}|\underline{Z}) \quad \text{given} \quad E_1(\underline{Z}) = E_2(\underline{Z}) \ .$$

Memon and Okamoto (1970) also considered the problem of discriminant analysis with covariance. Let

$$\begin{pmatrix} \underline{X}_1^{(i)} \\ \underline{Y}_1^{(i)} \end{pmatrix}, \quad \begin{pmatrix} \underline{X}_2^{(i)} \\ \underline{Y}_2^{(i)} \end{pmatrix}, \ldots, \begin{pmatrix} \underline{X}_{N_i}^{(i)} \\ \underline{Y}_{N_i}^{(i)} \end{pmatrix}$$

be two random samples drawn independently from

$$\pi_i: N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(i)} \\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}\right),$$

where $\underline{X}_\alpha^{(i)}$ is a (pxl) discriminator, $\underline{Y}_\alpha^{(i)}$ is a (qxl) covariate, i=1,2; α=1,2,...,N_i; and the covariance matrix

$$\underline{\Sigma} = \begin{bmatrix}\underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22}\end{bmatrix}$$

is positive-definite. Let $\hat{\underline{B}}$ be the sample estimate of the regression matrix $\underline{B}$ of the discriminator $\underline{X}$ on the covariate $\underline{Y}$. Define $\underline{X}^* = \underline{X} - \hat{\underline{B}}\,\underline{Y}$ and replace $\underline{X}$ by $\underline{X}^*$ in the Anderson discriminant function W given in (1.10). The modified criterion is given by

$$W^* = [\underline{X}^* - \tfrac{1}{2}(\overline{\underline{X}}^{*(1)} + \overline{\underline{X}}^{*(2)})]'(\underline{S}_{11} - \underline{S}_{12}\underline{S}_{22}^{-1}\underline{S}_{21})^{-1}\,[\overline{\underline{X}}^{*(1)} - \overline{\underline{X}}^{*(2)}], \tag{1.13}$$

where $\overline{\underline{X}}^{*(i)} = \overline{\underline{X}}^{(i)} - \hat{\underline{B}}\,\overline{\underline{Y}}^{(i)}$, i=1,2; $\hat{\underline{B}} = \underline{S}_{12}\underline{S}_{22}^{-1}$, $\overline{\underline{X}}^{(i)}$ and $\overline{\underline{Y}}^{(i)}$ denote the sample means and finally

$$\underline{S} = \begin{bmatrix}\underline{S}_{11} & \underline{S}_{12} \\ \underline{S}_{21} & \underline{S}_{22}\end{bmatrix}$$

is the best unbiased estimator of $\underline{\Sigma}$. Memon and Okamoto obtained the asymptotic expansion of the distribution of W*. They compared the three classification procedures (i), (ii), (iii), where one utilizes information on

(i) $\underline{X}$ only,

or

(ii) $\begin{pmatrix} \underline{X} \\ \underline{Y} \end{pmatrix}$,

or

(iii) $\underline{X}^*$,

using the minimization of the probabilities of misclassification as their criterion for obtaining the best procedure. They found that procedure (iii) is preferable to procedure (ii), while its superiority over (i) depends on the balance of the increased Mahalanobis' distance and the increased effective dimensionality, which they did not consider.

3. Quadratic discriminant function

Smith (1947) used the likelihood ratio procedure proposed by Welch (1939) to derive the discriminant function when there are two populations with different covariance matrices. He carried out the necessary computations for the bivariate case. He also compared the probabilities of misclassification between the linear discriminant function and the quadratic discriminant function for the bivariate case on two sets of data.

Okamoto (1961) has worked on the problem of discrimination between two populations with common mean vectors and different covariance matrices. He derived an asymptotic expansion of the first approximation to the quadratic discriminant function up to the linear term for the special case where the mean vector was known.

Bartlett and Please (1963) obtained the discriminant function when the mean differences between the two populations are all zero and the covariance matrices $\underline{\Sigma}^{(i)}$ are of the form

$$\underline{\Sigma}^{(i)} = \sigma_i^2 [(1-\rho_i)\underline{I} + \rho_i \underline{J}], \quad i=1,2,$$

i.e., they have intraclass correlation structure. Han (1968) extended the Bartlett-Please case and derived the discriminant function when the mean differences between the two populations is not zero.

Anderson and Bahadur (1962) treated the classification problem by determining the admissible linear procedures for classifying an observation as coming from one of two multivariate normal populations in the case that the two distributions differ both in mean vectors and covariance matrices. They assumed all parameters known.

In the case where the mean vectors are known and the covariance matrices $\underline{\Sigma}^{(i)}$ are general and known, finding the distribution of the discriminant function is often equivalent to finding the distribution of

$$V = A - B, \tag{1.14}$$

where A and B are given by

$$\begin{aligned} A &= \alpha[\chi'^2_{m_0}(d_0) + \sum_{i=1}^{r} a_i \chi'^2_{m_i}(d_i)] , \\ B &= \beta[\chi'^2_{n_o}(g_0) + \sum_{j=1}^{s} b_j \chi'^2_{n_j}(g_j)] , \end{aligned} \tag{1.15}$$

where $\alpha>0$, $\beta>0$, $a_i \geq 1$, $b_j \geq 1$, $d_i \geq 0$, $d_0 \geq 0$, $g_j \geq 0$, $g_0 \geq 0$, for all i and j, and where $\chi_m'^2(\lambda)$ denotes a non-central chi-square variate with m degrees of freedom and with non-centrality parameter λ, and all chi-square variates are independent. Gurland (1955) has obtained the distribution of an indefinite quadratic form of central chi-square variates using a complicated infinite series expansion involving Laguerre polynomials for the case in which the number of positive (or negative) coefficients is even. Shah (1963) extended his work to the non-central case.

Press (1966) obtained an explicit form for the distribution of V in (1.14). He has also obtained several asymptotic procedures for the case of unknown population parameters, assuming unequal covariance matrices whose difference is positive-definite. When the covariance matrices are unequal with intraclass correlation structure, Press (1964) has applied the method of classification developed by Anderson and Bahadur. Several of the situations developed by Press will be made more explicit in the following chapters.

Han (1969) has derived the discriminant function between two multivariate normal populations with mean vectors $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$, covariance matrices $\underline{\Sigma}$ and $\sigma^2\underline{\Sigma}$ $(\sigma^2>1)$, respectively. Hence the covariance matrix of the second population is a constant multiplier of the covariance matrix of the first population. He assumed that $\underline{\Sigma}$ and σ^2 are known; $\underline{\mu}^{(1)}$ and

$\underline{\mu}^{(2)}$ are known or unknown. When $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$ are known, he obtained the distribution of the discriminant function. When $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$ are unknown, an asymptotic expansion of the distribution of the discriminant function was obtained.

Han (1970) has also obtained the discriminant function when the two populations are multivariate normal with unequal circular covariance matrices. The discriminant function and its distribution were derived when all parameters are known. When the covariance matrices are known but the means are unknown, Han obtained the discriminant function and its distribution. An asymptotic expansion for the distribution of the discrimirant function was obtained when all parameters are unknown.

Gilbert (1969) compared Fisher's linear discriminant function and the quadratic discriminant function with respect to probabilities of misclassification when one covariance matrix is a constant multiple of the other. She assumed known population parameters. She used a chi-square distribution to approximate probabilities of misclassification for the quadratic discriminant function. Gilbert employed a wide range of parameter values and suggested situations when Fisher's linear discriminant function compared favorably with respect to the quadratic discriminant function.

Marks (1970) compared the performance of three discriminant functions in classifying individuals into two multivariate normally distributed populations when covariance matrices are unequal. The functions compared were Fisher's linear discriminant function, the Anderson-Bahadur best linear discriminant function, and the quadratic discriminant function. The expected probability of misclassification, based on known a priori probabilities, was used as a measure of performance. Monte Carlo methods were used. Parameters that were varied in the study included the relative position of the population means, the covariance matrices, number of dimensions, sample sizes, and a priori probabilities of origin from the populations. Recommendations are given, based on this study, for choosing the discriminant function which has the smallest expected probability of misclassification in the various situations.

II. DISTRIBUTION OF THE DISCRIMINANT FUNCTION UNDER INTRACLASS CORRELATION MODELS

A. Introduction

A (pxl) observation vector $\underline{X}$ is to be classified into one of two populations. Using the likelihood ratio procedure, we could assign the observation to π_1 if

$$\frac{p_1(\underline{X})}{p_2(\underline{X})} \geq k \quad \text{and to} \quad \pi_2 \text{ if } \quad \frac{p_1(\underline{X})}{p_2(\underline{X})} < k,$$

where k is a constant and $p_i(\underline{X})$ is proportional to the probability of $\underline{X}$ if the individual comes from π_i, i=1,2. We assume that $\underline{X} \sim N_p(\underline{\mu}^{(i)}, \underline{\Sigma}^{(i)})$ when the observation belongs to π_i, where $\underline{\mu}^{(i)}$ is a (pxl) vector and $\underline{\Sigma}^{(i)}$ is given by

$$\underline{\Sigma}^{(i)} = \sigma_i^2[(1-\rho_i)\underline{I} + \rho_i\underline{J}], \quad i=1,2, \tag{2.1}$$

where $\underline{I}$ is the (pxp) identity matrix, $\underline{J}$ is a (pxp) matrix of ones, and $\underline{\Sigma}^{(i)}$ is positive-definite for i=1,2. We shall assume $\sigma_1^2 \neq \sigma_2^2$. Hence, $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$ are unequal covariance matrices both having intraclass correlation structure.

Bartlett and Please (1963) considered this problem when the populations are π_i: $N_p(\underline{\mu}^{(i)}, \underline{\Sigma}^{(i)})$. They derived the discriminant function when $\underline{\mu}^{(1)} - \underline{\mu}^{(2)} = \underline{0}$ and $\underline{\Sigma}^{(i)}$ have the form (2.1). Han (1968) extended their work and derived the discriminant function when the mean difference is not zero.

Following Bartlett and Please, suppose $\underline{\Sigma}^{(1)}$ is standardized, then $\sigma_1^2=1$ and $\sigma_2^2=\sigma^2$ say. If $\underline{\mu}^{(1)}-\underline{\mu}^{(2)}=\underline{0}$ where $\underline{\mu}^{(1)}$, $\underline{\mu}^{(2)}$ are known, the logarithm of the likelihood ratio is proportional to

$$\underline{X}'[\underline{\Sigma}^{(1)}]^{-1}\underline{X} - \underline{X}'[\underline{\Sigma}^{(2)}]^{-1}\underline{X}. \qquad (2.2)$$

Since

$$[\underline{\Sigma}^{(1)}]^{-1} = \frac{1}{1-\rho_1}[\underline{I} - \frac{\rho_1}{1+(p-1)\rho_1}\underline{J}]$$

and

$$[\underline{\Sigma}^{(2)}]^{-1} = \frac{1}{\sigma^2(1-\rho_2)}[\underline{I} - \frac{\rho_2}{1+(p-1)\rho_2}\underline{J}],$$

(2.2) can be written as

$$[\frac{1}{1-\rho_1} - \frac{1}{\sigma^2(1-\rho_2)}]Z_1-[\frac{\rho_1}{1-\rho_1}\frac{1}{1+(p-1)\rho_1} - \frac{\rho_2}{\sigma^2(1-\rho_2)}\frac{1}{1+(p-1)\rho_2}]Z_2, \qquad (2.3)$$

where $Z_1 = \sum_{i=1}^{p} X_i^2$ and $Z_2 = (\sum_{i=1}^{p} X_i)^2$. Penrose (1947) called Z_2 the square of the "size" component, as $\sum_{i=1}^{p} X_i$ measures "total size." If

$$a = \frac{1}{1-\rho_1} - \frac{1}{\sigma^2(1-\rho_2)},$$

$$b = \frac{\rho_1}{1-\rho_1}\frac{1}{1+(p-1)\rho_1} - \frac{\rho_2}{\sigma^2(1-\rho_2)}\frac{1}{1+(p-1)\rho_2},$$

then the likelihood method leads to

$$aZ_1 - bZ_2 = c \qquad (2.4)$$

as the boundary separating the regions of misclassification

R_1 and R_2. The constant c in (2.4) is usually chosen so that the probabilities of misclassification are equal.

In the case of the further assumption $\rho_1 = \rho_2 = \rho$, (2.4) reduces to

$$Z_1 - \frac{\rho}{1+(p-1)\rho} Z_2 = c' \ . \tag{2.5}$$

Bartlett and Please (1963) suggested that the boundary (2.5) (or (2.4) if $\rho_1 \neq \rho_2$) is probably better fitted visually when only samples from the two populations are available and the parameters are unknown. Bartlett and Please give an interesting biological example of discriminating between monozygotic and dizygotic pairs of twins (with like sex) using the above procedure.

If the mean difference is not zero, using the likelihood ratio procedure, we obtain the quadratic discriminant function

$$U = (\underline{X}-\underline{\mu}^{(2)})'[\underline{\Sigma}^{(2)}]^{-1}(\underline{X}-\underline{\mu}^{(2)})-(\underline{X}-\underline{\mu}^{(1)})'[\underline{\Sigma}^{(1)}]^{-1}(\underline{X}-\underline{\mu}^{(1)}). \tag{2.6}$$

Press (1966) derived the distribution of U when all parameters are known and $\underline{\Sigma}^{(1)}-\underline{\Sigma}^{(2)}$ is positive-definite.

In Section B, we give the distribution of the discriminant function when all parameters are known. Section C obtains the distribution of the discriminant function when the mean vectors $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$, equal or unequal, are unknown, but the covariance matrices are known. In Section D, we derive an asymptotic expansion for the distribution of the discriminant function when all parameters are unknown.

B. Parameters Known

Let $\underline{X}$ be a (pxl) observation vector such that $\underline{X} \sim N_p(\underline{\mu}^{(i)}, \underline{\Sigma}^{(i)})$ when $\underline{X}$ belongs to π_i, i=1,2. We assume in this section that all parameters are known.

Let $\underline{\Gamma}$ be a (pxp) orthogonal matrix whose first row is $p^{-1/2}\underline{e}'$ where $\underline{e}$ is a (pxl) vector of ones. Then,

$$\underline{\Gamma}\ \underline{\Sigma}^{(1)}\underline{\Gamma}' = \underline{D}^{(1)}, \quad \underline{\Gamma}\ \underline{\Sigma}^{(2)}\underline{\Gamma}' = \underline{D}^{(2)}, \tag{2.7}$$

where $\underline{D}^{(i)}$ is a (pxp) diagonal matrix, which we write as $\underline{D}^{(i)} = \mathrm{diag}(\alpha_i, \beta_i, \ldots, \beta_i)$, and where

$$\alpha_i = \sigma_i^2\,[1+(p-1)\rho_i],$$

$$\beta_i = \sigma_i^2\,(1-\rho_i),\ i=1,2 \tag{2.8}$$

(see, for example, Press, 1972, p. 48). Since $\underline{\Gamma}$ is independent of the elements in $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$, the discriminant function is equivalent to that when the covariance matrices are diagonal. This is true because the discriminant function given in (2.6) is invariant under any orthogonal transformation. To see this, let $\underline{Y} = \underline{\Gamma}\ \underline{X}$ so $\underline{Y} \sim N_p(\underline{\nu}^{(i)}, \underline{D}^{(i)})$ when $\underline{X}$ belongs to π_i, where $\underline{\nu}^{(i)} = \underline{\Gamma}\underline{\mu}^{(i)}$, $\underline{D}^{(i)} = \mathrm{diag}\ (\alpha_i, \beta_i, \ldots, \beta_i)$, i=1,2. It follows from (2.6) that the discriminant function in terms of $\underline{Y}$ is

$$(\underline{Y}-\underline{\nu}^{(2)})'[\underline{D}^{(2)}]^{-1}(\underline{Y}-\underline{\nu}^{(2)})-(\underline{Y}-\underline{\nu}^{(1)})'[\underline{D}^{(1)}]^{-1}(\underline{Y}-\underline{\nu}^{(1)})$$

$$= [\underline{\Gamma}\ \underline{X}-\underline{\Gamma}\ \underline{\mu}^{(2)})'[\underline{\Gamma}\ \underline{\Sigma}^{(2)}\underline{\Gamma}']^{-1}(\underline{\Gamma}\ \underline{X}-\underline{\Gamma}\ \underline{\mu}^{(2)})$$

$$-(\underline{\Gamma}\ \underline{X}-\underline{\Gamma}\ \underline{\mu}^{(1)})'[\underline{\Gamma}\ \underline{\Sigma}^{(1)}\underline{\Gamma}']^{-1}(\underline{\Gamma}\ \underline{X}-\underline{\Gamma}\ \underline{\mu}^{(1)})$$

$$= (\underline{X}-\underline{\mu}^{(2)})'\underline{\Gamma}'(\underline{\Gamma}')^{-1}[\underline{\Sigma}^{(2)}]^{-1}\underline{\Gamma}^{-1}\underline{\Gamma}(\underline{X}-\underline{\mu}^{(2)})$$

$$-(\underline{X}-\underline{\mu}^{(1)})'\underline{\Gamma}'(\underline{\Gamma}')^{-1}[\underline{\Sigma}^{(1)}]^{-1}\underline{\Gamma}^{-1}\underline{\Gamma}(\underline{X}-\underline{\mu}^{(1)})$$

$$= (\underline{X}-\underline{\mu}^{(2)})'[\underline{\Sigma}^{(2)}]^{-1}(\underline{X}-\underline{\mu}^{(2)})-(\underline{X}-\underline{\mu}^{(1)})'[\underline{\Sigma}^{(1)}]^{-1}(\underline{X}-\underline{\mu}^{(1)})$$

$$= U.$$

Hence the discriminant function can be written as

$$U = (\underline{Y}-\underline{\nu}^{(2)})'[\underline{D}^{(2)}]^{-1}(\underline{Y}-\underline{\nu}^{(2)})-(\underline{Y}-\underline{\nu}^{(1)})'[\underline{D}^{(1)}]^{-1}(\underline{Y}-\underline{\nu}^{(1)}). \quad (2.9)$$

Substituting $\underline{\nu}^{(i)}$ and $\underline{D}^{(i)}$ we obtain, apart from a constant,

$$V= \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\left(Y_1- \frac{\frac{\nu_1^{(2)}}{\alpha_2} - \frac{\nu_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right)^2$$

$$+\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\sum_{j=2}^{p}\left(Y_j- \frac{\frac{\nu_j^{(2)}}{\beta_2} - \frac{\nu_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right)^2, \quad (2.10)$$

where Y_j and $\nu_j^{(i)}$ are the jth components of $\underline{Y}$ and $\underline{\nu}^{(i)}$, i=1,2, respectively. We shall classify $\underline{X}$ into π_1 if $V \geq c$

and into π_2 if $V<c$ for some suitable choice of the constant c.

To find the distribution of V, we shall assume that $\alpha_1>\alpha_2$, $\beta_1>\beta_2$, or equivalently that $\underline{\Sigma}^{(1)}-\underline{\Sigma}^{(2)}$ is positive-definite. Hence $\frac{1}{\alpha_2} - \frac{1}{\alpha_1} > 0$ and $\frac{1}{\beta_2} - \frac{1}{\beta_1} > 0$. Let

$$Z_1 = \sqrt{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\,\left(Y_1 - \frac{\dfrac{\nu_1^{(2)}}{\alpha_2} - \dfrac{\nu_1^{(1)}}{\alpha_1}}{\dfrac{1}{\alpha_2} - \dfrac{1}{\alpha_1}}\right),$$

$$Z_j = \sqrt{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\,\left(Y_j - \frac{\dfrac{\nu_j^{(2)}}{\beta_2} - \dfrac{\nu_j^{(1)}}{\beta_1}}{\dfrac{1}{\beta_2} - \dfrac{1}{\beta_1}}\right), \quad j=2,\ldots,p. \tag{2.11}$$

Then $V = \sum_{j=1}^{p} Z_j^2$. When $\underline{X}$ comes from π_i, $i=1$ or 2, Z_j are independently distributed as $N(\xi_j^{(i)}, \tau_{ij}^2)$ where

$$\xi_1^{(i)} = \sqrt{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\,\left(\nu_1^{(i)} - \frac{\dfrac{\nu_1^{(2)}}{\alpha_2} - \dfrac{\nu_1^{(1)}}{\alpha_1}}{\dfrac{1}{\alpha_2} - \dfrac{1}{\alpha_1}}\right),$$

$$\tau_{i1}^2 = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\alpha_i \quad \text{for } i=1,2;$$

$$\xi_j^{(i)} = \sqrt{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\,\left(\nu_j^{(i)} - \frac{\dfrac{\nu_j^{(2)}}{\beta_2} - \dfrac{\nu_j^{(1)}}{\beta_1}}{\dfrac{1}{\beta_2} - \dfrac{1}{\beta_1}}\right),$$

$$\tau_{ij}^2 = (\frac{1}{\beta_2} - \frac{1}{\beta_1})\beta_i \quad \text{for} \quad j=2,\ldots,p; \quad i=1,2. \qquad (2.12)$$

Therefore, V is distributed as the sum of $\tau_{ij}^2 \chi_1'^2(\delta_{ij}^2)$, which we write as $V \sim \sum_{j=1}^{p} \tau_{ij}^2 \chi_1'^2(\delta_{ij}^2)$ for $i=1,2$, where $\chi_1'^2(\delta_{ij}^2)$ denotes a non-central chi-square distribution with 1 degree of freedom and non-centrality parameter

$$\delta_{ij}^2 = \frac{[\xi_j^{(i)}]^2}{\tau_{ij}^2}.$$

It is not easy, in general, to obtain the distribution of V in a closed form.

Several special cases are of interest and will be considered separately in the following:

1. Case (i): $\underline{\nu}^{(1)} = \underline{\nu}^{(2)}$, $\rho_1 = \rho_2$.

Suppose that in addition to the above assumptions concerning the parameters, we assume $\underline{\nu}^{(1)} = \underline{\nu}^{(2)} = \underline{\nu}$ and $\rho_1 = \rho_2 = \rho$. Therefore, V becomes

$$V = [\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}]\{\frac{1}{1+(p-1)\rho}(Y_1-\nu_1)^2 + \frac{1}{1-\rho}\sum_{j=2}^{p}(Y_j-\nu_j)^2\}, \qquad (2.13)$$

where ν_j is the jth component of $\underline{\nu}$ and (2.12) becomes

$$\xi_j^{(i)} = 0,$$

$$\tau_{ij}^2 = \left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2\sigma_2^2}\right)\sigma_i^2, \quad j=1,2,\ldots,p; \quad i=1,2. \tag{2.14}$$

Hence $V \sim \sum_{j=1}^{p} \tau_{ij}^2 \chi_1^2$, where χ_1^2 denotes a central chi-square distribution with 1 degree of freedom. Therefore,

$$V \sim \frac{\sigma_1^2 - \sigma_2^2}{\sigma_2^2} \chi_p^2, \text{ if } \underline{X} \text{ belongs to } \pi_1, \tag{2.15}$$

and

$$V \sim \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2} \chi_p^2, \text{ if } \underline{X} \text{ belongs to } \pi_2. \tag{2.16}$$

2. Case (ii): $\underline{\nu}^{(1)} = \underline{\nu}^{(2)}, \quad \rho_1 \neq \rho_2$.

If we drop the assumption $\rho_1=\rho_2$ in case (i) but keep the remaining assumptions concerning the parameters, we obtain V in the form

$$V = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)(Y_1-\nu_1)^2 + \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\sum_{j=2}^{p}(Y_j-\nu_j)^2, \tag{2.17}$$

where ν_j is the jth component of $\underline{\nu} = \underline{\nu}^{(1)} = \underline{\nu}^{(2)}$. The equations in (2.12) become

$$\xi_j^{(i)} = 0, \quad j=1,\ldots,p; \quad i=1,2;$$

$$\tau_{i1}^2 = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\alpha_i,$$

$$\tau_{ij}^2 = \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\beta_i, \quad j=2,\ldots,p; \quad i=1,2. \tag{2.18}$$

Therefore,

$$V \sim \frac{\alpha_1-\alpha_2}{\alpha_2}\chi_1^2 + \frac{\beta_1-\beta_2}{\beta_2}\chi_{p-1}^2, \quad \text{if } \underline{X} \text{ belongs to } \pi_1, \tag{2.19}$$

and

$$V \sim \frac{\alpha_1-\alpha_2}{\alpha_1}\chi_1^2 + \frac{\beta_1-\beta_2}{\beta_1}\chi_{p-1}^2, \quad \text{if } \underline{X} \text{ belongs to } \pi_2. \tag{2.20}$$

3. Case (iii): $\underline{\nu}^{(1)} \neq \underline{\nu}^{(2)}$, $\rho_1 = \rho_2$.

If we drop the assumption $\underline{\nu}^{(1)}=\underline{\nu}^{(2)}$ in Case (i), but keep the remaining assumptions concerning the parameters including $\rho_1=\rho_2 = \rho$, we obtain V in the form

$$V = [\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}]\{\frac{1}{1+(p-1)\rho}(Y_1 - \frac{\frac{\nu_1^{(2)}}{\sigma_2^2} - \frac{\nu_1^{(1)}}{\sigma_1^2}}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}})^2$$

$$+ \frac{1}{1-\rho}\sum_{j=2}^{p}(Y_j - \frac{\frac{\nu_j^{(2)}}{\sigma_2^2} - \frac{\nu_j^{(1)}}{\sigma_1^2}}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}})^2\}. \tag{2.21}$$

The equations in (2.12) become

$$\xi_1^{(1)} = \sqrt{\frac{\sigma_1^2}{\sigma_2^2(1+(p-1)\rho)(\sigma_1^2 - \sigma_2^2)}}\,(\nu_1^{(1)}-\nu_1^{(2)}),$$

$$\xi_1^{(2)} = \sqrt{\frac{\sigma_2^2}{\sigma_1^2(1+(p-1)\rho)(\sigma_1^2 - \sigma_2^2)}}\,(\nu_1^{(1)}-\nu_1^{(2)}),$$

$$\xi_j^{(1)} = \sqrt{\frac{\sigma_1^2}{\sigma_2^2(1-\rho)(\sigma_1^2-\sigma_2^2)}}\,(\nu_j^{(1)}-\nu_j^{(2)}),$$

$$\xi_j^{(2)} = \sqrt{\frac{\sigma_2^2}{\sigma_1^2(1-\rho)(\sigma_1^2-\sigma_2^2)}}\,(\nu_j^{(1)}-\nu_j^{(2)}), \quad j=2,\ldots,p;$$

$$\tau_{ij}^2 = \left(\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2\sigma_2^2}\right)\sigma_i^2, \quad j=1,\ldots,p; \quad i=1,2. \tag{2.22}$$

Therefore,

$$V \sim \frac{\sigma_1^2-\sigma_2^2}{\sigma_2^2}\{\chi_p'^2[\sum_{j=1}^{p}(\delta_{ij}^2)]\}, \text{ if } \underline{X} \text{ belongs to } \pi_1, \tag{2.23}$$

and

$$V \sim \frac{\sigma_1^2-\sigma_2^2}{\sigma_{\underline{i}}^2}\{\chi_p'^2[\sum_{j=1}^{p}(\delta_{ij}^2)]\}, \text{ if } \underline{X} \text{ belongs to } \pi_2, \tag{2.24}$$

where

$$\delta_{ij}^2 = \frac{[\xi_j^{(i)}]^2}{\tau_{ij}^2} .$$

4. Case (iv): $\underline{\nu}^{(1)} \neq \underline{\nu}^{(2)}$, $\rho_1 \neq \rho_2$.

Suppose we have the assumptions concerning the parameters given at the beginning of Section B. In addition, assume $\underline{\nu}^{(1)} \neq \underline{\nu}^{(2)}$ and $\rho_1 \neq \rho_2$. From (2.12) we see that

$$V \sim \{(\frac{\alpha_1-\alpha_2}{\alpha_2})\chi_1'^2(\delta_{11}^2) + (\frac{\beta_1-\beta_2}{\beta_2})\chi_{p-1}'^2[\sum_{j=2}^{p}(\delta_{1j}^2)]\},$$

if $\underline{X}$ belongs to π_1, (2.25)

and

$$V \sim \{(\frac{\alpha_1-\alpha_2}{\alpha_1})\chi_1'^2(\delta_{21}^2) + (\frac{\beta_1-\beta_2}{\beta_1})\chi_{p-1}'^2[\sum_{j=2}^{p}(\delta_{2j}^2)]\},$$

if $\underline{X}$ belongs to π_2. (2.26)

Both (2.25) and (2.26) can be written in the form

$$A = \alpha[\chi_{m_0}'^2(d_0) + \sum_{i=1}^{r} a_i \chi_{m_i}'^2(d_i)], \tag{2.27}$$

where A is a positive-definite quadratic form with $\alpha>0$, $a_i \geq 1$, $d_0 \geq 0$, $d_i \geq 0$ for all i, and all chi-square variates are independent. Specifically, if

$$\alpha = \frac{\beta_1-\beta_2}{\beta_2},$$

$$a_1 = (\frac{\beta_2}{\alpha_2})(\frac{\alpha_1-\alpha_2}{\beta_1-\beta_2}), \tag{2.28}$$

then (2.25) can be written as

$$\alpha\{\chi_{p-1}'^2[\sum_{j=2}^{p}(\delta_{ij}^2)] + a_1\chi_1'^2(\delta_{11}^2)\}, \tag{2.29}$$

if the parameters are labeled so that $\alpha_1/\alpha_2 > \beta_1/\beta_2$. Similarly, (2.26) could be written in the form (2.27).

Press (1966) has shown that if F(x) is the c.d.f. of A in (2.27) and $F_\nu(x)$ denotes the c.d.f. of a central chi-square variate with ν degrees of freedom, then one possible representation for F(x) is

$$F(x) = \sum_{i=0}^{\infty} q_i F_{M+2i}(x/\alpha), \tag{2.30}$$

where

$$M = \sum_{i=0}^{r} m_i, \; q_i > 0, \; \sum_{i=0}^{\infty} q_i = 1, \text{ and the } q_i \text{ are constants}$$

depending on m_i, d_i, a_i. Hence the c.d.f. of V, (2.25) or (2.26), can be obtained by this method. It is perhaps important to note that V is expressible as the sum of two, instead of p, chi-squared variates.

Patnaik (1949) has considered a chi-square approximation to the distribution of (2.25) and (2.26) by fitting the first two moments. Pearson (1959) suggested an improvement of the chi-square approximation to the distribution of this sum by fitting the first three moments. These approximations and their applications will be discussed more fully in Chapter V.

C. Means Unknown, Covariance Matrices Known

In this section, we shall assume that $\underline{\mu}^{(1)}$ and $\underline{\mu}^{(2)}$ are unknown, but $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$ are known. Suppose that we have a random sample $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \ldots, \underline{X}_{N_1}^{(1)}$ from π_1 and an independent random sample $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \ldots, \underline{X}_{N_2}^{(2)}$ from π_2. Clearly, our estimate of $\underline{\mu}^{(i)}$ is the sample mean

$\underline{\bar{X}}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \underline{X}_{\alpha}^{(i)}$, $i=1,2$. The sample means $\underline{\bar{X}}^{(i)} \sim N_p(\underline{\mu}^{(i)}, \frac{1}{N_i} \underline{\Sigma}^{(i)})$, $i=1,2$. If $\underline{Y} = \underline{\Gamma}\, \underline{X}$, where $\underline{\Gamma}$ is the (pxp) orthogonal matrix given in Section B, then we estimate $\underline{\nu}^{(i)} = \underline{\Gamma}\, \underline{\mu}^{(i)}$ by $\underline{\bar{Y}}^{(i)} = \underline{\Gamma}\, \underline{\bar{X}}^{(i)}$, $i=1,2$. The sample means $\underline{\bar{Y}}^{(i)}$ are independently distributed as $N_p(\underline{\nu}^{(i)}, \frac{1}{N_i} \underline{D}^{(i)})$, $i=1,2$, and independent of the distribution of $\underline{Y}$. Further, since the $\underline{D}^{(i)}$ are diagonal matrices, the components of $\underline{\bar{Y}}^{(i)}$ are also independently distributed. The discriminant function V in (2.10), after substituting the unknown parameters by estimates, becomes

$$V = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\left(Y_1 - \frac{\frac{\bar{Y}_1^{(2)}}{\alpha_2} - \frac{\bar{Y}_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right)^2$$

$$+ \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right) \sum_{j=2}^{p} \left(Y_j - \frac{\frac{\bar{Y}_j^{(2)}}{\beta_2} - \frac{\bar{Y}_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right)^2, \tag{2.31}$$

where Y_j and $\bar{Y}_j^{(i)}$ are the jth components of $\underline{Y}$ and $\underline{\bar{Y}}^{(i)}$, $i=1,2$, respectively. Let

$$W_1 = \sqrt{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\left(Y_1 - \frac{\frac{\bar{Y}_1^{(2)}}{\alpha_2} - \frac{\bar{Y}_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right),$$

$$W_j = \sqrt{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\,\left(Y_j - \frac{\frac{\overline{Y}_j^{(2)}}{\beta_2} - \frac{\overline{Y}_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right),\ j=2,3,\ldots,p. \qquad (2.32)$$

Then

$$V = \sum_{j=1}^{p} W_j^2 .$$

The distribution of V given in (2.31) can be found in a manner similar to that used in Section B. When $\underline{X}$ comes from π_i, i=1 or 2, then W_j are independently distributed as $N(\xi_j^{(i)}, \tau_{ij}^{*2})$, j=1,2,...,p, where $\xi_j^{(i)}$ are given in (2.12) and

$$\tau_{i1}^{*2} = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\alpha_i + \frac{N_1\alpha_1+N_2\alpha_2}{N_1N_2(\alpha_1-\alpha_2)},$$

$$\tau_{ij}^{*2} = \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\beta_i + \frac{N_1\beta_1+N_2\beta_2}{N_1N_2(\beta_1-\beta_2)},\ j=2,\ldots,p;\quad i=1,2. \qquad (2.33)$$

The second term on the right side of τ_{ij}^{*2} is the increase in variance accounted for by the unknown means. Therefore,

$$V \sim \{\tau_{i1}^{*2}\,\chi_1'^2(\delta_{i1}^{*2}) + \tau_{ij}^{*2}\,\chi_{p-1}'^2[\sum_{j=2}^{p}(\delta_{ij}^{*2})]\}, \qquad (2.34)$$

if $\underline{X}$ comes from π_i, i=1,2, where

$$\delta_{ij}^{*2} = \frac{[\xi_j^{(i)}]^2}{\tau_{ij}^{*2}},$$ j=1,2,...,p, and where $\xi_j^{(i)}$ and τ_{ij}^{*2} are given in (2.12) and (2.33), respectively.

Consider the special case in which the means $\underline{\nu}^{(1)}$ and $\underline{\nu}^{(2)}$ are equal but unknown. Let $\underline{\nu}$ denote their common value. The maximum likelihood estimate of $\underline{\nu}$ is

$$\underline{\bar{Y}} = (N_1[\underline{D}^{(1)}]^{-1}+N_2[\underline{D}^{(2)}]^{-1})^{-1}(N_1[\underline{D}^{(1)}]^{-1}\underline{\bar{Y}}^{(1)}+N_2[\underline{D}^{(2)}]^{-1}\underline{\bar{Y}}^{(2)}).$$

On simplifying we obtain

$$\underline{\bar{Y}} = (\frac{\alpha_1\alpha_2}{N_1\alpha_2+N_2\alpha_1}[\frac{N_1}{\alpha_1}\bar{Y}_1^{(1)} + \frac{N_2}{\alpha_2}\bar{Y}_1^{(2)}],$$

$$\frac{\beta_1\beta_2}{N_1\beta_2+N_2\beta_1}[\frac{N_1}{\beta_1}\bar{Y}_2^{(1)} + \frac{N_2}{\beta_2}\bar{Y}_2^{(2)}] \;,\ldots,$$

$$\frac{\beta_1\beta_2}{N_1\beta_2+N_2\beta_1}[\frac{N_1}{\beta_1}\bar{Y}_p^{(1)} + \frac{N_2}{\beta_2}\bar{Y}_p^{(2)}])' .$$

Hence V given in (2.31) is of the form

$$V = (\frac{1}{\alpha_2} - \frac{1}{\alpha_1})(Y_1-\bar{Y}_1)^2 + (\frac{1}{\beta_2} - \frac{1}{\beta_1})\sum_{j=2}^{p}(Y_j-\bar{Y}_j)^2, \tag{2.35}$$

where Y_j and $\bar{Y}_j$ are the jth components of $\underline{Y}$ and $\underline{\bar{Y}}$, respectively. Then

$$\xi_j^{(i)} = 0,$$

$$\tau_{i1}^{*2} = (\frac{1}{\alpha_2} - \frac{1}{\alpha_1})[\alpha_i + \frac{\alpha_1\alpha_2}{N_1\alpha_2+N_2\alpha_1}],$$

$$\tau_{ij}^{*2} = (\frac{1}{\beta_2} - \frac{1}{\beta_1})[\beta_i + \frac{\beta_1\beta_2}{N_1\beta_2+N_2\beta_1}], \quad j=2,\ldots,p; \quad i=1,2. \tag{2.36}$$

Hence, if $\underline{X}$ belongs to π_1,

$$V \sim \left\{\frac{\alpha_1-\alpha_2}{\alpha_2}\left[1 + \frac{\alpha_2}{N_1\alpha_2+N_2\alpha_1}\right]\chi_1^2 + \frac{\beta_1-\beta_2}{\beta_2}\left[1+ \frac{\beta_2}{N_1\beta_2+N_2\beta_1}\right]\chi_{p-1}^2\right\}, \tag{2.37}$$

and if $\underline{X}$ belongs to π_2,

$$V \sim \left\{\frac{\alpha_1-\alpha_2}{\alpha_1}\left[1+ \frac{\alpha_1}{N_1\alpha_2+N_2\alpha_1}\right]\chi_1^2 + \frac{\beta_1-\beta_2}{\beta_1}\left[1+ \frac{\beta_1}{N_1\beta_2+N_2\beta_1}\right]\chi_{p-1}^2\right\}. \tag{2.38}$$

One could determine the c.d.f. of V in (2.37) or (2.38) by Press' method. Also, Patnaik's or Pearson's method could be used for approximating the distribution of V for the situation presented in this section.

D. Parameters Unknown

When the parameters are all unknown, we shall derive an asymptotic expansion for the distribution of the discriminant function. The technique used for the expansion is the "studentization" method of Hartley (1938) and of Welch (1947). Suppose a random sample $\underline{X}_1^{(1)}$, $\underline{X}_2^{(1)}$, ..., $\underline{X}_{N_1}^{(1)}$ is taken from π_1, and an independent random sample $\underline{X}_1^{(2)}$, $\underline{X}_2^{(2)}$, ..., $\underline{X}_{N_2}^{(2)}$ is taken from π_2. Our estimate of $\underline{\mu}^{(i)}$ is the sample mean $\bar{\underline{X}}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \underline{X}_\alpha^{(i)}$ and of $\underline{\Sigma}^{(i)}$ is the sample variance

$$\underline{S}^{(i)} = \frac{1}{n_i} \sum_{\alpha=1}^{N_i} (\underline{X}_\alpha^{(i)} - \overline{\underline{X}}^{(i)})(\underline{X}_\alpha^{(i)} - \overline{\underline{X}}^{(i)})', \tag{2.39}$$

where $n_i = N_i - 1$, $i=1,2$. If $\underline{Y} = \underline{\Gamma}\ \underline{X}$, where $\underline{\Gamma}$ is the (pxp) orthogonal matrix defined previously, then we will estimate $\underline{\nu}^{(i)} = \underline{\Gamma}\ \underline{\mu}^{(i)}$ by $\overline{\underline{Y}}^{(i)} = \underline{\Gamma}\ \overline{\underline{X}}^{(i)}$ and $\underline{D}^{(i)} = \mathrm{diag}(\alpha_i, \beta_i, \ldots, \beta_i)$ by

$\tilde{\underline{D}}^{(i)} = \mathrm{diag}(\tilde{\alpha}_i, \tilde{\beta}_i, \ldots, \tilde{\beta}_i)$ where

$$\tilde{\alpha}_i = \frac{\sum_{\alpha=1}^{N_i} (Y_{1\alpha}^{(i)} - \overline{Y}_1^{(i)})^2}{n_i},$$

$$\tilde{\beta}_i = \frac{\sum_{j=2}^{p} \sum_{\alpha=1}^{N_i} (Y_{j\alpha}^{(i)} - \overline{Y}_j^{(i)})^2}{(p-1)n_i}, \quad i=1,2. \tag{2.40}$$

The $\overline{Y}_j^{(i)}$ and $\tilde{\alpha}_i$, $\tilde{\beta}_i$ are independently distributed. After substituting for the unknown parameters by their respective estimates, the discriminant function in (2.10) becomes

$$V = \left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)\left(Y_1 - \frac{\dfrac{\overline{Y}_1^{(2)}}{\tilde{\alpha}_2} - \dfrac{\overline{Y}_1^{(1)}}{\tilde{\alpha}_1}}{\dfrac{1}{\tilde{\alpha}_2} - \dfrac{1}{\tilde{\alpha}_1}}\right)^2$$

$$+ \left(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}\right) \sum_{j=2}^{p} \left(Y_j - \frac{\dfrac{\overline{Y}_j^{(2)}}{\tilde{\beta}_2} - \dfrac{\overline{Y}_j^{(1)}}{\tilde{\beta}_1}}{\dfrac{1}{\tilde{\beta}_2} - \dfrac{1}{\tilde{\beta}_1}}\right)^2. \tag{2.41}$$

It is easily seen that V is invariant under any linear transformation. Hence, without loss of generality, we may let $\underline{\nu}^{(1)} = \underline{0}$, $\underline{D}^{(1)} = \underline{I}$, $\underline{\nu}^{(2)} = \underline{\nu}_0 = (\nu_{01}, \nu_{02}, \ldots, \nu_{0p})$, and $\underline{D}^{(2)} = \underline{D}_0 = \text{diag}\,(\alpha_0, \beta_0, \ldots, \beta_0)$. We shall derive the cumulative distribution function (c.d.f.) of V, $F_i(v)$, given that $\underline{X}$ comes from π_i, i=1,2.

The characteristic function (c.f.) of V when $\underline{X}$ comes from π_1 is

$$\phi(t/\pi_1) = E(e^{itV}/\pi_1), \tag{2.42}$$

which can be written as

$$\begin{aligned}\phi(t/\pi_1) &= E^{\underline{\overline{Y}}^{(1)}, \underline{\overline{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)}}\{E(e^{itV} | \underline{\overline{Y}}^{(1)}, \underline{\overline{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)}; \pi_1\} \\ &= E^{\underline{\overline{Y}}^{(1)}, \underline{\overline{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)}}\{\psi(\underline{\overline{Y}}^{(1)}, \underline{\overline{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)})\},\end{aligned} \tag{2.43}$$

where $\psi(\underline{\overline{Y}}^{(1)}, \underline{\overline{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)})$ is the conditional c.f. of V given $\underline{\overline{Y}}^{(1)}$, $\underline{\overline{Y}}^{(2)}$, $\underline{\tilde{D}}^{(1)}$, and $\underline{\tilde{D}}^{(2)}$. Recall that if λ is the non-centrality parameter of a non-central chi-square variate X with p degrees of freedom then the c.f. of X is given by

$$\phi_X(u) = (1-2iu)^{-p/2} \exp\{iu\lambda/(1-2iu)\}. \tag{2.44}$$

Noting that $\bar{\underline{Y}}^{(1)}$, $\bar{\underline{Y}}^{(2)}$, $\tilde{\underline{D}}^{(1)}$, $\tilde{\underline{D}}^{(2)}$ are fixed and that V is the sum of p independent non-central chi-square variates each with 1 degree of freedom, where

$$u_1 = \left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)t, \quad \lambda_1 = \frac{\left(\frac{\bar{Y}_1^{(2)}}{\tilde{\alpha}_2} - \frac{\bar{Y}_1^{(1)}}{\tilde{\alpha}_1}\right)^2}{\left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)^2}$$

for the first variate, and where

$$u_j = \left(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}\right)t, \quad \lambda_j = \frac{\left(\frac{\bar{Y}_j^{(2)}}{\tilde{\beta}_2} - \frac{\bar{Y}_j^{(1)}}{\tilde{\beta}_1}\right)^2}{\left(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}\right)^2}$$

for the remaining $j(j=2,\ldots,p)$ variates, we obtain

$$\psi(\bar{\underline{Y}}^{(1)}, \bar{\underline{Y}}^{(2)}, \tilde{\underline{D}}^{(1)}, \tilde{\underline{D}}^{(2)}) = \left(1-2it\left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)\right)^{-1/2}$$

$$\times \exp\left\{it\left(\frac{\bar{Y}_1^{(2)}}{\tilde{\alpha}_2} - \frac{\bar{Y}_1^{(1)}}{\tilde{\alpha}_1}\right)^2\left[\left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)\left(1-2it\left(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}\right)\right)\right]^{-1}\right\}$$

$$\times \left(1-2it\left(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}\right)\right)^{-\frac{(p-1)}{2}}$$

$$\times \exp\left\{it\sum_{j=2}^{p}\left(\frac{\bar{Y}_j^{(2)}}{\tilde{\beta}_2} - \frac{\bar{Y}_j^{(1)}}{\tilde{\beta}_1}\right)^2\right.$$

$$\times\ [(\frac{1}{\tilde\beta_2}-\frac{1}{\tilde\beta_1})(1-2it(\frac{1}{\tilde\beta_2}-\frac{1}{\tilde\beta_1}))]^{-1}\}. \qquad (2.45)$$

Since the function ψ is analytic about the point ($\underline{0}$, $\underline{\nu}_0$, $\underline{I}$, $\underline{D}_0$), expanding ψ into a Taylor's series, we have

$$\psi(\underline{\bar Y}^{(1)}, \underline{\bar Y}^{(2)}, \underline{\tilde D}^{(1)}, \underline{\tilde D}^{(2)})$$

$$= \exp\{[\sum_{j=1}^{p} \bar Y_j^{(1)} \frac{\partial}{\partial\nu_j^{(1)}} + \sum_{j=1}^{p} (\bar Y_j^{(2)} - \nu_{0j})\frac{\partial}{\partial\nu_j^{(2)}}$$

$$+(\tilde\alpha_1-1)\frac{\partial}{\partial\alpha_1} + (p-1)(\tilde\beta_1-1)\frac{\partial}{\partial\beta_1} + (\tilde\alpha_2-\alpha_0)\frac{\partial}{\partial\alpha_2}$$

$$+(p-1)(\tilde\beta_2-\beta_0)\frac{\partial}{\partial\beta_2}]\}\ \psi(\underline{\nu}^{(1)},\underline{\nu}^{(2)},\underline{D}^{(1)},\underline{D}^{(2)})|_0\ , \qquad (2.46)$$

where $|_0$ denotes that the expression is evaluated at ($\underline{0}$, $\underline{\nu}_0$, $\underline{I}$, $\underline{D}_0$). The c.f. of V is then

$$\phi(t/\pi_1) = E^{\underline{\bar Y}^{(1)},\underline{\bar Y}^{(2)},\underline{\tilde D}^{(1)},\underline{\tilde D}^{(2)}}\{\psi(\underline{\bar Y}^{(1)},\underline{\bar Y}^{(2)},\underline{\tilde D}^{(1)},\underline{\tilde D}^{(2)})\}$$

$$= (H)\,\psi(\underline{\nu}^{(1)},\underline{\nu}^{(2)},\underline{D}^{(1)},\underline{D}^{(2)})|_0, \qquad (2.47)$$

where (H) is the differential operator

$$\textcircled{H} = E^{\underline{\overline{Y}}^{(1)},\underline{\overline{Y}}^{(2)},\underline{\tilde{D}}^{(1)},\underline{\tilde{D}}^{(2)}}\{\exp[\sum_{j=1}^{p}\overline{Y}_j^{(1)} \frac{\partial}{\partial\nu_j^{(1)}}$$

$$+ \sum_{j=1}^{p}(\overline{Y}_j^{(2)} - \nu_{0j})\frac{\partial}{\partial\nu_j^{(2)}} + (\tilde{\alpha}_1-1)\frac{\partial}{\partial\alpha_1} + (p-1)(\tilde{\beta}_1-1)\frac{\partial}{\partial\beta_1}$$

$$+ (\tilde{\alpha}_2-\alpha_0)\frac{\partial}{\partial\alpha_2} + (p-1)(\tilde{\beta}_2-\beta_0)\frac{\partial}{\partial\beta_2}]\}. \tag{2.48}$$

Making use of the fact that $\underline{\overline{Y}}^{(i)}$ and $\underline{\tilde{D}}^{(i)}$ are independent, we have

$$\textcircled{H} = E^{\underline{\overline{Y}}^{(1)}}\{\exp[\sum_{j=1}^{p}\overline{Y}_j^{(1)} \frac{\partial}{\partial\nu_j^{(1)}}]\}$$

$$\times E^{\underline{\overline{Y}}^{(2)}}\{\exp[\sum_{j=1}^{p}(\overline{Y}_j^{(2)} - \nu_{0j})\frac{\partial}{\partial\nu_j^{(2)}}]\}$$

$$\times E^{\underline{\tilde{D}}^{(1)}}\{\exp[(\tilde{\alpha}_1-1)\frac{\partial}{\partial\alpha_1} + (p-1)(\tilde{\beta}_1-1)\frac{\partial}{\partial\beta_1}]\}$$

$$\times E^{\underline{\tilde{D}}^{(2)}}\{\exp[(\tilde{\alpha}_2-\alpha_0)\frac{\partial}{\partial\alpha_2} + (p-1)(\tilde{\beta}_2-\beta_0)\frac{\partial}{\partial\beta_2}]\}. \tag{2.49}$$

Since $\overline{Y}_j^{(i)}$ and $\tilde{\alpha}_i$, $\tilde{\beta}_i$ are independently distributed as normal and χ^2, respectively, using the moment generating functions (m.g.f.), we have

$$\textcircled{H} = \exp\{\sum_{j=1}^{p}\frac{1}{2N_1}\frac{\partial^2}{\partial(\nu_j^{(1)})^2}\}\times\exp\{\frac{1}{2N_2}[\alpha_0\frac{\partial^2}{\partial(\nu_1^{(2)})^2}$$

$$+\beta_0\sum_{j=2}^{p}\frac{\partial^2}{\partial(\nu_j^{(2)})^2}]\}\times\exp\{-\frac{\partial}{\partial\alpha_1}-(p-1)\frac{\partial}{\partial\beta_1}$$

$$-\frac{n_1}{2}[\log(1-\frac{2}{n_1}\frac{\partial}{\partial\alpha_1})+(p-1)\log(1-\frac{2}{n_1}\frac{\partial}{\partial\beta_1})]\}$$

$$\times\exp\{-\alpha_0\frac{\partial}{\partial\alpha_2}-(p-1)\beta_0\frac{\partial}{\partial\beta_2}-\frac{n_2}{2}[\log(1-\frac{2\alpha_0}{n_2}\frac{\partial}{\partial\alpha_2})$$

$$+(p-1)\log(1-\frac{2\beta_0}{n_2}\frac{\partial}{\partial\beta_2})]\}. \tag{2.50}$$

Substituting $\log(1-x) = -x-\frac{1}{2}x^2-\frac{1}{3}x^3-\ldots$ in (2.50) we obtain

$$\textcircled{H} = \exp\{[\frac{1}{2N_1}\sum_{j=1}^{p}\frac{\partial^2}{\partial(\nu_1^{(1)})^2}+\frac{\alpha_0}{2N_2}\frac{\partial^2}{\partial(\nu_1^{(2)})^2}+\frac{1}{n_1}\frac{\partial^2}{\partial\alpha_1^2}$$

$$+\frac{1}{n_2}\alpha_0^2\frac{\partial^2}{\partial\alpha_2^2}+\frac{\beta_0}{2N_2}\sum_{j=2}^{p}\frac{\partial^2}{\partial(\nu_j^{(2)})^2}+\frac{(p-1)}{n_1}\frac{\partial^2}{\partial\beta_1^2}$$

$$+\frac{(p-1)\beta_0^2}{n_2}\frac{\partial^2}{\partial\beta_2^2}]+O_2^*\}, \tag{2.51}$$

where O_2^* stands for terms of second order with respect to

(n_1^{-1}, n_2^{-1}). Expanding the exponential in (2.51), we have

$$\text{Ⓗ} = 1 + \frac{1}{2N_1}\sum_{j=1}^{p}\frac{\partial^2}{\partial(\nu_j^{(1)})^2} + \frac{\alpha_0}{2N_2}\frac{\partial^2}{\partial(\nu_1^{(2)})^2} + \frac{1}{n_1}\frac{\partial^2}{\partial\alpha_1^2}$$

$$+ \frac{1}{n_2}\alpha_0^2\frac{\partial^2}{\partial\alpha_2^2} + \frac{\beta_0}{2N_2}\sum_{j=2}^{p}\frac{\partial^2}{\partial(\nu_j^{(2)})^2} + \frac{(p-1)}{n_1}\frac{\partial^2}{\partial\beta_1^2}$$

$$+ \frac{(p-1)}{n_2}\beta_0^2\frac{\partial^2}{\partial\beta_2^2} + O_2 , \qquad (2.52)$$

where O_2 stands for terms of second order with respect to $(N_1^{-1}, N_2^{-1}, n_1^{-1}, n_2^{-1})$. We now find the individual terms in (2.47). The principal term of $\phi(t/\pi_1)$ is

$$\psi(\underline{0},\underline{\nu}_0,\underline{I},\underline{D}_0) = (1-2ita_1)^{-1/2}\exp\{it\,\nu_{01}^2\,[\alpha_0^2 a_1(1-2ita_1)]^{-1}\}$$

$$\times (1-2ita_2)^{-\frac{p-1}{2}}\exp\{it\sum_{j=2}^{p}\nu_{0j}^2\,[\beta_0^2 a_2(1-ita_2)]^{-1}\} = \psi_0, \qquad (2.53)$$

where $a_1 = \frac{1}{\alpha_0} - 1$, $a_2 = \frac{1}{\beta_0} - 1$. ψ_0 is the c.f. of $\sum_{j=1}^{p} Z_j^2$, where Z_j are independently distributed as

$$Z_1 \sim N(\nu_{01}/\alpha_0\sqrt{a_1},\, a_1),$$

$$Z_j \sim N(\nu_{0j}/\beta_0\sqrt{a_2},\, a_2), \quad j=2,\dots,p . \qquad (2.54)$$

To find the term associated with N_1^{-1}, we have to differentiate the function $\psi(\underline{\nu}^{(1)}, \underline{\nu}^{(2)}, \underline{D}^{(1)}, \underline{D}^{(2)})$ with respect to $\nu_j^{(1)}$. The function $\psi(\underline{\nu}^{(1)}, \underline{\nu}^{(2)}, \underline{D}^{(1)}, \underline{D}^{(2)})$ is given by

$$\psi(\underline{\nu}^{(1)},\underline{\nu}^{(2)},\underline{D}^{(1)},\underline{D}^{(2)})=(1-2it(\frac{1}{\alpha_2}-\frac{1}{\alpha_1}))^{-\frac{1}{2}}$$

$$\times \exp\{it(\frac{\nu_1^{(2)}}{\alpha_2}-\frac{\nu_1^{(1)}}{\alpha_1})^2[(\frac{1}{\alpha_2}-\frac{1}{\alpha_1})(1-2it(\frac{1}{\alpha_2}-\frac{1}{\alpha_1}))]^{-1}\}$$

$$\times (1-2it(\frac{1}{\beta_2}-\frac{1}{\beta_1}))^{-\frac{p-1}{2}}\exp\{it\sum_{j=2}^{p}(\frac{\nu_j^{(2)}}{\beta_2}-\frac{\nu_j^{(1)}}{\beta_1})^2$$

$$\times [(\frac{1}{\beta_2}-\frac{1}{\beta_1})(1-2it(\frac{1}{\beta_2}-\frac{1}{\beta_1}))]^{-1}\}. \qquad (2.55)$$

Noting that the last (p-1) factors of $\psi(\underline{\nu}^{(1)}, \underline{\nu}^{(2)}, \underline{D}^{(1)}, \underline{D}^{(2)})$ in (2.55) can be obtained from the first factor, say

$$L= (1-2it(\frac{1}{\alpha_2}-\frac{1}{\alpha_1}))^{-\frac{1}{2}}\exp\{it(\frac{\nu_1^{(2)}}{\alpha_2}-\frac{\nu_1^{(1)}}{\alpha_1})^2$$

$$[(\frac{1}{\alpha_2}-\frac{1}{\alpha_1})(1-2it(\frac{1}{\alpha_2}-\frac{1}{\alpha_1}))]^{-1}\}, \qquad (2.56)$$

by replacing α_i by β_i and $\nu_1^{(i)}$ by $\nu_j^{(i)}$, $i=1,2$; $j=2,\ldots,p$, we shall, for convenience in differentiation, only differentiate L. Since differentiation is concerned only with $\nu_1^{(1)}$ in L it makes the calculation simpler to put $\nu_1^{(2)} = \nu_{01}$, $\alpha_1 = 1$, $\alpha_2 = \alpha_0$ before differentiation and $\nu_1^{(1)} = 0$ after differentiation.

Thus,

$$\frac{\partial L}{\partial \nu_1^{(1)}} = (1-2ita_1)^{-1/2}\ 2it\left(\frac{\nu_{01}}{\alpha_0} - \nu_1^{(1)}\right)(-1)$$

$$\times\ [a_1(1-2ita_1)]^{-1}\exp\{A\}, \tag{2.57}$$

where

$$A = it\left(\frac{\nu_{01}}{\alpha_0} - \nu_1^{(1)}\right)^2 [a_1(1-2ita_1)]^{-1} .$$

Further,

$$\frac{\partial^2 L}{\partial (\nu_1^{(1)})^2} = (1-2ita_1)^{-1/2}\ 2it[a_1(1-2ita_1)]^{-1}\exp\{A\}$$

$$+\ (1-2ita_1)^{-1/2}\ 4(it)^2\left(\frac{\nu_{01}}{\alpha_0} - \nu_1^{(1)}\right)^2$$

$$\times\ [a_1(1-2ita_1)]^{-2}\ \exp\{A\}. \tag{2.58}$$

Therefore,

$$\frac{1}{2N_1}\ \frac{\partial^2}{\partial (\nu_j^{(1)})^2}\ \psi(\underline{\nu}^{(1)}, \underline{\nu}^{(2)}, \underline{D}^{(1)}, \underline{D}^{(2)})\big|_0$$

$$= \frac{1}{N_1}\left\{\left[\frac{it}{(1-2ita_1)a_1} + \frac{2(it)^2\nu_{01}^2}{(1-2ita_1)^2\alpha_0^2 a_1^2}\right]\right.$$

$$\left. +\ \left[(p-1)\frac{it}{(1-2ita_2)a_2} + \frac{2(it)^2\sum_{j=2}^{p}\nu_{0j}^2}{(1-2ita_2)^2\beta_0^2 a_2^2}\right]\right\}\psi_0\ . \tag{2.59}$$

Next we shall find the term associated with N_2^{-1}. Since differentiation is concerned only with $\nu_1^{(2)}$ in L, for L in (2.56), we put $\nu_1^{(1)} = 0$, $\alpha_1 = 1$, $\alpha_2 = \alpha_0$ before differentiation and $\nu_1^{(2)} = \nu_{01}$ after differentiation. Then,

$$\frac{\partial L}{\partial \nu_1^{(2)}} = (1-2ita_1)^{-1/2}\, 2it\frac{\nu_1^{(2)}}{\alpha_0^2}\,[a_1(1-2ita_1)]^{-1}\exp\{B\}, \quad (2.60)$$

where

$$B = it\left(\frac{\nu_1^{(2)}}{\alpha_0}\right)^2[a_1(1-2ita_1)]^{-1},$$

and

$$\frac{\partial^2 L}{\partial(\nu_1^{(2)})^2} = (1-2ita_1)^{-1/2}\,\frac{2it}{\alpha_0^2}\,[a_1(1-2ita_1)]^{-1}\exp\{B\}$$

$$+ (1-2ita_1)^{-1/2}\,\frac{4(it)^2}{\alpha_0^2}\left(\frac{\nu_1^{(2)}}{\alpha_0}\right)^2[a_1(1-2ita_1)]^{-1}\exp\{B\}. \quad (2.61)$$

Therefore,

$$\left[\frac{\alpha_0}{2N_2}\frac{\partial^2}{\partial(\nu_1^{(2)})^2} + \frac{\beta_0}{2N_2}\sum_{j=2}^{p}\frac{\partial^2}{\partial(\nu_j^{(2)})^2}\right]\psi(\underline{\nu}^{(1)},\underline{\nu}^{(2)},\underline{D}^{(1)},\underline{D}^{(2)})\Big|_0$$

$$= \frac{1}{N_2}\left\{\left[\frac{it}{(1-2ita_1)\alpha_0 a_1} + \frac{2(it)^2\nu_{01}^2}{(1-2ita_1)^2\alpha_0^3 a_1^2}\right]\right.$$

$$\left. +\left[\frac{(p-1)it}{(1-2ita_2)\beta_0 a_2} + \frac{2(it)^2\sum_{j=2}^{p}\nu_{0j}^2}{(1-2ita_2)^2\beta_0^3 a_2^2}\right]\right\}\psi_0 . \quad (2.62)$$

To obtain the term associated with n_1^{-1}, which involves differentiating L with respect to α_1, we set $\nu_1^{(2)} = \nu_{01}$, $\alpha_2=\alpha_0$ before differentiation and $\nu_1^{(1)} = 0$, $\alpha_1 = 1$ after differentiation. Hence,

$$\frac{\partial L}{\partial \alpha_1} = (1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-3/2} \frac{it}{\alpha_1^2} \exp\{C\}$$

$$+ (1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-1/2} \exp\{C\}$$

$$\times \{2it(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})\frac{\nu_1^{(1)}}{\alpha_1^2}[(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))]^{-1}$$

$$- it(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2[(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))]^{-2}$$

$$\times [\frac{1}{\alpha_1^2}(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})) - \frac{2it}{\alpha_1^2}(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})]\}, \tag{2.63}$$

where

$$C = it(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2[(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))]^{-1}. \tag{2.64}$$

Combining terms, (2.63) can be written as

$$\frac{\partial L}{\partial \alpha_1} = (1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-1/2} \exp\{C\} \times [D], \tag{2.65}$$

where

$$D = it(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-1}[\frac{1}{\alpha_1^2} + \frac{2\nu_1^{(1)}}{\alpha_1^2} (\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1}$$

$$- \frac{1}{\alpha_1^2}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-2}]$$

$$+ (it)^2(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-2}[\frac{2}{\alpha_1^2}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1}]. \tag{2.66}$$

Therefore,

$$\frac{\partial^2 L}{\partial(\alpha_1)^2} = (1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-1/2} \exp\{C\}\times\{[D]^2 + \frac{\partial}{\partial\alpha_1}[D]\}, \tag{2.67}$$

where C and D are given in (2.64) and (2.66), respectively.

Expanding (2.67) we have

$$\frac{\partial^2 L}{\partial(\alpha_1)^2} = (1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))^{-1/2}\exp\{C\}\times\{[D]^2$$

$$+ \frac{2it\alpha_1^{-3}}{(1-2it(\frac{1}{\alpha_0} - \frac{1}{\alpha_1}))}(-1-2\nu_1^{(1)} (\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1}$$

$$+ \frac{(\nu_1^{(1)})^2}{\alpha_1}(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1} + (\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-2}$$

$$-2\frac{\nu_1^{(1)}}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-2} + \frac{1}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-3})$$

$$+ \frac{2(it)^2\alpha_1^{-3}}{(1-2it(\frac{1}{\alpha_0}-\frac{1}{\alpha_1}))^2}(\frac{1}{\alpha_1} + 2\frac{\nu_1^{(1)}}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1}$$

$$- \frac{1}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-2}$$

$$-2(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1} + 2\frac{\nu_1^{(1)}}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1}$$

$$- \frac{1}{\alpha_1}(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-2})$$

$$+ \frac{2(it)^3\alpha_1^{-4}}{(1-2it(\frac{1}{\alpha_0}-\frac{1}{\alpha_1}))^3}(4(\frac{\nu_{01}}{\alpha_0} - \frac{\nu_1^{(1)}}{\alpha_1})^2(\frac{1}{\alpha_0} - \frac{1}{\alpha_1})^{-1})\}. \qquad (2.68)$$

Setting $\nu_1^{(1)} = 0$ and $\alpha_1 = 1$ in (2.68) we have

$$\left.\frac{\partial^2 L}{\partial(\alpha_1)^2}\right|_{(0,\nu_{01},1,\alpha_0)} = (1-2ita_1)^{-\frac{1}{2}}\exp\{it\ \nu_{01}^2\ [\alpha_0^2 a_1(1-2ita_1)]^{-1}\}$$

$$\times\{[\frac{it}{1-2ita_1}(1-\frac{\nu_{01}^2}{\alpha_0^2 a_1^2}) + \frac{(it)^2}{(1-2ita_1)^2}(2\frac{\nu_{01}^2}{\alpha_0^2\ a_1})]^2$$

$$+ \frac{2(it)}{1-2ita_1}(-1 + \frac{\nu_{01}^2}{\alpha_0^2 a_1^2} + \frac{\nu_{01}^2}{\alpha_0^2 a_1^3})$$

$$+ \frac{2(it)^2}{(1-2ita_1)^2}(1-\frac{\nu_{01}^2}{\alpha_0^2 a_1^2} - 2\frac{\nu_{01}^2}{\alpha_0^2 a_1} - \frac{\nu_{01}^2}{\alpha_0^2 a_1^2})$$

$$+ \frac{2(it)^3}{(1-2ita_1)^3}\left(4\frac{\nu_{01}^2}{\alpha_0^2 a_1}\right)\}. \tag{2.69}$$

It is easily seen from the first factor L that the term associated with n_1^{-1} is

$$\frac{1}{n_1}\{[\frac{it}{1-2ita_1} c_1' + \frac{it}{1-2ita_2}\sum_{j=2}^{p} c_j']$$

$$+ [\frac{(it)^2}{(1-2ita_1)^2}(c_1^2+2c_1+b_1') + \frac{(it)^2}{(1-2ita_2)^2}\sum_{j=2}^{p}(c_j^2+2c_j+b_j')]$$

$$+ [\frac{(it)^3}{(1-2ita_1)^3}(4b_1+2b_1c_1) + \frac{(it)^3}{(1-2ita_2)^3}\sum_{j=2}^{p}(4b_j+2b_jc_j)]$$

$$+ [\frac{(it)^4}{(1-2ita_1)^4} b_1^2 + \frac{(it)^4}{(1-2ita_2)^4}\sum_{j=2}^{p} b_j^2]\}\psi_0, \tag{2.70}$$

where

$$b_1 = \frac{2\nu_{01}^2}{\alpha_0^2 a_1}, \quad b_j = \frac{2\nu_{0j}^2}{\beta_0^2 a_2},$$

$$b_1' = \frac{-2\nu_{01}^2}{\alpha_0^2\, a_1}\left(\frac{1}{a_1} + 2\right), \quad b_j' = \frac{-2\nu_{0j}^2}{\beta_0^2\, a_2}\left(\frac{1}{a_2} + 2\right),$$

$$c_1 = \frac{-\nu_{01}^2}{\alpha_0^2 a_1^2} + 1, \quad c_j = \frac{-\nu_{0j}^2}{\beta_0^2 a_2^2} + 1,$$

$$c_1' = 2\left(\frac{\nu_{01}^2}{\alpha_0^3 a_1^3} - 1\right), \quad c_j' = 2\left(\frac{\nu_{0j}^2}{\beta_0^3 a_2^3} - 1\right), \quad j=2,\ldots,p. \tag{2.71}$$

Similarly, one could obtain the term associated with n_2^{-1}, which involves differentiating the first factor L with respect to α_2. To simplify the calculations we set $\nu_1^{(1)} = 0$, $\nu_1^{(2)} = \nu_{01}$, $\alpha_1=1$ before differentiation and $\alpha_2 = \alpha_0$ after differentiation. Consequently, we find the term associated with n_2^{-1} is

$$\frac{1}{n_2}\Big\{\Big[\frac{it}{1-2ita_1}\alpha_0^2 h_1' + \frac{it}{1-2ita_2}\beta_0^2 \sum_{j=2}^{p} h_j'\Big]$$

$$+\Big[\frac{(it)^2}{(1-2ita_1)^2}(\alpha_0^2 h_1^2 - 2h_1 - \alpha_0^2 g_1')$$

$$+\frac{(it)^2}{(1-2ita_2)^2}\sum_{j=2}^{p}(\beta_0^2 h_j^2 - 2h_j - \beta_0^2 g_j')\Big]$$

$$+\Big[\frac{(it)^3}{(1-2ita_1)^3}(4g_1 - 2\alpha_0^2 g_1 h_1) + \frac{(it)^3}{(1-2ita_2)^3}\sum_{j=2}^{p}(4g_j - 2\beta_0^2 g_j h_j)\Big]$$

$$+\Big[\frac{(it)^4}{(1-2ita_1)^4}\alpha_0^2 g_1^2 + \frac{(it)^4}{(1-2ita_2)^4}\beta_0^2 \sum_{j=2}^{p} g_j^2\Big]\Big\}\psi_0, \tag{2.72}$$

where

$$g_1 = \frac{2\nu_{01}^2}{\alpha_0^4 a_1}, \quad g_j = \frac{2\nu_{0j}^2}{\beta_0^4 a_2},$$

$$g_1' = \frac{2\nu_{01}^2}{\alpha_0^5 a_1}\left(\frac{1}{\alpha_0 a_1} - 4\right), \quad g_j' = \frac{2\nu_{0j}^2}{\beta_0^5 a_2}\left(\frac{1}{\beta_0 a_2} - 4\right),$$

$$h_1 = \frac{\nu_{01}^2}{\alpha_0^3 a_1}\left(\frac{1}{a_1} - 1\right) - \frac{1}{\alpha_0^2}, \quad h_j = \frac{\nu_{0j}^2}{\beta_0^3 a_2}\left(\frac{1}{a_2} - 1\right) - \frac{1}{\beta_0^2},$$

$$h_1' = \frac{2}{\alpha_0^3}\left(\frac{\nu_{01}^2}{\alpha_0^3 a_1^3} - \frac{3\nu_{01}^2}{\alpha_0 a_1^2} + 1\right), \quad h_j' = \frac{2}{\beta_0^3}\left(\frac{\nu_{0j}^2}{\beta_0^3 a_2^3} - \frac{3\nu_{0j}^2}{\beta_0 a_2^2} + 1\right),$$

$$j=2,\ldots,p. \tag{2.73}$$

The c.f. of V, after collecting terms, is

$$\phi(t/\pi_1) = \{1 + it \sum_{k=1}^{2} \frac{m_{1k}}{(1-2ita_k)} + (it)^2 \sum_{k=1}^{2} \frac{m_{2k}}{(1-2ita_k)^2}$$
$$+ (it)^3 \sum_{k=1}^{2} \frac{m_{3k}}{(1-2ita_k)^3} + (it)^4 \sum_{k=1}^{2} \frac{m_{4k}}{(1-2ita_k)^4}\}\psi_0 + O_2, \tag{2.74}$$

where

$$m_{11} = \frac{1}{N_1 a_1} + \frac{1}{N_2 \alpha_0 a_1} + \frac{1}{n_1} c_1' + \frac{1}{n_2} \alpha_0^2 h_1',$$

$$m_{12} = \frac{p-1}{N_1 a_2} + \frac{(p-1)}{N_2 \beta_0 a_2} + \frac{1}{n_1} \sum_{j=2}^{p} c_j' + \frac{1}{n_2} \beta_0^2 \sum_{j=2}^{p} h_j',$$

$$m_{21} = \frac{2\nu_{01}^2}{N_1\alpha_0^2 a_1^2} + \frac{2\nu_{01}^2}{N_2\alpha_0^3 a_1^2} + \frac{1}{n_1}(c_1^2+2c_1+b_1') + \frac{1}{n_2}(\alpha_0^2 h_1^2 - 2h_1 - \alpha_0^2 g_1'),$$

$$m_{22} = \frac{2\sum_{j=2}^{p}\nu_{0j}^2}{N_1\beta_0^2 a_2^2} + \frac{2\sum_{j=2}^{p}\nu_{0j}^2}{N_2\beta_0^3 a_2^2} + \frac{1}{n_1}\sum_{j=2}^{p}(c_j^2+2c_j+b_j') + \frac{1}{n_2}\sum_{j=2}^{p}(\beta_0^2 h_j^2 - 2h_j - \beta_0^2 g_j'),$$

$$m_{31} = \frac{1}{n_1}(4b_1+2b_1c_1) + \frac{1}{n_2}(4g_1-2\alpha_0^2 g_1h_1)$$

$$m_{32} = \frac{1}{n_1}\sum_{j=2}^{p}(4b_j+2b_jc_j) + \frac{1}{n_2}\sum_{j=2}^{p}(4g_j-2\beta_0^2 g_jh_j)$$

$$m_{41} = \frac{1}{n_1}b_1^2 + \frac{1}{n_2}\alpha_0^2 g_1^2,$$

$$m_{42} = \frac{1}{n_1}\sum_{j=2}^{p}b_j^2 + \frac{1}{n_2}\beta_0^2\sum_{j=2}^{p}g_j^2. \tag{2.75}$$

In order to invert the c.f. to obtain the c.d.f. $F_1(v)$, we use the method given by Wallace (1958) which was used by Ito (1960) and Han (1969, 1970) for similar problems. If $F(x)$ is the c.d.f. of a statistic and $\phi(t)$ is its c.f., then the c.d.f. corresponding to $(-it)^r \phi(t)$ is $F^{(r)}(x)$ where $F^{(r)}(x)$ is the rth derivative of $F(x)$. Now let $G_{m,k}(x)$ be the c.d.f. of a non-central chi-square variate with m degrees of freedom and non-centrality parameter

$\frac{\nu_{01}^2}{\alpha_0^2 a_1^2}$ if k=1 and $\frac{\sum_{j=2}^{p} \nu_{oj}^2}{\beta_0^2 a_2^2}$ if k=2. Then

$$F_1(v) = G_{1,1}(v) + G_{(p-1),2}(v) - m_{11}G_{3,1}^{(1)}(v) - m_{12}G_{3(p-1),2}^{(1)}(v)$$
$$+ m_{21}G_{5,1}^{(2)}(v) + m_{22}G_{5(p-1),2}^{(2)}(v) - m_{31}G_{7,1}^{(3)}(v)$$
$$- m_{32}G_{7(p-1),2}^{(3)}(v) + m_{41}G_{9,1}^{(4)}(v) + m_{42}G_{9(p-1),2}^{(4)}(v) + O_2, \tag{2.76}$$

where $G_{m,k}^{(r)}(v)$ is the rth derivative of $G_{m,k}(v)$.

To find the c.d.f. $F_2(v)$ when $\underline{X}$ comes from π_2, a similar procedure is used. When $\underline{X}$ comes from π_1, the conditional c.f. is given in (2.45). Using a procedure similar to that used in obtaining (2.45), we can find the conditional c.f. $\psi(\underline{\bar{Y}}^{(1)}, \underline{\bar{Y}}^{(2)}, \underline{\tilde{D}}^{(1)}, \underline{\tilde{D}}^{(2)})$ when $\underline{X}$ comes from π_2. This conditional c.f. is

$$\psi(\underline{\bar{Y}}^{(1)},\underline{\bar{Y}}^{(2)},\underline{\tilde{D}}^{(1)},\underline{\tilde{D}}^{(2)}) = (1-2it\alpha_0(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}))^{-1/2}$$

$$\times \exp\{(it)[\nu_{01}(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1})-(\frac{\bar{Y}_1^{(2)}}{\tilde{\alpha}_2} - \frac{\bar{Y}_1^{(1)}}{\tilde{\alpha}_1})]^2$$

$$\times [(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1})(1-2it\alpha_0(\frac{1}{\tilde{\alpha}_2} - \frac{1}{\tilde{\alpha}_1}))]^{-1}\}$$

$$\times (1-2it\beta_0(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}))^{-\frac{(p-1)}{2}}$$

$$\times \exp\{(it)\sum_{j=2}^{p}[\nu_{0j}(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1})-(\frac{\bar{Y}_j^{(2)}}{\tilde{\beta}_2} - \frac{\bar{Y}_j^{(1)}}{\tilde{\beta}_1})]^2$$

$$\times [(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1})(1-2it\beta_0(\frac{1}{\tilde{\beta}_2} - \frac{1}{\tilde{\beta}_1}))]^{-1}\}. \qquad (2.77)$$

Again expanding ψ in a Taylor's series about $(\underline{0}, \underline{\nu}_0, \underline{I}, \underline{D}_0)$, we obtain as the principal term

$$\psi_0 = (1-2it\alpha_0 a_1)^{-\frac{1}{2}} \exp\{it\nu_{01}^2[a_1(1-2it\alpha_0 a_1)]^{-1}\}$$

$$\times (1-2it\beta_0 a_2)^{-\frac{(p-1)}{2}} \exp\{it\sum_{j=2}^{p}\nu_{0j}^2[a_2(1-2it\beta_0 a_2)]^{-1}\}, \qquad (2.78)$$

which is the c.f. of $\sum_{j=1}^{p} Z_j^2$ where Z_j are independently distributed as

$$Z_1 \sim N(\nu_{01}/\sqrt{a_1},\ \alpha_0 a_1),$$

$$Z_j \sim N(\nu_{0j}/\sqrt{a_2},\ \beta_0 a_2),\quad j=2,\ldots,p\ . \tag{2.79}$$

As was done when $\underline{X}$ came from π_1, we also want to obtain the terms associated with N_1^{-1}, N_2^{-1}, n_1^{-1}, and n_2^{-1} for this case. These terms are found to be

$$\frac{1}{N_1}\{[\frac{it}{(1-2it\alpha_0 a_1)a_1} + \frac{2(it)^2\nu_{01}^2}{(1-2it\alpha_0 a_1)^2 a_1^2}] + [\frac{(p-1)it}{(1-2it\beta_0 a_2)a_2}$$

$$+ \frac{2(it)^2 \sum_{j=2}^{p} \nu_{oj}^2}{(1-2it\beta_0 a_2)^2 a_2^2}]\}\psi_0, \tag{2.80}$$

$$\frac{1}{N_2}\{[\frac{it}{(1-2it\alpha_0 a_1)\alpha_0^2 a_1} + \frac{2(it)^2\nu_{01}^2}{(1-2it\alpha_0 a_1)^2\alpha_0^2 a_1^2}] + [\frac{(p-1)it}{(1-2it\beta_0 a_2)\beta_0^2 a_2}$$

$$+ \frac{2(it)^2 \sum_{j=2}^{p} \nu_{oj}^2}{(1-2it\beta_0 a_2)^2\beta_0^2 a_2^2}]\}\psi_0, \tag{2.81}$$

$$\frac{1}{n_1}\{[\frac{it}{(1-2it\alpha_0 a_1)}\ C_1' + \frac{it}{(1-2it\beta_0 a_2)}\ \sum_{j=2}^{p} C_j']$$

$$+ [\frac{(it)^2}{(1-2it\alpha_0 a_1)^2}\ (C_1^2 + 2\alpha_0 C_1 + B_1')$$

$$+\frac{(it)^2}{(1-2it\beta_0 a_2)^2}\sum_{j=2}^{p}(C_j^2+2\beta_0 C_j+B_j')]$$

$$+[\frac{(it)^3}{(1-2it\alpha_0 a_1)^3}(4\alpha_0 B_1+2B_1C_1)+\frac{(it)^3}{(1-2it\beta_0 a_2)^3}\sum_{j=2}^{p}(4\beta_0 B_j+2B_jC_j)]$$

$$+[\frac{(it)^4}{(1-2it\alpha_0 a_1)^4}B_1^2+\frac{(it)^4}{(1-2it\beta_0 a_2)^4}\sum_{j=2}^{p}B_j^2]\}\psi_0, \qquad (2.82)$$

$$\frac{1}{n_2}\{[\frac{it}{(1-2it\alpha_0 a_1)}\alpha_0^2 H_1'+\frac{it}{(1-2it\beta_0 a_2)}\beta_0^2\sum_{j=2}^{p}H_j']$$

$$+[\frac{(it)^2}{(1-2it\alpha_0 a_1)^2}(\alpha_0^2 H_1^2-2\alpha_0 H_1-\alpha_0^2 G_1')$$

$$+\frac{(it)^2}{(1-2it\beta_0 a_2)^2}\sum_{j=2}^{p}(\beta_0^2 H_j^2-2\beta_0 H_j-\beta_0^2 G_j')]$$

$$+[\frac{(it)^3}{(1-2it\alpha_0 a_1)^3}(4\alpha_0 G_1-2\alpha_0^2 G_1 H_1)$$

$$+\frac{(it)^3}{(1-2it\beta_0 a_2)^3}\sum_{j=2}^{p}(4\beta_0 G_j-2\beta_0^2 G_j H_j)]$$

$$+[\frac{(it)^4}{(1-2it\alpha_0 a_1)^4}\alpha_0^2 G_1^2+\frac{(it)^4}{(1-2it\beta_0 a_2)^4}\beta_0^2\sum_{j=1}^{p}G_j^2]\}\psi_0, \qquad (2.83)$$

where

$$B_1 = \frac{2\nu_{01}^2\alpha_0}{a_1}, \quad B_j = \frac{2\nu_{0j}^2\beta_0}{a_2},$$

$$B_1' = -\frac{2\nu_{01}^2\alpha_0}{a_1}\left(\frac{1}{a_1} + 4\right), \quad B_j' = -\frac{2\nu_{0j}^2\beta_0}{a_2}\left(\frac{1}{a_2} + 4\right),$$

$$C_1 = -\frac{\nu_{01}^2}{a_1^2} - \frac{2\nu_{01}^2}{a_1} + \alpha_0, \quad C_j = -\frac{\nu_{0j}^2}{a_2^2} - \frac{2\nu_{0j}^2}{a_2} + \beta_0,$$

$$C_1' = 2\left[\frac{\nu_{01}^2}{a_1^3} + \frac{3\nu_{01}^2}{a_1^2} + \frac{3\nu_{01}^2}{a_1} - \alpha_0\right],$$

$$C_j' = 2\left[\frac{\nu_{0j}^2}{a_2^3} + \frac{3\nu_{0j}^2}{a_2^2} + \frac{3\nu_{0j}^2}{a_2} - \beta_0\right],$$

$$G_1 = \frac{2\nu_{01}^2}{\alpha_0 a_1}, \quad G_j = \frac{2\nu_{0j}^2}{\beta_0 a_2},$$

$$G_1' = \frac{2\nu_{01}^2}{\alpha_0^2 a_1}\left[\frac{1}{\alpha_0 a_1} - 2\right], \quad G_j' = \frac{2\nu_{0j}^2}{\beta_0^2 a_2}\left[\frac{1}{\beta_0 a_2} - 2\right],$$

$$H_1 = \frac{\nu_{01}^2}{\alpha_0^2 a_1^2} - \frac{1}{\alpha_0}, \quad H_j = \frac{\nu_{0j}^2}{\beta_0^2 a_2^2} - \frac{1}{\beta_0},$$

$$H_1' = \frac{2\nu_{01}^2}{\alpha_0^3 a_1^3} + \frac{2}{\alpha_0^2}, \quad H_j' = \frac{2\nu_{0j}^2}{\beta_0^3 a_2^3} + \frac{2}{\beta_0^2}, \quad j=2,\ldots,p. \qquad (2.84)$$

The c.f. of V given that $\underline{X}$ comes from π_2, after collecting terms, is

$$\phi(t/\pi_2) = \{1 + it \sum_{k=1}^{2} \frac{M_{1k}}{(1-2ita_k/(a_k+1))}$$

$$+ (it)^2 \sum_{k=1}^{2} \frac{M_{2k}}{(1-2ita_k/(a_k+1))^2} + (it)^3 \sum_{k=1}^{2} \frac{M_{3k}}{(1-2ita_k/(a_k+1))^3}$$

$$+ (it)^4 \sum_{k=1}^{2} \frac{M_{4k}}{(1-2ita_k/(a_k+1))^4}\}\psi_0 + O_2, \tag{2.85}$$

where

$$M_{11} = \frac{1}{N_1 a_1} + \frac{1}{N_2\alpha_0^2 a_1} + \frac{1}{n_1} C_1' + \frac{1}{n_2} \alpha_0^2 H_1' ,$$

$$M_{12} = \frac{(p-1)}{N_1 a_2} + \frac{(p-1)}{N_2\beta_0^2 a_2} + \frac{1}{n_1} \sum_{j=2}^{p} C_j' + \frac{1}{n_2} \beta_0^2 \sum_{j=2}^{p} H_j' ,$$

$$M_{21} = \frac{2\nu_{01}^2}{N_1 a_1^2} + \frac{2\nu_{01}^2}{N_2\alpha_0^2 a_1^2} + \frac{1}{n_1}(C_1^2+2\alpha_0 C_1+B_1') + \frac{1}{n_2}(\alpha_0^2 H_1^2-2\alpha_0 H_1-\alpha_0^2 G_1'),$$

$$M_{22} = \frac{2\sum_{j=2}^{p}\nu_{0j}^2}{N_1 a_2^2} + \frac{2\sum_{j=2}^{p}\nu_{0j}^2}{N_2\beta_0^2 a_2^2} + \frac{1}{n_1}\sum_{j=2}^{p}(C_j^2+2\beta_0 C_j+B_j')$$

$$+ \frac{1}{n_2}\sum_{j=2}^{p}(\beta_0^2 H_j^2-2\beta_0 H_j-\beta_0^2 G_j'),$$

$$M_{31} = \frac{1}{n_1}(4\alpha_0 B_1+2B_1 C_1) + \frac{1}{n_2}(4\alpha_0 G_1-2\alpha_0^2 G_1 H_1),$$

$$M_{32} = \frac{1}{n_1} \sum_{j=2}^{p} (4\beta_0 B_j + 2B_j C_j) + \frac{1}{n_2} \sum_{j=2}^{p} (4\beta_0 G_j - 2\beta_0^2 G_j H_j),$$

$$M_{41} = \frac{1}{n_1} B_1^2 + \frac{1}{n_2} \alpha_0^2 G_1^2,$$

$$M_{42} = \frac{1}{n_1} \sum_{j=2}^{p} B_j^2 + \frac{1}{n_2} \beta_0^2 \sum_{j=2}^{p} G_j^2 . \qquad (2.86)$$

The c.d.f. $F_2(v)$ is obtained by inverting $\phi(t/\pi_2)$. Let $G_{m,k}(x)$ be the c.d.f. of a non-central chi-square variate with m degrees of freedom and non-centrality parameter $\frac{\nu_{01}^2}{\alpha_0 a_1^2}$ when k=1 and $\frac{\sum_{j=2}^{p} \nu_{0j}^2}{\beta_0 a_2^2}$ when k=2. Employing the same method used in obtaining $F_1(v)$, we have

$$F_2(v) = G_{1,1}(v) + G_{(p-1),2}(v) - M_{11}G_{3,1}^{(1)}(v) - M_{12}G_{3(p-1),2}^{(1)}(v)$$

$$+ M_{21}G_{5,1}^{(2)}(v) + M_{22}G_{5(p-1),2}^{(2)}(v) - M_{31}G_{7,1}^{(3)}(v) - M_{32}G_{7(p-1),2}^{(3)}(v)$$

$$+ M_{41}G_{9,1}^{(4)}(v) + M_{42}G_{9(p-1),2}^{(4)}(v) + O_2 , \qquad (2.87)$$

where $G_{m,k}^{(r)}(v)$ is the rth derivative of $G_{m,k}(v)$.

III. DISTRIBUTION OF THE DISCRIMINANT FUNCTION FOR P ADJUSTED (FOR Q COVARIATES) DISCRIMINATORS UNDER INTRACLASS CORRELATION MODELS

A. Introduction

In this chapter, we assume that, in addition to p discriminators, q covariates are available for the discriminant function. Though the covariates do not have discriminating power by themselves, it is demonstrated that it is advantageous to use them in linear discriminant function (see, for example, Cochran and Bliss, 1948 and Cochran, 1964a). We shall show in this thesis that it is also advantageous to use covariates in the case of quadratic discriminant function with intraclass correlation structure.

Assume a ((p+q)x1) observation vector $\underline{X}$ is of the form $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$, where $\underline{X}_1 = (X_1, X_2, \ldots, X_p)'$ and $\underline{X}_2 = (X_{p+1}, X_{p+2}, \ldots, X_{p+q})'$. Suppose $\underline{X}$ belongs to one of two populations π_i: $N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(i)}\right)$, where $\underline{\mu}^{(i)}$ is a (px1) vector, $\underline{\mu}$ is a (qx1) vector, and

$$\underline{\Sigma}^{(i)} = \begin{bmatrix} \sigma_i^2[(1-\rho_i)\underline{I}+\rho_i\underline{J}] & \sigma_i\sigma\rho_i'\underline{J} \\ \sigma_i\sigma\rho_i'\underline{J} & \sigma^2[(1-\rho)\underline{I}+\rho\underline{J}] \end{bmatrix} = \begin{bmatrix} \underline{\Sigma}_{11}^{(i)} & \underline{\Sigma}_{12}^{(i)} \\ \underline{\Sigma}_{21}^{(i)} & \underline{\Sigma}_{22} \end{bmatrix}, \quad (3.1)$$

for i=1,2, where $\sigma_1^2 \neq \sigma_2^2$. For i=1,2, the matrix

$\underline{\Sigma}_{11}^{(i)} = \sigma_i^2[(1-\rho_i)\underline{I}+\rho_i\underline{J}]$ is a (pxp) positive-definite matrix,

$\underline{\Sigma}_{12}^{(i)} = \sigma_i\sigma\rho_i'\underline{J}$ is a (pxq) matrix, $\underline{\Sigma}_{21}^{(i)} = [\underline{\Sigma}_{12}^{(i)}]'$, and

$\underline{\Sigma}_{22} = \sigma^2[(1-\rho)\underline{I}+\rho\underline{J}]$ is a (qxq) positive-definite matrix.

The p variates, $\underline{X}_1$, are discriminators and the q variates, $\underline{X}_2$, are covariates. Notice that $\underline{X}_2$ has the same mean and the same covariance matrix in both populations. Hence $\underline{X}_2$ has no discriminating power by itself.

In Chapter II, we ignored $\underline{X}_2$ completely. The rule for classifying an observation in Chapter II was based only on $\underline{X}_1$ (written as $\underline{X}$), where $\underline{X}_1 \sim N_p(\underline{\mu}^{(i)}, \underline{\Sigma}_{11}^{(i)})$ when the observation belongs to π_i, i=1,2. The matrix $\underline{\Sigma}_{11}^{(i)}$ (written as $\underline{\Sigma}^{(i)}$ in Chapter II) is given in (3.1) for i=1,2.

In this chapter, we shall make use of the information provided by $\underline{X}_2$, by adjusting the discriminators, $\underline{X}_1$. Since the regression coefficient matrices $\underline{B}^{(1)} = \underline{\Sigma}_{12}^{(1)}\underline{\Sigma}_{22}^{-1}$ and $\underline{B}^{(2)} = \underline{\Sigma}_{12}^{(2)}\underline{\Sigma}_{22}^{-1}$ for the two respective populations π_1 and π_2 are, in general, not equal, a natural procedure is to form the variates $\underline{Z}_i$ where

$$\underline{Z}_i = \underline{X}_1 - \underline{B}^{(i)}\underline{X}_2 = \underline{X}_1 - \underline{\Sigma}_{12}^{(i)}\underline{\Sigma}_{22}^{-1}\underline{X}_2, \tag{3.2}$$

when $\underline{X}\varepsilon\pi_i$, i=1,2. Ordinarily, one does not know to which population π_i the observation belongs. Using the variates $\underline{Z}_i$ given in (3.2), one could form the weighted variates $\underline{Z}$ where

$$\underline{Z} = W_1\underline{Z}_1 + W_2\underline{Z}_2 \tag{3.3}$$

and W_i are weights such that $W_1 + W_2 = 1$. Therefore, the variates $\underline{Z}$ are a weighted sum of the variates $\underline{Z}_i$. In terms of the observation vector $\underline{X}$, (3.3) can be written as

$$\underline{Z} = \underline{X}_1 - (W_1\underline{\Sigma}_{12}^{(1)} + W_2\underline{\Sigma}_{12}^{(2)})\underline{\Sigma}_{22}^{-1}\underline{X}_2 \ . \tag{3.4}$$

As mentioned previously, Cochran and Bliss (1948) and Cochran (1964a) considered a similar situation when both populations have the same covariance matrix. For their situation, the adjusted variates are given by $\underline{Z} = \underline{X}_1 - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}(\underline{X}_2-\underline{\mu})$.

Memon and Okamoto (1970) considered the problem of classifying an observation $\begin{pmatrix}\underline{X}_1\\ \underline{X}_2\end{pmatrix}$ into one of two populations π_i: $N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(i)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}\right)$, i=1,2, where $\underline{\Sigma}$ is an arbitrary positive-definite matrix. They obtained an asymptotic expansion of the distribution of the classification statistic W* given by (1.13) in Chapter I.

In Section B of this chapter, the discriminant function for the adjusted variates $\underline{Z}$, (3.3) or (3.4), is obtained using the likelihood ratio procedure. The distribution of the discriminant function is given when all parameters are known. The distribution of the discriminant function when the mean vectors are unknown, but the covariance matrices are known is obtained in Section C. In Section D, the limiting distribution is found for the discriminant function when all parameters are unknown.

B. Parameters Known

Let $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$ be a $((p+q)x1)$ observation vector where $\underline{X}_1$ is a $(px1)$ vector and $\underline{X}_2$ is a $(qx1)$ vector such that $\underline{X} \sim N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(i)}\right)$ when $\underline{X}$ belongs to π_i, $i=1,2$. We assume in this section that all parameters are known.

Let $\underline{Z} = W_1\underline{Z}_1 + W_2\underline{Z}_2$ where $\underline{Z}_i$ are given in (3.2) and where W_i are known constant weights with $W_1+W_2 = 1$. Let $\underline{\Sigma}^{(i)}_{11.2} = \underline{\Sigma}^{(i)}_{11} - \underline{\Sigma}^{(i)}_{12}\underline{\Sigma}^{-1}_{22}\underline{\Sigma}^{(i)}_{21}$, $i=1,2$. If $\underline{X}\epsilon\pi_1$, then

$$
\begin{aligned}
E(\underline{Z}) &= W_1E(\underline{Z}_1)+W_2E(\underline{Z}_2) \\
&= \underline{\mu}^{(1)} - (W_1\underline{\Sigma}^{(1)}_{12} + W_2\underline{\Sigma}^{(2)}_{12})\underline{\Sigma}^{-1}_{22}\underline{\mu} = \underline{\mu}_Z^{(1)}
\end{aligned} \tag{3.5}
$$

and

$$
\begin{aligned}
E(\underline{Z}-E(\underline{Z}))(\underline{Z}-E(\underline{Z}))' &= W_1^2E(\underline{Z}_1-E(\underline{Z}_1))(\underline{Z}_1-E(\underline{Z}_1))' \\
&\quad + 2W_1W_2E(\underline{Z}_1-E(\underline{Z}_1))(\underline{Z}_2-E(\underline{Z}_2))' \\
&\quad + W_2^2E(\underline{Z}_2-E(\underline{Z}_2))(\underline{Z}_2-E(\underline{Z}_2))' \\
&= W_1^2\underline{\Sigma}^{(1)}_{11.2} + 2W_1W_2\underline{\Sigma}^{(1)}_{11.2} + W_2^2[\underline{\Sigma}^{(1)}_{11} + \underline{\Sigma}^{(2)}_{12}\underline{\Sigma}^{-1}_{22}(\underline{\Sigma}^{(2)}_{21} - 2\underline{\Sigma}^{(1)}_{21})] \\
&= W_1(W_1+2W_2)\{\sigma_1^2[(1-\rho_1)\underline{I} + (\rho_1 - \frac{\rho_1'^2 q}{1+(q-1)\rho})\underline{J}]_{pxp}\} \\
&\quad + W_2^2\{\sigma_1^2[(1-\rho_1)\underline{I} + \rho_1\underline{J}]_{pxp}
\end{aligned}
$$

$$+ \sigma_2\sigma\rho_2'\underline{J}_{pxq}\ \frac{1}{\sigma^2(1-\rho)}\ [\underline{I} - \frac{\rho}{1+(q-1)\rho}\underline{J}]_{qxq}$$

$$\times\ [\sigma_2\sigma\rho_2'-2\sigma_1\sigma\rho_1']\underline{J}_{pxq}\}$$

$$= W_1(W_1+2W_2)\{\sigma_1^2[(1-\rho_1)\underline{I}+(\rho_1-\frac{\rho_1'^2 q}{1+(q-1)\rho})\underline{J}]_{pxp}\}$$

$$+ W_2^2\{\sigma_1^2[(1-\rho_1)\underline{I}+\rho_1\underline{J}]_{pxp} + \frac{\sigma_2\rho_2'q(\sigma_2\rho_2'-2\sigma_1\rho_1')}{1+(q-1)\rho}\underline{J}_{pxp}\}$$

$$= \sigma_1^2(1-\rho_1)\underline{I}_{pxp}+\{\sigma_1^2\rho_1-\frac{1}{1+(q-1)\rho}[W_1(W_1+2W_2)\sigma_1^2\rho_1'^2 q$$

$$-W_2^2\sigma_2\rho_2'q(\sigma_2\rho_2'-2\sigma_1\rho_1')]\}\underline{J}_{pxp}$$

$$= (a_1-b_1)\underline{I} + b_1\underline{J} = \underline{V}^{(1)}, \tag{3.6}$$

where

$$a_1 = \sigma_1^2-B_1,$$

$$b_1 = \sigma_1^2\rho_1-B_1,$$

and

$$B_1 = \frac{1}{1+(q-1)\rho}\{W_1(W_1+2W_2)\sigma_1^2\rho_1'^2 q-W_2^2[\sigma_2\rho_2'q(\sigma_2\rho_2'-2\sigma_1\rho_1')]\}. \tag{3.7}$$

Similarly, if $\underline{X}\varepsilon\pi_2$, then

$$E(\underline{Z}) = \underline{\mu}^{(2)}-(W_1\underline{\Sigma}_{12}^{(1)}+W_2\underline{\Sigma}_{12}^{(2)})\underline{\Sigma}_{22}^{-1}\underline{\mu} = \underline{\mu}_Z^{(2)} \tag{3.8}$$

and

$$E(\underline{Z}-E(\underline{Z}))(\underline{Z}-E(\underline{Z}))'=W_1^2[\underline{\Sigma}_{11}^{(2)}+\underline{\Sigma}_{12}^{(1)}\underline{\Sigma}_{22}^{-1}(\underline{\Sigma}_{21}^{(1)}-2\underline{\Sigma}_{21}^{(2)})]$$

$$+\ 2W_1W_2\underline{\Sigma}_{11.2}^{(2)}+W_2^2\underline{\Sigma}_{11.2}^{(2)}$$

$$= W_1^2\{\sigma_2^2[(1-\rho_2)\underline{I}+\rho_2\underline{J}]_{pxp}+\sigma_1\sigma\rho_1'\underline{J}_{pxq}\ \frac{1}{\sigma^2(1-\rho)}[\underline{I}-\ \frac{\rho}{1+(q-1)\rho}\underline{J}]_{qxq}$$

$$\times\ [\sigma_1\sigma\rho_1'-2\sigma_2\sigma\rho_2']\underline{J}_{pxp}\}$$

$$+\ W_2(W_2+2W_1)\{\sigma_2^2[(1-\rho_2)\underline{I}+(\rho_2-\ \frac{\rho_2'^2q}{1+(q-1)\rho})\underline{J}]_{pxp}\}$$

$$= \sigma_2^2(1-\rho_2)\underline{I}_{pxp}+\{\sigma_2^2\rho_2-\ \frac{1}{1+(q-1)\rho}[-W_1^2\sigma_1\rho_1'q(\sigma_1\rho_1'-2\sigma_2\rho_2')$$

$$+\ W_2(W_2+2W_1)\sigma_2^2\rho_2'^2q]\}\underline{J}_{pxp}$$

$$= (a_2-b_2)\underline{I}\ +\ b_2\underline{J}\ =\ \underline{V}^{(2)}, \qquad (3.9)$$

where

$$a_2\ =\ \sigma_2^2-B_2,$$

$$b_2\ =\ \sigma_2^2\rho_2-B_2,$$

and

$$B_2\ =\ \frac{1}{1+(q-1)\rho}\{-W_1^2\sigma_1\rho_1'q(\sigma_1\rho_1'-2\sigma_2\rho_2')+W_2(W_2+2W_1)\sigma_2^2\rho_2'^2q\}. \qquad (3.10)$$

The covariance matrices $\underline{V}^{(1)}$ and $\underline{V}^{(2)}$ given above have intraclass correlation structure. Let $\underline{\Gamma}$ be a (pxp) orthogonal matrix with first row $p^{-\frac{1}{2}}\underline{e}'$ where $\underline{e}$ is a (pxl) vector of ones. Then $\underline{\Gamma}$ simultaneously diagonalizes $\underline{V}^{(1)}$ and $\underline{V}^{(2)}$, that is,

$$\underline{\Gamma}\ \underline{V}^{(1)}\underline{\Gamma}' = \underline{D}^{(1)}, \quad \underline{\Gamma}\ \underline{V}^{(2)}\underline{\Gamma}' = \underline{D}^{(2)}, \tag{3.11}$$

where $\underline{D}^{(i)} = \mathrm{diag}(\alpha_i, \beta_i, \ldots, \beta_i)$, and

$$\alpha_i = a_i + (p-1)b_i = \sigma_i^2[1+(p-1)\rho_i] - pB_i,$$

$$\beta_i = a_i - b_i = \sigma_i^2(1-\rho_i), \quad i=1,2. \tag{3.12}$$

Since $\underline{\Gamma}$ is independent of the elements in $\underline{V}^{(1)}$ and $\underline{V}^{(2)}$, the discriminant function is equivalent to that when the covariance matrices are diagonal. This is true because the discriminant function is invariant under any orthogonal transformation as shown previously.

Let $\underline{Y} = \underline{\Gamma}\ \underline{Z}$ so $\underline{Y} \sim N_p(\underline{\nu}^{(i)}, \underline{D}^{(i)})$

when $\underline{X}$ belongs to π_i where

$$\underline{\nu}^{(i)} = \underline{\Gamma}[\underline{\mu}^{(i)} - (W_1\underline{\Sigma}_{12}^{(1)} + W_2\underline{\Sigma}_{12}^{(2)})\underline{\Sigma}_{22}^{-1}\underline{\mu}] = \underline{\Gamma}\ \underline{\mu}_Z^{(i)}, \tag{3.13}$$

with $\underline{\mu}_Z^{(i)}$ given in (3.5) or (3.8) and $\underline{D}^{(i)} = \mathrm{diag}(\alpha_i, \beta_i, \ldots, \beta_i)$ with α_i, β_i, given in (3.12) for i=1,2.

The discriminant function for the adjusted variates $\underline{Z}$, using the likelihood ratio procedure, is

$$U = (\underline{Z}-\underline{\mu}_Z^{(2)})'[\underline{V}^{(2)}]^{-1}(\underline{Z}-\underline{\mu}_Z^{(2)}) - (\underline{Z}-\underline{\mu}_Z^{(1)})'[\underline{V}^{(1)}]^{-1}(\underline{Z}-\underline{\mu}_Z^{(1)}). \tag{3.14}$$

The discriminant function U in (3.14) can be written in terms of $\underline{Y}$ as

$$U = (\underline{Y}-\underline{\nu}^{(2)})'[\underline{D}^{(2)}]^{-1}(\underline{Y}-\underline{\nu}^{(2)})-(\underline{Y}-\underline{\nu}^{(1)})'[\underline{D}^{(1)}]^{-1}(\underline{Y}-\underline{\nu}^{(1)}). \tag{3.15}$$

Substituting $\underline{\nu}^{(i)}$ and $\underline{D}^{(i)}$, we obtain, apart from a constant,

$$V= \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\left(Y_1- \frac{\frac{\nu_1^{(2)}}{\alpha_2} - \frac{\nu_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right)^2+\left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\sum_{j=2}^{p}\left(Y_j- \frac{\frac{\nu_j^{(2)}}{\beta_2} - \frac{\nu_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right)^2, \tag{3.16}$$

where Y_j and $\nu_j^{(i)}$ are the jth components of $\underline{Y}$ and $\underline{\nu}^{(i)}$, i=1,2, respectively. We shall classify $\underline{X}$ into π_1 if $V\geq c$ and into π_2 if $V<c$ for some suitable choice of the constant c.

To find the distribution of V in (3.16), we shall assume that $\alpha_1>\alpha_2$, $\beta_1>\beta_2$, or equivalently that $\underline{V}^{(1)}-\underline{V}^{(2)}$ is positive-definite. Since V is in the form given by (2.10), we can carry out the procedure of obtaining the distribution of the discriminant function in precisely the same manner as for the case of known parameters in Chapter II. Let

$$Z_1 =\sqrt{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\left(Y_1- \frac{\frac{\nu_1^{(2)}}{\alpha_2} - \frac{\nu_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right),$$

$$Z_j =\sqrt{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\left(Y_j- \frac{\frac{\nu_j^{(2)}}{\beta_2} - \frac{\nu_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right), \quad j=2,\dots,p, \tag{3.17}$$

where α_i, β_i are given in (3.12) and $\underline{\nu}^{(i)}$ are given in (3.13). Then $V = \sum_{j=1}^{p} Z_j^2$. When $\underline{X}$ comes from π_i, i=1 or 2, Z_j are independently distributed as $N(\xi_j^{(i)}, \tau_{ij}^2)$ where

$$\xi_1^{(i)} = \sqrt{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\,\left(\nu_1^{(i)} - \frac{\frac{\nu_1^{(2)}}{\alpha_2} - \frac{\nu_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right),$$

$$\tau_{i1}^2 = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\alpha_i \quad \text{for} \quad i=1,2;$$

$$\xi_j^{(i)} = \sqrt{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\,\left(\nu_j^{(i)} - \frac{\frac{\nu_j^{(2)}}{\beta_2} - \frac{\nu_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right),$$

$$\tau_{ij}^2 = \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\beta_i \quad \text{for} \quad j=2,\ldots,p; \quad i=1,2. \tag{3.18}$$

Therefore, we have

$$V \sim \sum_{j=1}^{p} \tau_{ij}^2 \chi_1'^2(\delta_{ij}^2) , \tag{3.19}$$

where $\delta_{ij}^2 = \dfrac{[\xi_j^{(i)}]^2}{\tau_{ij}^2}$, for i=1,2.

The c.d.f. of V in (3.16) could be found by using Press' method as explained in Chapter II. Also the distribution of V could be approximated by either Patnaik's or Pearson's approximation.

C. Means Unknown, Covariance Matrices Known

In this section, we shall assume that $\begin{pmatrix}\underline{\mu}^{(1)}\\ \underline{\mu}\end{pmatrix}$ and $\begin{pmatrix}\underline{\mu}^{(2)}\\ \underline{\mu}\end{pmatrix}$ are unknown, but $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$ are known. Suppose that we have a random sample $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \ldots, \underline{X}_{N_1}^{(1)}$ from $\pi_1: N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(1)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}^{(1)}\right)$ and an independent random sample $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \ldots, \underline{X}_{N_2}^{(2)}$ from $\pi_2: N_{p+q}\left(\begin{pmatrix}\underline{\mu}^{(2)}\\ \underline{\mu}\end{pmatrix}, \underline{\Sigma}^{(2)}\right)$. Our estimate of $\begin{pmatrix}\underline{\mu}^{(i)}\\ \underline{\mu}\end{pmatrix}$ is the sample mean $\begin{pmatrix}\overline{\underline{X}}_1^{(i)}\\ \overline{\underline{X}}_2^{(i)}\end{pmatrix}$ where

$$\overline{\underline{X}}_1^{(i)} = \frac{1}{N_i}\sum_{\alpha=1}^{N_i} \underline{X}_{1\alpha}^{(i)}, \quad \overline{\underline{X}}_2^{(i)} = \frac{1}{N_i}\sum_{\alpha=1}^{N_i} \underline{X}_{2\alpha}^{(i)}, \quad i=1,2.$$

Then our estimate of $\underline{\mu}_Z^{(i)}$, (3.5) or (3.8), is given by

$$\overline{\underline{Z}}^{(i)} = W_1[\overline{\underline{X}}_1^{(i)} - \underline{\Sigma}_{12}^{-1}\underline{\Sigma}_{22}^{-1}\overline{\underline{X}}_2^{(i)}] + W_2[\overline{\underline{X}}_1^{(i)} - \underline{\Sigma}_{12}^{(2)}\underline{\Sigma}_{22}^{-1}\overline{\underline{X}}_2^{(i)}]$$

$$= \overline{\underline{X}}_1^{(i)} - [W_1\underline{\Sigma}_{12}^{(1)} + W_2\underline{\Sigma}_{12}^{(2)}]\underline{\Sigma}_{22}^{-1}\overline{\underline{X}}_2^{(i)} \quad . \tag{3.20}$$

The sample means $\overline{\underline{Z}}^{(i)}$ are independently distributed as $N(\underline{\mu}_Z^{(i)}, \frac{1}{N_i}\underline{V}^{(i)})$, $i=1,2$, where $\underline{V}^{(i)}$ are given in (3.6) and (3.9). The sample means are also independent of the distribution of $\underline{Z}$. If $\underline{Y} = \underline{\Gamma}\ \underline{Z}$ where $\underline{\Gamma}$ is the (pxp) orthogonal matrix given in Section B, then we estimate

$\underline{\nu}^{(i)} = \underline{\Gamma}\,\underline{\mu}_Z^{(i)}$ by $\overline{\underline{Y}}^{(i)} = \underline{\Gamma}\,\overline{\underline{Z}}^{(i)}$, i=1,2. The sample means $\overline{\underline{Y}}^{(i)}$ are independently distributed as $N(\underline{\nu}^{(i)}, \frac{1}{N_i}\underline{D}^{(i)})$, i=1,2, and independent of the distribution of $\underline{Y}$. Further, since the $\underline{D}^{(i)}$ are diagonal matrices, the components of $\overline{\underline{Y}}^{(i)}$ are also independently distributed. The discriminant function V in (3.16), after substituting for the unknown parameters by their respective estimates, becomes

$$V = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\left(Y_1 - \frac{\frac{\overline{Y}_1^{(2)}}{\alpha_2} - \frac{\overline{Y}_1^{(1)}}{\alpha_1}}{\frac{1}{\alpha_2} - \frac{1}{\alpha_1}}\right)^2$$

$$+ \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\sum_{j=2}^{p}\left(Y_j - \frac{\frac{\overline{Y}_j^{(2)}}{\beta_2} - \frac{\overline{Y}_j^{(1)}}{\beta_1}}{\frac{1}{\beta_2} - \frac{1}{\beta_1}}\right)^2, \qquad (3.21)$$

where Y_j and $\overline{Y}_j^{(i)}$ are the jth components of $\underline{Y}$ and $\overline{\underline{Y}}^{(i)}$ and α_i, β_i are given in (3.12), i=1,2, respectively. Since V in (3.21) is similar to the discriminant function in (2.31), one can proceed in the same manner as in Chapter II-C.

D. Parameters Unknown

In this section, we assume all parameters $\begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu} \end{pmatrix}$, $\begin{pmatrix} \underline{\mu}^{(2)} \\ \underline{\mu} \end{pmatrix}$, $\underline{\Sigma}^{(1)}$, and $\underline{\Sigma}^{(2)}$ are unknown. Suppose that we have a random sample $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \ldots, \underline{X}_{N_1}^{(1)}$ from $\pi_1: N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(1)}\right)$ and an independent random sample $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \ldots, \underline{X}_{N_2}^{(2)}$ from $\pi_2: N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(2)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(2)}\right)$. Our estimate of $\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}$ is the sample mean $\overline{\underline{X}}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \underline{X}_\alpha^{(i)}$ and of $\underline{\Sigma}^{(i)}$ is the sample variance

$$\underline{S}^{(i)} = \frac{1}{n_i} \sum_{\alpha=1}^{N_i} (\underline{X}_\alpha^{(i)} - \overline{\underline{X}}^{(i)})(\underline{X}_\alpha^{(i)} - \overline{\underline{X}}^{(i)})', \tag{3.22}$$

where $n_i = N_i - 1$, $i=1,2$.

In Sections B and C, we did not specify any values for the known constant weights W_i, $i=1,2$. If $\underline{\Sigma}^{(1)}$, $\underline{\Sigma}^{(2)}$ are known, one possible choice for the weights is

$$W_1 = \frac{|\underline{\Sigma}_{11.2}^{(2)}|}{|\underline{\Sigma}_{11.2}^{(1)}| + |\underline{\Sigma}_{11.2}^{(2)}|}$$

and

$$W_2 = \frac{|\underline{\Sigma}_{11.2}^{(1)}|}{|\underline{\Sigma}_{11.2}^{(1)}| + |\underline{\Sigma}_{11.2}^{(2)}|}, \tag{3.23}$$

where

$\underline{\Sigma}_{11.2}^{(i)} = \underline{\Sigma}_{11}^{(i)} - \underline{\Sigma}_{12}^{(i)} \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21}^{(i)}$ for i=1,2. These are the weights used in Chapter V for the two bivariate (p=1, q=1) normal populations and for the two trivariate (p=1, q=2) normal populations.

If the parameters are unknown, the estimates for W_1 and W_2 shall be defined, respectively, as

$$\tilde{W}_1 = \frac{|\underline{S}_{11.2}^{(2)}|}{|\underline{S}_{11.2}^{(1)}| + |\underline{S}_{11.2}^{(2)}|}$$

and

$$\tilde{W}_2 = \frac{|\underline{S}_{11.2}^{(1)}|}{|\underline{S}_{11.2}^{(1)}| + |\underline{S}_{11.2}^{(2)}|}, \tag{3.24}$$

where $\underline{S}_{11.2}^{(i)} = \underline{S}_{11}^{(i)} - \underline{S}_{12}^{(i)} \underline{S}_{22}^{-1} \underline{S}_{21}^{(i)}$ and $\underline{S}_{22} = \dfrac{n_1 \underline{S}_{22}^{(1)} + n_2 \underline{S}_{22}^{(2)}}{n_1 + n_2}$.

The discriminant function (3.16), for the adjusted variates $\underline{Y}$, when all parameters are known is

$$V = \left(\frac{1}{\alpha_2} - \frac{1}{\alpha_1}\right)\left(Y_1 - \frac{\dfrac{\nu_1^{(2)}}{\alpha_2} - \dfrac{\nu_1^{(1)}}{\alpha_1}}{\dfrac{1}{\alpha_2} - \dfrac{1}{\alpha_1}}\right)^2 + \left(\frac{1}{\beta_2} - \frac{1}{\beta_1}\right)\sum_{j=2}^{p}\left(Y_j - \frac{\dfrac{\nu_j^{(2)}}{\beta_2} - \dfrac{\nu_j^{(1)}}{\beta_1}}{\dfrac{1}{\beta_2} - \dfrac{1}{\beta_1}}\right)^2.$$

when all parameters are unknown, let V* denote the discriminant function obtained if we replace any unknown parameters by their respective estimates in the discriminant function V. Hence

$$V^* = \left(\frac{1}{\tilde{\alpha}_2^*} - \frac{1}{\tilde{\alpha}_1^*}\right)\left(Y_1^* - \frac{\dfrac{\bar{Y}_1^{*(2)}}{\tilde{\alpha}_2^*} - \dfrac{\bar{Y}_1^{*(1)}}{\tilde{\alpha}_1^*}}{\dfrac{1}{\tilde{\alpha}_2^*} - \dfrac{1}{\tilde{\alpha}_1^*}}\right)^2$$

$$+ \left(\frac{1}{\tilde{\beta}_2^*} - \frac{1}{\tilde{\beta}_1^*}\right) \sum_{j=2}^{p} \left(Y_j^* - \frac{\dfrac{\bar{Y}_j^{*(2)}}{\tilde{\beta}_2^*} - \dfrac{\bar{Y}_j^{*(1)}}{\tilde{\beta}_1^*}}{\dfrac{1}{\tilde{\beta}_2^*} - \dfrac{1}{\tilde{\beta}_1^*}}\right)^2, \qquad (3.25)$$

where

$$\underline{Y}^* = (Y_1^*, Y_2^*, \ldots, Y_p^*)' = \underline{\Gamma}\ \underline{Z}^* = \underline{\Gamma}[\underline{X}_1 - (\tilde{W}_1 \underline{S}_{12}^{(1)} + \tilde{W}_2 \underline{S}_{12}^{(2)}) \underline{S}_{22}^{-1} \underline{X}_2], \qquad (3.26)$$

$$\bar{\underline{Y}}^{*(i)} = (\bar{Y}_1^{*(i)}, \bar{Y}_2^{*(i)}, \ldots, \bar{Y}_p^{*(i)})' = \underline{\Gamma}\ \bar{\underline{Z}}^{*(i)}$$

$$= \underline{\Gamma}[\bar{\underline{X}}_1^{(i)} - (\tilde{W}_1 \underline{S}_{12}^{(1)} + \tilde{W}_2 \underline{S}_{12}^{(2)}) \underline{S}_{22}^{-1} \bar{\underline{X}}_2^{(i)}], \qquad (3.27)$$

and

$$\tilde{\alpha}_i^* = \frac{\sum_{\alpha=1}^{N_i} (Y_{1\alpha}^{*(i)} - \bar{Y}_1^{*(i)})^2}{n_i},$$

$$\tilde{\beta}_i^* = \frac{\sum_{j=2}^{p} \sum_{\alpha=1}^{N_i} (Y_{j\alpha}^{*(i)} - \bar{Y}_j^{*(i)})^2}{(p-1) n_i}, \quad i=1,2. \qquad (3.28)$$

We have as N_1, $N_2 \to \infty$

$$\text{plim}\ \bar{\underline{Y}}^{*(i)} = \underline{\nu}^{(i)},$$

$$\text{plim}\ \tilde{\alpha}_i^* = \alpha_i,$$

$$\text{plim}\ \tilde{\beta}_i^* = \beta_i, \quad i=1,2, \tag{3.29}$$

where "plim" denotes convergence in probability (Cramér, 1946, p. 255). Also, we know from (3.26) that the distribution of $\underline{Y}^*$ tends to that of $\underline{Y}$ when N_1, $N_2 \to \infty$. Therefore, from Cramér's Theorem (Cramér, 1946, p. 254) for stochastic convergence we find that the distribution of V^* tends to that of V, as N_1, $N_2 \to \infty$ in such a way that $\frac{N_2}{N_1} \to$ constant. Consequently, for sufficiently large samples from π_1 and π_2, we can replace any unknown parameter values by their respective estimates and use the discriminant function V^* in the classification procedure.

IV. DISTRIBUTION OF THE DISCRIMINANT FUNCTION FOR P+Q UNADJUSTED DISCRIMINATORS UNDER INTRACLASS CORRELATION MODELS

A. Introduction

As in Chapter III, assume $\underline{X}$ is a ((p+q)x1) observation vector of the form $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$, where $\underline{X}_1 = (X_1, X_2, \ldots, X_p)'$are considered as discriminators and $\underline{X}_2 = (X_{p+1}, X_{p+2}, \ldots, X_{p+q})'$ are viewed as covariates. Suppose $\underline{X}$ belongs to one of two populations π_i: $N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(i)}\right)$, where $\underline{\mu}^{(i)}$ is a (px1) vector, $\underline{\mu}$ is a (qx1) vector, and $\underline{\Sigma}^{(i)}$ has the form (3.1) for i=1,2. Since $\underline{X}_2$ has the same mean and the same covariance matrix in both populations, it has no discriminating power by itself.

In Chapter III, the discriminators $\underline{X}_1$ were adjusted for these covariates, $\underline{X}_2$. In Chapter II, the covariates $\underline{X}_2$ were ignored. In this chapter, we shall include the q covariates $\underline{X}_2$ in the discriminant function in exactly the same way as the p discriminators, $\underline{X}_1$. Hence $\underline{X}$ will be viewed as p+q discriminators.

In Section B, the discriminant function for the p+q variates $\underline{X}$ is given using the likelihood ratio procedure. The distribution of the discriminant function is given when all parameters are known. We obtain in Section C the distribution of the discriminant function when the mean vectors are unknown, but the covariance matrices are known. In Section D, the limiting distribution is found for the discriminant function when all parameters are unknown.

B. Parameters Known

Let $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$ be a $((p+q)\times 1)$ observation vector where $\underline{X}_1$ is a $(p\times 1)$ vector and $\underline{X}_2$ is a $(q\times 1)$ vector such that $\underline{X} \sim N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(i)}\right)$ when $\underline{X}$ belongs to π_i, $i=1,2$. We assume in this section that $\underline{\mu}^{(i)}$, $\underline{\mu}$, and $\underline{\Sigma}^{(i)}$ have the structure given in Section A and that all parameters are known. Since $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$ are positive-definite matrices, there exists a non-singular matrix $\underline{P}$ such that

$$\underline{P}'\underline{\Sigma}^{(1)}\underline{P} = \underline{I}, \quad \underline{P}'\underline{\Sigma}^{(2)}\underline{P} = \underline{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_{p+q}), \tag{4.1}$$

where the λ_j's are the roots of $|\underline{\Sigma}^{(2)} - \lambda\underline{\Sigma}^{(1)}| = 0$. The eigenvalues λ_j are all positive since both $\underline{\Sigma}^{(1)}$ and $\underline{\Sigma}^{(2)}$ are positive-definite.

In the known parameter case, the logarithm of the likelihood ratio procedure is proportional to

$$U = (\underline{X}-\underline{\mu}_X^{(2)})'[\underline{\Sigma}^{(2)}]^{-1}(\underline{X}-\underline{\mu}_X^{(2)}) - (\underline{X}-\underline{\mu}_X^{(1)})'[\underline{\Sigma}^{(1)}]^{-1}(\underline{X}-\underline{\mu}_X^{(1)}), \tag{4.2}$$

where $\underline{\mu}_X^{(i)} = \begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}$, $i=1,2$. Let $\underline{Y} = \underline{P}'\underline{X}$. Since the discriminant function U in (4.2) is invariant under any non-singular linear transformation, the discriminant function can be written in terms of $\underline{Y}$ as

$$U = (\underline{Y}-\underline{\nu}^{(2)})'\underline{\Lambda}^{-1}(\underline{Y}-\underline{\nu}^{(2)})-(\underline{Y}-\underline{\nu}^{(1)})'(\underline{Y}-\underline{\nu}^{(1)}) \tag{4.3}$$

where $\underline{\nu}^{(i)} = \underline{P}'\underline{\mu}_X^{(i)}$.

Substituting $\underline{\nu}^{(i)}$ and $\underline{\Lambda}$, we obtain

$$U = \sum_{j=1}^{p+q}[\frac{1}{\lambda_j}(Y_j-\nu_j^{(2)})^2-(Y_j-\nu_j^{(1)})^2], \tag{4.4}$$

where Y_j and $\nu_j^{(i)}$ are the jth components of $\underline{Y}$ and $\underline{\nu}^{(i)}$, i=1,2, respectively. If $\lambda_j \neq 1$ for each j=1,2,...,p+q, we obtain, apart from a constant,

$$V = \sum_{j=1}^{p+q}(\frac{1}{\lambda_j} - 1)(Y_j - \frac{\frac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\frac{1}{\lambda_j} - 1})^2 \quad . \tag{4.5}$$

If $\lambda_{j'} = 1$, for some j=j', the distribution of U in (4.4) is difficult to obtain. However, if the first p components of the eigenvector associated with $\lambda_{j'}$ are all zero, then the terms involving $Y_{j'}$ in (4.4) cancel out, and we can complete the square as in (4.5). A sufficient condition for the first p components of the eigenvector associated with $\lambda_{j'}$ to be zero is

$$(\sigma_2^2-\sigma_1^2) + (\sigma_1^2\rho_1-\sigma_2^2\rho_2) \neq 0, \quad \text{(for p=2,3,...)}$$

and

$(\sigma_2\rho_2' - \sigma_1\rho_1') \neq 0$.

The case when $\underline{\Sigma}_{12}^{(1)} = \underline{\Sigma}_{12}^{(2)}$, hence $\sigma_2\rho_2' - \sigma_1\rho_1' = 0$, will be considered in Chapter V-D-1. Assuming the sufficient condition obtains, we write

$$V = \sum_{j=1}^{s} \left(\frac{1}{\lambda_j} - 1\right)\left(Y_j - \frac{\frac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right)^2, \tag{4.6}$$

where s is the number of eigenvectors λ_j of $|\underline{\Sigma}^{(2)} - \lambda\underline{\Sigma}^{(1)}| = 0$ not equal to one. We shall classify $\underline{X}$ into π_1 if $V \geq c$ and into π_2 if $V < c$ for some suitable choice of the constant c where V is given in (4.6).

To find the distribution of V, we shall assume, without loss of generality, that the coefficients $(\frac{1}{\lambda_j} - 1)$ are labeled so that the first r are positive and the remaining s-r coefficients, $(\frac{1}{\lambda_j} - 1)$, are negative. Then V can be written as

$$V = \sum_{j=1}^{r} \left(\frac{1}{\lambda_j} - 1\right)\left(Y_j - \frac{\frac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right)^2$$

$$- \sum_{j=r+1}^{s} \left(1 - \frac{1}{\lambda_j}\right)\left(Y_j - \frac{\frac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right)^2. \tag{4.7}$$

Let

$$A_j = \sqrt{\frac{1}{\lambda_j} - 1}\,\left(Y_j - \frac{\dfrac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\dfrac{1}{\lambda_j} - 1}\right), \quad j=1,2,\ldots,r, \tag{4.8}$$

and

$$B_j = \sqrt{1 - \frac{1}{\lambda_j}}\,\left(Y_j - \frac{\dfrac{\nu_j^{(2)}}{\lambda_j} - \nu_j^{(1)}}{\dfrac{1}{\lambda_j} - 1}\right), \quad j=r+1,\ldots,s. \tag{4.9}$$

Then

$V = \sum_{j=1}^{r} A_j^2 - \sum_{j=r+1}^{s} B_j^2$. When $\underline{X}$ comes from π_i, i=1 or 2, all A_j's and B_j's are independently distributed as $N(\xi_j^{(i)}, \tau_{ij}^2)$ where

$$\xi_j^{(1)} = \frac{1}{\sqrt{\lambda_j(1-\lambda_j)}}(\nu_j^{(1)} - \nu_j^{(2)}),$$

$$\xi_j^{(2)} = \sqrt{\frac{\lambda_j}{1-\lambda_j}}\,(\nu_j^{(1)} - \nu_j^{(2)}),$$

$$\tau_{1j}^2 = \frac{1-\lambda_j}{\lambda_j},$$

$$\tau_{2j}^2 = 1-\lambda_j, \quad j=1,2,\ldots,r; \tag{4.10}$$

and

$$\xi_j^{(1)} = \frac{1}{\sqrt{\lambda_j(\lambda_j-1)}}\,(\nu_j^{(2)} - \nu_j^{(1)}),$$

$$\xi_j^{(2)} = \sqrt{\frac{\lambda_j}{\lambda_j - 1}} \; (\nu_j^{(2)} - \nu_j^{(1)}),$$

$$\tau_{1j}^2 = \frac{\lambda_j - 1}{\lambda_j},$$

$$\tau_{2j}^2 = (\lambda_j - 1), \quad j = r+1, \ldots, s. \tag{4.11}$$

Now

$$V \sim \sum_{j=1}^{r} \tau_{ij}^2 \chi_1'^2(\delta_{ij}^2) - \sum_{j=r+1}^{s} \tau_{ij}^2 \chi_1'^2(\delta_{ij}^2) \tag{4.12}$$

when $\underline{X}$ comes from π_i, $i=1,2$, where $\chi_1'^2(\delta_{ij}^2)$ denotes a non-central chi-square variate with 1 degree of freedom and non-centrality parameter $\delta_{ij}^2 = \dfrac{[\xi_j^{(i)}]^2}{\tau_{ij}^2}$.

Hence V in (4.12) can be written as

$$V = A - B, \tag{4.13}$$

where A and B are given in (1.15) in Chapter I and are independent positive-definite quadratic forms in non-central chi-square variates. In particular, if $\underline{\mu}^{(1)} = \underline{\mu}^{(2)}$, then all variates are central chi-square variates. In general, it is not easy to obtain the distribution of V given in (4.12) or (4.13) in closed form. Gurland (1955), in considering the problem for central chi-square 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. Shah (1963) extended his work to the non-central case. Press (1966) obtained the c.d.f. of forms such as V.

C. Means Unknown, Covariance Matrices Known

In this section, we shall assume that the means $\underline{\mu}_X^{(i)} = \begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}$ are unknown, while $\underline{\Sigma}^{(i)}$ are known. Suppose we have a random sample $\underline{X}_1^{(1)}, \underline{X}_2^{(1)}, \ldots, \underline{X}_{N_1}^{(1)}$ from π_1: $N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(1)}\right)$ and an independent random sample $\underline{X}_1^{(2)}, \underline{X}_2^{(2)}, \ldots, \underline{X}_{N_2}^{(2)}$ from π_2: $N_{p+q}\left(\begin{pmatrix} \underline{\mu}^{(2)} \\ \underline{\mu} \end{pmatrix}, \underline{\Sigma}^{(2)}\right)$. Our estimate of $\underline{\mu}_X^{(i)}$ is the sample mean $\overline{\underline{X}}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \underline{X}_\alpha^{(i)}$. The sample means $\overline{\underline{X}}^{(i)}$ are distributed independently as $N_{p+q}(\underline{\mu}_X^{(i)}, \frac{1}{N_i} \underline{\Sigma}^{(i)})$, i=1,2. The sample means are also independent of the distribution of $\underline{X}$. If $\underline{Y} = \underline{P}'\underline{X}$ where $\underline{P}$ is a non-singular matrix such that

$$\underline{P}'\underline{\Sigma}^{(1)}\underline{P} = \underline{I},\ \underline{P}'\underline{\Sigma}^{(2)}\underline{P} = \underline{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{p+q}), \quad (4.14)$$

where λ_j's are the roots of $|\underline{\Sigma}^{(2)} - \lambda\underline{\Sigma}^{(1)}| = 0$, then we estimate $\underline{\nu}^{(i)} = \underline{P}'\underline{\mu}_X^{(i)}$ by $\overline{\underline{Y}}^{(i)} = \underline{P}'\overline{\underline{X}}^{(i)}$, i=1,2. The sample means $\overline{\underline{Y}}^{(i)}$ are independently distributed as $N_{p+q}(\underline{\nu}^{(1)}, \frac{1}{N_1}\underline{I})$ and $N_{p+q}(\underline{\nu}^{(2)}, \frac{1}{N_2}\underline{\Lambda})$ for i=1,2, respectively, and independent of the distribution of $\underline{Y}$. Further, since $\underline{I}$ and $\underline{\Lambda}$ are diagonal matrices, the components of $\underline{\bar{Y}}^{(i)}$ are also independently distributed. The discriminant function V in (4.6), after substituting for the unknown parameters by their respective estimates, becomes

$$V = \sum_{j=1}^{r} \left(\frac{1}{\lambda_j} - 1\right)\left(Y_j - \frac{\frac{\bar{Y}_j^{(2)}}{\lambda_j} - \bar{Y}_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right)^2$$

$$- \sum_{j=r+1}^{s} \left(1 - \frac{1}{\lambda_j}\right)\left(Y_j - \frac{\frac{\bar{Y}_j^{(2)}}{\lambda_j} - \bar{Y}_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right)^2, \qquad (4.15)$$

where Y_j and $\bar{Y}_j^{(i)}$ are the jth components of $\underline{Y}$ and $\underline{\bar{Y}}^{(i)}$, i=1,2, respectively. Let

$$A_j^* = \sqrt{\frac{1}{\lambda_j} - 1}\left(Y_j - \frac{\frac{\bar{Y}_j^{(2)}}{\lambda_j} - \bar{Y}_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right), \quad j=1,2,\ldots,r, \qquad (4.16)$$

and

$$B_j^* = \sqrt{1 - \frac{1}{\lambda_j}}\left(Y_j - \frac{\frac{\bar{Y}_j^{(2)}}{\lambda_j} - \bar{Y}_j^{(1)}}{\frac{1}{\lambda_j} - 1}\right), \quad j=r+1,\ldots,s . \qquad (4.17)$$

Then $V = \sum_{j=1}^{r} A_j^{*2} - \sum_{j=r+1}^{s} B_j^{*2}$. The distribution of V in

(4.15) can be found in a manner analogous to that in Section B. When $\underline{X}$ comes from π_i, i=1 or 2, all A_j^*'s and B_j^*'s are independently distributed as $N(\xi_j^{(i)}, \tau_{ij}^{*2})$, j=1,2,...,s, where $\xi_j^{(i)}$ are given in (4.10) and (4.11) and

$$\tau_{1j}^{*2} = (\frac{1}{\lambda_j} - 1) + \frac{N_1+N_2\lambda_j}{N_1N_2(1-\lambda_j)} ,$$

$$\tau_{2j}^{*2} = (\frac{1}{\lambda_j} - 1)\lambda_j + \frac{N_1+N_2\lambda_j}{N_1N_2(1-\lambda_j)} , \quad j=1,2,\ldots,r; \qquad (4.18)$$

$$\tau_{1j}^{*2} = (1- \frac{1}{\lambda_j}) + \frac{N_1+N_2\lambda_j}{N_1N_2(\lambda_j-1)} ,$$

$$\tau_{2j}^{*2} = (1- \frac{1}{\lambda_j})\lambda_j + \frac{N_1+N_2\lambda_j}{N_1N_2(\lambda_j-1)} , \quad j=r+1,\ldots,s. \qquad (4.19)$$

The second term on the right side of τ_{ij}^{*2} is the increase in variance accounted for by the unknown means. Therefore, V in (4.15) is distributed as

$$\sum_{j=1}^{r} \tau_{ij}^{*2}\chi_1'^2(\delta_{ij}^{*2}) - \sum_{j=r+1}^{s} \tau_{ij}^{*2}\chi_1'^2(\delta_{ij}^{*2}) \quad \text{when } \underline{X} \text{ comes}$$

from π_i, i=1,2, where $\delta_{ij}^{*2} = \frac{[\xi_j^{(i)}]^2}{\tau_{ij}^{*2}}$ and where $\xi_j^{(i)}$ and τ_{ij}^{*2} are specified above.

D. Parameters Unknown

In this section, we assume all parameters $\underline{\mu}_X^{(i)}$, $\underline{\Sigma}^{(i)}$ are unknown. Suppose that we have a random sample $\underline{X}_1^{(1)}$, $\underline{X}_2^{(1)}$, ..., $\underline{X}_{N_1}^{(1)}$ from π_1: $N_{p+q}(\underline{\mu}_X^{(1)}, \underline{\Sigma}^{(1)})$ and an independent random sample $\underline{X}_1^{(2)}$, $\underline{X}_2^{(2)}$, ..., $\underline{X}_{N_2}^{(2)}$ from π_2: N_{p+q} $(\underline{\mu}_X^{(2)}, \underline{\Sigma}^{(2)})$. Our estimate of $\underline{\mu}_X^{(i)}$ is the sample mean $\underline{\bar{X}}^{(i)} = \frac{1}{N_i}\sum_{\alpha=1}^{N_i} \underline{X}_\alpha^{(i)}$ and of $\underline{\Sigma}^{(i)}$ is the sample variance

$$\underline{S}^{(i)} = \frac{1}{n_i}\sum_{\alpha=1}^{N_i} (\underline{X}_\alpha^{(i)}-\underline{\bar{X}}^{(i)})(\underline{X}_\alpha^{(i)}-\underline{\bar{X}}^{(i)})', \tag{4.20}$$

where $n_i = N_i-1$, $i=1,2$. When all parameters are unknown, let U* denote the discriminant function obtained if we replace any unknown parameters in the discriminant function U in (4.2) by their respective estimates.

Hence,

$$U^* = (\underline{X}-\underline{\bar{X}}^{(2)})'[\underline{S}^{(2)}]^{-1}(\underline{X}-\underline{\bar{X}}^{(2)})-(\underline{X}-\underline{\bar{X}}^{(1)})'[\underline{S}^{(1)}]^{-1}(\underline{X}-\underline{\bar{X}}^{(1)}). \tag{4.21}$$

As in Section B, we know there exists a non-singular matrix $\underline{\hat{P}}$ and a diagonal matrix $\underline{\hat{\Lambda}}$ such that

$$\underline{\hat{P}}'\underline{S}^{(1)}\underline{\hat{P}} = \underline{I},\ \underline{\hat{P}}'\underline{S}^{(2)}\underline{\hat{P}} = \hat{\Lambda} = \text{diag}(\hat{\lambda}_1,\hat{\lambda}_2,\ldots,\hat{\lambda}_{p+q}), \tag{4.22}$$

where the $\hat{\lambda}_j$'s are the roots of $|\underline{S}^{(2)}-\hat{\lambda}\underline{S}^{(1)}| = 0$. We adopt $(\underline{\hat{P}}, \underline{\hat{\Lambda}})$ as an estimate of $(\underline{P}, \underline{\Lambda})$. Let $\underline{Y}^* = \underline{\hat{P}}'\underline{X}$ and

$\underline{\bar{Y}}^{*(i)} = \underline{\hat{P}}'\underline{\bar{X}}^{(i)}$, i=1,2. Therefore, the discriminant function U* in (4.21) can be written as

$$U^* = (\underline{Y}^*-\underline{\bar{Y}}^{*(2)})'\underline{\hat{\Lambda}}^{-1}(\underline{Y}^*-\underline{\bar{Y}}^{*(2)})-(\underline{Y}^*-\underline{\bar{Y}}^{*(1)})'(\underline{Y}^*-\underline{\bar{Y}}^{*(1)}) . \quad (4.23)$$

Okamoto (1961) has shown that as N_1, $N_2 \to \infty$ such that $\frac{N_2}{N_1} \to$ constant,

$$\text{plim } \underline{\hat{\Lambda}} = \underline{\Lambda},$$
$$\text{plim } \underline{\hat{P}} = \underline{P} . \quad (4.24)$$

The distribution of $(\underline{Y}^*-\underline{\bar{Y}}^{*(i)})$ tends to the distribution of $(\underline{Y}-\underline{\nu}^{(i)})$ as $N_1, N_2 \to \infty$ for i=1,2, respectively.

Therefore, from Cramér's Theorem (Cramér, 1946, p. 254) for stochastic convergence we find that the distribution of U* tends to the distribution of U. It follows that the discriminant function U* in (4.23) is asymptotically equivalent to the discriminant function U in (4.3).

V. COMPARISONS OF PROBABILITIES OF MISCLASSIFICATION

A. Introduction

In the previous three chapters, we have discussed the distributions of the discriminant functions when there are

(a) p unadjusted discriminators,

(b) p adjusted (for q covariates) discriminators, and

(c) p+q unadjusted discriminators.

In this chapter, we shall compare probabilities of misclassification for these three classification procedures. Numerical comparisons are made for the following cases: a bivariate (p=1, q=1) case and a trivariate (p=1, q=2) case. In both cases all parameters $\begin{pmatrix} \underline{\mu}^{(i)} \\ \underline{\mu} \end{pmatrix}$, $\underline{\Sigma}^{(i)}$, i=1,2, are assumed known. Various sets of parameters will be considered for each case. The probabilities of misclassification are approximated by numerical integration for some parameter situations. But it is not possible to obtain approximations for all cases. Hence a Monte Carlo study was conducted to compare the three classification procedures empirically.

B. Methods of Computing Probabilities of Misclassification for the Bivariate (p=1, q=1) Case

Consider the problem of classification of a bivariate observation $\underline{X} = (X_1, X_2)'$ into one of two normal populations π_1 and π_2 where π_i: $N_2\left(\begin{pmatrix} \mu^{(i)} \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \sigma_i \sigma \rho_i' \\ \sigma_i \sigma \rho_i' & \sigma^2 \end{pmatrix}\right)$, $= 1,2$. We shall assume all parameters are known. Since the parameters are known, we could make the transformation

$$X_1 \to \frac{1}{\sigma_2}(X_1 - \mu^{(1)})$$

$$X_2 \to \frac{1}{\sigma}(X_2 - \mu)$$

and consequently the populations π_1 and π_2 could be written as

$$\pi_1\colon N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_0 \rho_1' \\ \sigma_0 \rho_1' & 1 \end{pmatrix}\right) \text{ and } \pi_2\colon N_2\left(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2' \\ \rho_2' & 1 \end{pmatrix}\right),$$

where $\theta = \dfrac{\mu^{(2)} - \mu^{(1)}}{\sigma_2}$ and $\sigma_0^2 = \dfrac{\sigma_1^2}{\sigma_2^2}$. We shall assume $\sigma_0^2 > 1$.

In this section, we shall determine the probabilities of misclassification for each of the following three procedures:

(a) the univariate (p=1) unadjusted discriminator,

(b) the univariate adjusted (for q=1 covariate) discriminator, and

(c) the bivariate $(p+q=2)$ unadjusted discriminator.

Numerical integration procedures will be obtained for use in approximating the probabilities of misclassification for each of the three procedures under some parameter situations. In addition, Monte Carlo procedures are used to estimate the probabilities of misclassification empirically for each of the three classification procedures.

1. Numerical integration procedures

a. The univariate $(p=1)$ unadjusted discriminator In this situation, we consider only the variate X_1 as a discriminator ignoring the covariate X_2. Hence if $\underline{X}\varepsilon\pi_1$, then $X_1 \sim N(0, \sigma_0^2)$ and if $\underline{X}\varepsilon\pi_2$, then $X_1 \sim N(\theta,1)$.

Following Chapter II, we shall classify the observation into π_1 if

$$(X_1-\theta)^2 - \frac{1}{\sigma_0^2} X_1^2 + \log \frac{1}{\sigma_0^2} \geq 0, \tag{5.1}$$

and into π_2, otherwise. Equivalently, classify the observation into π_1 if

$$\left(\frac{X_1}{\sigma_0} - \frac{\theta\sigma_0}{\sigma_0^2-1}\right)^2 \geq \frac{1}{\sigma_0^2-1} \left(\log \sigma_0^2 + \frac{\theta^2}{\sigma_0^2-1}\right) = \frac{k}{\sigma_0^2-1} , \tag{5.2}$$

where

$$k = \log \sigma_0^2 + \frac{\theta^2}{\sigma_0^2-1} .$$

If $\underline{X}\varepsilon\pi_1$, then

$$(\frac{X_1}{\sigma_0} - \frac{\theta\sigma_0}{\sigma_0^2-1}) \sim N(\frac{-\theta\sigma_0}{\sigma_0^2-1}\ ,\ 1)$$

so

$$(\frac{X_1}{\sigma_0} - \frac{\theta\sigma_0}{\sigma_0^2-1})^2 \sim \chi_1'^2(\frac{\theta^2\sigma_0^2}{(\sigma_0^2-1)^2})\ .$$

Therefore,

$$P(2/1) = P[\chi_1'^2(\frac{\theta^2\sigma_0^2}{(\sigma_0^2-1)^2}) < \frac{k}{\sigma_0^2-1}]\ . \qquad (5.3)$$

If $\underline{X}\epsilon\pi_2$, then

$$(X_1 - \frac{\theta\sigma_0^2}{\sigma_0^2-1}) \sim N(\frac{-\theta}{\sigma_0^2-1}\ ,\ 1)$$

so

$$(X_1 - \frac{\theta\sigma_0^2}{\sigma_0^2-1})^2 \sim \chi_1'^2(\frac{\theta^2}{(\sigma_0^2-1)^2})\ .$$

Therefore,

$$P(1/2) = P[\chi_1'^2(\frac{\theta^2}{(\sigma_0^2-1)^2}) \geq \frac{k\sigma_0^2}{(\sigma_0^2-1)}]. \qquad (5.4)$$

To obtain P(2/1) and P(1/2) when $\theta \neq 0$, one could use non-central chi-square tables where convenient. We shall use an approximation to the non-central chi-square distribution for determining the probabilities of misclassification. Many approximations have been suggested. The simplest approximation consists of replacing χ'^2 by a multiple of a central χ^2, $a\chi_g^2$ say, with a and g chosen so that the first two moments of the two variables, $\chi_\eta'^2(\delta)$ and $a\chi_g^2$ agree. The appropriate values of a and g are

$$a = \frac{\eta+2\delta}{\eta+\delta} \quad , \quad g = \frac{(\eta+\delta)^2}{\eta+2\delta} .$$

This approximation was suggested by Patnaik (1949). Pearson (1959) suggested an improvement of this approximation, introducing an additional constant b, and choosing b, c, and f so that the first three cumulants of $\chi_\nu'^2(\lambda)$ and $(c\chi_f^2 + b)$ agree. Let K_i and K_i^* denote the ith cumulants of $(c\chi_f^2 + b)$ and $\chi_\nu'^2(\lambda)$, respectively for i=1,2,3. Equating $K_i = K_i^*$, i=1,2,3, we have

$$cf + b = \nu+\lambda$$
$$2c^2f = 2(\nu+2\lambda)$$
$$8c^3f = 8(\nu+3\lambda).$$

Solving for b, c, and f gives

$$b = \frac{-\lambda^2}{\nu+3\lambda} ,$$

$$c = \frac{\nu+3\lambda}{\nu+2\lambda} ,$$

and

$$f = \frac{(\nu+2\lambda)^3}{(\nu+3\lambda)^2} . \tag{5.5}$$

In both Patnaik's and Pearson's approximations the degrees of freedom g and f are usually fractional, thus interpolation is needed if standard χ^2 tables are used.

In the univariate unadjusted discriminator situation considered, we will use Pearson's approximation to determine the probabilities of misclassification P(2/1) and P(1/2) when $\theta \neq 0$. For determining P(2/1):

$$\nu = 1, \quad \lambda = \frac{\theta^2 \sigma_0^2}{(\sigma_0^2 - 1)^2}$$

and for determining P(1/2): $\nu = 1$, $\lambda = \dfrac{\theta^2}{(\sigma_0^2 - 1)^2}$.

To compute P(2/1) and P(1/2), the IBM Scientific Subroutine Program CDTR was used. This subroutine determines $P[Y \leq y]$ where $Y \sim \chi_f^2$. Computation results are shown in the tables in the Appendix for given values of parameters.

b. <u>The univariate adjusted (for q=1 covariate) discriminator</u> As before we have the bivariate normal populations

$$\pi_1: N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_0 \rho_1' \\ \sigma_0 \rho_1' & 1 \end{pmatrix}\right) \text{ and } \pi_2: N_2\left(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2' \\ \rho_2' & 1 \end{pmatrix}\right) .$$

As specified in Chapter III, we form the variates

$$Z_1 = X_1 - \underline{\Sigma}_{12}^{(1)} \underline{\Sigma}_{22}^{-1} X_2 = X_1 - \sigma_0 \rho_1' X_2 ,$$

$$Z_2 = X_1 - \underline{\Sigma}_{12}^{(2)} \underline{\Sigma}_{22}^{-1} X_2 = X_1 - \rho_2' X_2 . \tag{5.6}$$

Then use the weighted variate Z where

$$Z = W_1 Z_1 + W_2 Z_2 . \tag{5.7}$$

Possible choices for weights will be discussed in Section D. Two of these choices are the constant weights

$$W_1 = \frac{|\underline{\Sigma}^{(2)}_{11.2}|}{|\underline{\Sigma}^{(1)}_{11.2}| + |\underline{\Sigma}^{(2)}_{11.2}|}, \quad W_2 = \frac{|\underline{\Sigma}^{(1)}_{11.2}|}{|\underline{\Sigma}^{(1)}_{11.2}| + |\underline{\Sigma}^{(2)}_{11.2}|},$$

and the matrix weights

$$\underline{W}_1 = [(\underline{\Sigma}^{(1)}_{11.2})^{-1} + (\underline{\Sigma}^{(2)}_{11.2})^{-1}]^{-1}[\underline{\Sigma}^{(1)}_{11.2}]^{-1},$$

$$\underline{W}_2 = [(\underline{\Sigma}^{(1)}_{11.2})^{-1} + (\underline{\Sigma}^{(2)}_{11.2})^{-1}]^{-1}[\underline{\Sigma}^{(2)}_{11.2}]^{-1}.$$

For the case p=1, these weights are the same and are given by

$$W_1 = \frac{\sigma^{(2)}_{11.2}}{\sigma^{(1)}_{11.2} + \sigma^{(2)}_{11.2}} = \frac{1-\rho_2'^2}{\sigma_0^2(1-\rho_1'^2) + (1-\rho_2'^2)}, \quad W_2 = 1 - W_1. \tag{5.8}$$

Simplifying (5.7) we obtain

$$Z = X_1 - (W_1\sigma_0\rho_1' + W_2\rho_2')X_2, \tag{5.9}$$

where W_i are given in (5.8). We consider Z as our univariate adjusted discriminator.

If $\underline{X}\varepsilon\pi_1$, then $Z \sim N(0, V_1)$ where

$$V_1 = \sigma_0^2 + (W_1\sigma_0\rho_1' + W_2\rho_2')^2 - 2(W_1\sigma_0\rho_1' + W_2\rho_2')\sigma_0\rho_1'. \tag{5.10}$$

If $\underline{X}\varepsilon\pi_2$, then $Z \sim N(\theta, V_2)$ where

$$V_2 = 1 + (W_1\sigma_0\rho_1'+W_2\rho_2')^2 - 2(W_1\sigma_0\rho_1'+W_2\rho_2')\rho_2' . \tag{5.11}$$

By the results in Chapter III, we shall classify $\underline{X}$ into π_1 if

$$\frac{(Z-\theta)^2}{V_2} - \frac{Z^2}{V_1} + \log \frac{V_2}{V_1} \geq 0. \tag{5.12}$$

Suppose $V_1 > V_2$. Then we classify $\underline{X}$ into π_2 if

$$\frac{1}{V_1}(Z - \frac{V_1\theta}{V_1-V_2})^2 < \frac{V_2}{V_1-V_2} k_1 , \tag{5.13}$$

where

$$k_1 = \log \frac{V_1}{V_2} - \frac{\theta^2}{V_2} + \frac{V_1\theta^2}{V_2(V_1-V_2)} .$$

If $\underline{X}\varepsilon\pi_1$, then

$$\frac{1}{\sqrt{V_1}} (Z - \frac{V_1\theta}{V_1-V_2}) \sim N(\frac{-\sqrt{V_1}\theta}{V_1-V_2} ,1)$$

so

$$\frac{1}{V_1} (Z - \frac{V_1\theta}{V_1-V_2})^2 \sim \chi_1'^2(\frac{V_1\theta^2}{(V_1-V_2)^2}) .$$

Therefore,

$$P(2/1) = P[\chi_1'^2(\frac{V_1\theta^2}{(V_1-V_2)^2}) < \frac{V_2}{V_1-V_2} k_1] . \tag{5.14}$$

From (5.13), if $V_1>V_2$, classify $\underline{X}$ into π_1 if

$$\frac{1}{V_2}(Z - \frac{V_1\theta}{V_1-V_2})^2 \geq \frac{V_1}{V_1-V_2} k_1 . \tag{5.15}$$

If $\underline{X}\varepsilon\pi_2$, then

$$\frac{1}{\sqrt{V_2}}\,(Z - \frac{V_1\theta}{V_1-V_2}) \sim N(\frac{-\sqrt{V_2}\,\theta}{V_1-V_2}\,,\,1)$$

so

$$\frac{1}{V_2}(Z - \frac{V_1\theta}{V_1-V_2})^2 \sim \chi_1'^2(\frac{V_2\theta^2}{(V_1-V_2)^2}) \ .$$

Therefore,

$$P(1/2) = P[\chi_1'^2(\frac{V_2\theta^2}{(V_1-V_2)^2}) \geq \frac{V_1}{V_1-V_2}\,k_1]. \qquad (5.16)$$

If $V_1<V_2$, then we would have to reverse the inequalities in (5.14) and (5.16).

To determine P(2/1) and P(1/2), Pearson's approximation was again used. Computation results are given in the tables in the Appendix.

c. <u>The bivariate (p+q=2) unadjusted discriminator</u> In this procedure, we consider both X_1 and X_2 as discriminators ignoring the fact that X_2 has the same mean and covariance matrix in both populations π_1 and π_2.

Using the likelihood ratio procedure our rule is to classify $\underline{X}$ into π_1 if

$$U = (\underline{X}-\underline{\mu}^{(2)})'[\underline{\Sigma}^{(2)}]^{-1}(\underline{X}-\underline{\mu}^{(2)}) - \underline{X}'[\underline{\Sigma}^{(1)}]^{-1}\underline{X} + \log\frac{|\underline{\Sigma}^{(2)}|}{|\underline{\Sigma}^{(1)}|} \geq 0 \qquad (5.17)$$

and into π_2 if $U < 0$.

As mentioned in Chapter IV, there exists a non-singular matrix $\underline{P}$ such that

$$\underline{P}'\underline{\Sigma}^{(1)}\underline{P} = \underline{I},\ \underline{P}'\underline{\Sigma}^{(2)}\underline{P} = \underline{\Lambda} = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_{p+q})\ , \qquad (5.18)$$

where the λ_j's are roots of $|\underline{\Sigma}^{(2)}-\lambda\underline{\Sigma}^{(1)}| = 0$. Let $\underline{Y} = \underline{P}'\underline{X}$. Then (5.18) becomes classify $\underline{X}$ into π_1 if

$$(\underline{Y}-\underline{\nu}^{(2)})'\underline{\Lambda}^{-1}(\underline{Y}-\underline{\nu}^{(2)}) - \underline{Y}'\underline{Y} + \log|\underline{\Lambda}| \geq 0, \qquad (5.19)$$

where $\underline{\nu}^{(2)} = \underline{P}'\underline{\mu}^{(2)}$.

To determine the matrix $\underline{P}$ in general, the following procedure given in condensed form by Cooley and Lohnes (1962) was used. We used the IBM Scientific Subroutine Program EIGEN to determine matrices $\underline{A}$ and $\underline{D}$ such that

$$\underline{\Sigma}^{(1)} = \underline{A}\ \underline{D}\ \underline{A}' \qquad (5.20)$$

where $\underline{A}$ is the orthogonal matrix whose columns are the eigenvectors of $\underline{\Sigma}^{(1)}$ and $\underline{D}$ is the diagonal matrix whose elements are the eigenvalues of $\underline{\Sigma}^{(1)}$. The method, which used the IBM Scientific Subroutine Program NROOT, involves finding the eigenvalues and eigenvectors of the non-symmetric matrix, $[\underline{\Sigma}^{(1)}]^{-1}\underline{\Sigma}^{(2)}$. If

$$(\underline{\Sigma}^{(2)}-\lambda\underline{\Sigma}^{(1)})\underline{p} = \underline{0}$$

the λ's are the eigenvalues of $[\underline{\Sigma}^{(1)}]^{-1}\underline{\Sigma}^{(2)}$ because the roots of

$$|\underline{\Sigma}^{(2)}-\lambda\underline{\Sigma}^{(1)}|= 0$$

are equal to the roots of

$$|[\underline{\Sigma}^{(1)}]^{-1}\underline{\Sigma}^{(2)}-\lambda\underline{I}| = 0.$$

In view of (5.20),

$$\underline{\Sigma}^{(1)} = \underline{A}\ \underline{D}^{\frac{1}{2}}\ \underline{D}^{\frac{1}{2}}\underline{A}' ,$$

$$(\underline{\Sigma}^{(2)}-\lambda\underline{A}\ \underline{D}^{\frac{1}{2}}\underline{D}^{\frac{1}{2}}A')\underline{p} = \underline{0},$$

$$(\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}-\lambda\underline{D}^{\frac{1}{2}}\underline{A}')\underline{p} = \underline{0} ,$$

$$(\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}}\underline{D}^{\frac{1}{2}}A'-\lambda\underline{D}^{\frac{1}{2}}\underline{A}')\underline{p} = \underline{0},$$

and

$$(\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}}-\lambda\underline{I})\underline{D}^{\frac{1}{2}}\underline{A}'\underline{p} = \underline{0}\ . \tag{5.21}$$

Let the eigenvectors $\underline{p}$ be chosen so that $\underline{b} = \underline{D}^{\frac{1}{2}}\underline{A}'\underline{p}$ are of unit length. Then (5.21) becomes

$$(\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}}-\lambda\underline{I})\underline{b} = 0 \tag{5.22}$$

and, therefore,

$$\underline{B}'(\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}})\underline{B} = \underline{\Lambda}, \tag{5.23}$$

where $\underline{B}$ is the orthogonal matrix whose columns are the eigenvectors of $\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}}$ and $\underline{\Lambda}$ is the diagonal matrix whose diagonal elements are the eigenvalues of

$$\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\ \underline{D}^{-\frac{1}{2}} .$$

From $\underline{B} = \underline{D}^{\frac{1}{2}}\underline{A}'\underline{P}$, we have $\underline{P} = \underline{A}\,\underline{D}^{-\frac{1}{2}}\underline{B}$. Thus,

$$\begin{aligned} \underline{P}'\underline{\Sigma}^{(1)}\underline{P} &= \underline{B}'\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(1)}\underline{A}\,\underline{D}^{-\frac{1}{2}}\underline{B} \\ &= \underline{B}'\underline{D}^{-\frac{1}{2}}\underline{D}\,\underline{D}^{-\frac{1}{2}}\underline{B} \\ &= \underline{I}\ . \end{aligned} \tag{5.24}$$

Also,

$$\begin{aligned} \underline{P}'\underline{\Sigma}^{(2)}\underline{P} &= \underline{B}'\underline{D}^{-\frac{1}{2}}\underline{A}'\underline{\Sigma}^{(2)}\underline{A}\,\underline{D}^{-\frac{1}{2}}\underline{B} \\ &= \underline{\Lambda}\ . \end{aligned} \tag{5.25}$$

Hence the matrix $\underline{P}$ obtained, is the desired matrix. Thus far, our development could have been for a multivariate unadjusted discriminator situation.

In the bivariate case, for the parameter values specified, the matrices $\underline{P}$ and $\underline{\Lambda}$ were determined by the method described above. One of the roots, say λ_1, was less than one and the other root, say λ_2, was greater than one. The roots depend on σ_0^2 and (ρ_1', ρ_2'). The roots obtained from $(\rho_1', -\rho_2')$ and from $(-\rho_1', \rho_2')$ are the same. Also the roots from (ρ_1', ρ_2') and from $(-\rho_1', -\rho_2')$ are the same. Therefore, we only considered $(\rho_1', -\rho_2')$ and (ρ_1', ρ_2') for the specified parameters.

Specifically for the bivariate (p+q=2) unadjusted discriminator situation where $\underline{\mu}^{(2)} = \begin{pmatrix}\theta \\ 0\end{pmatrix}$ and hence $\underline{\nu}^{(2)} = \underline{P}'\underline{\mu}^{(2)} = \begin{pmatrix}\nu_1^{(2)} \\ \nu_2^{(2)}\end{pmatrix}$, we have from (5.19) to classify $\underline{X}$ into π_1 if

$$\frac{1}{\lambda_1}(Y_1-\nu_1^{(2)})^2 + \frac{1}{\lambda_2}(Y_2-\nu_2^{(2)})^2-Y_1^2-Y_2^2+\log(\lambda_1\lambda_2) \geq 0, \quad (5.26)$$

where $\underline{Y} = (Y_1,Y_2)'$. Letting $\lambda_1<1$, $\lambda_2>1$, we can write (5.26) as

$$(\frac{1}{\lambda_1} - 1)(Y_1- \frac{\nu_1^{(2)}}{1-\lambda_1})^2-(1- \frac{1}{\lambda_2})(Y_2- \frac{\nu_2^{(2)}}{1-\lambda_2})^2- \frac{(\nu_1^{(2)})^2}{1-\lambda_1}$$

$$- \frac{(\nu_2^{(2)})^2}{1-\lambda_2} + \log(\lambda_1\lambda_2) \geq 0 . \quad (5.27)$$

If $\underline{X}\varepsilon\pi_1$, then

$$(Y_1 - \frac{\nu_1^{(2)}}{1-\lambda_1})^2 \sim \chi_1'^2(\frac{(\nu_1^{(2)})^2}{(1-\lambda_1)^2})$$

and

$$(Y_2 - \frac{\nu_2^{(2)}}{1-\lambda_2})^2 \sim \chi_1'^2(\frac{(\nu_2^{(2)})^2}{(1-\lambda_2)^2}) .$$

If $\underline{X}\varepsilon\pi_2$, then

$$\frac{1}{\lambda_1}(Y_1 - \frac{\nu_1^{(2)}}{1-\lambda_1})^2 \sim \chi_1'^2(\frac{\lambda_1(\nu_1^{(2)})^2}{(1-\lambda_1)^2})$$

and

$$\frac{1}{\lambda_2}(Y_2 - \frac{\nu_2^{(2)}}{1-\lambda_2})^2 \sim \chi_1'^2(\frac{\lambda_2(\nu_2^{(2)})^2}{(1-\lambda_2)^2}) .$$

Therefore, the probabilities of misclassification are given by

$$P(2/1) = P[(\frac{1}{\lambda_1} - 1)\chi_1'^2(\frac{(\nu_1^{(2)})^2}{(1-\lambda_1)^2})-(1- \frac{1}{\lambda_2})\chi_1'^2(\frac{(\nu_2^{(2)})^2}{(1-\lambda_2)^2})< k_2] \quad (5.28)$$

and

$$P(1/2) = P[(1-\lambda_1)\chi_1'^2(\frac{\lambda_1(\nu_1^{(2)})^2}{(1-\lambda_1)^2})-(\lambda_2-1)\chi_1'^2(\frac{\lambda_2(\nu_2^{(2)})^2}{(1-\lambda_2)^2}) \geq k_2], \tag{5.29}$$

where

$$k_2 = \frac{(\nu_1^{(2)})^2}{(1-\lambda_1)} + \frac{(\nu_2^{(2)})^2}{(1-\lambda_2)} - \log(\lambda_1\lambda_2).$$

If $\theta=0$, then $\underline{\nu}^{(2)}=\underline{0}$. Thus the probabilities of misclassification P(2/1) and P(1/2) become

$$P(2/1) = P[(\frac{1}{\lambda_1} - 1)\chi_1^2-(1-\frac{1}{\lambda_2})\chi_1^2 < -\log(\lambda_1\lambda_2)] \tag{5.30}$$

and

$$P(1/2) = P[(1-\lambda_1)\chi_1^2-(\lambda_2-1)\chi_1^2 \geq -\log(\lambda_1\lambda_2)]. \tag{5.31}$$

To determine the probabilities of misclassification in (5.30) and (5.31) we have to determine probabilities of the form

$P[\alpha X-\beta Y>c]$

where c is a constant, $\alpha>0$, $\beta>0$, $X\sim\chi_1^2$, $Y\sim\chi_1^2$, and X and Y are independent.

More generally, we may wish to determine probabilities of the form

$P[\alpha X-\beta Y>c]$

where $\alpha>0$, $\beta>0$, $X\sim\chi_m^2$, $Y\sim\chi_n^2$, and X and Y are independent. Press (1966) obtained the probability density of $Z = \alpha X-\beta Y$ where $\alpha>0$, $\beta>0$, $X\sim\chi_m^2$, $Y\sim\chi_n^2$, and X and Y are independent as

$$p_Z(t) = \begin{cases} \dfrac{1}{2^{\frac{m+n}{2}} \alpha^{\frac{m}{2}} \beta^{\frac{n}{2}} \Gamma(\frac{m}{2})} \, t^{\frac{m+n-2}{2}} \, e^{-\frac{t}{2\alpha}} \, \psi(\frac{n}{2}, \frac{m+n}{2}; \frac{\alpha+\beta}{2\alpha\beta} t), & t \geq 0 \\ \\ \dfrac{1}{2^{\frac{m+n}{2}} \alpha^{\frac{m}{2}} \beta^{\frac{n}{2}} \Gamma(\frac{n}{2})} \, (-t)^{\frac{m+n-2}{2}} \, e^{\frac{t}{2\beta}} \, \psi(\frac{m}{2}, \frac{m+n}{2}; -\frac{\alpha+\beta}{2\alpha\beta} t), & t \leq 0 \end{cases} \tag{5.32}$$

where ψ is the finite version of the confluent hypergeometric function, which can be found, for example, in (Slater, 1960) and is given by

$$\psi(a,b;x) \equiv \frac{1}{\Gamma(a)} \int_0^\infty e^{-xt} t^{a-1} (1+t)^{b-a-1} dt \tag{5.33}$$

for $a>0$, $x>0$.

In our situation, $m=n=1$ so

$$p_Z(t) = \begin{cases} \dfrac{1}{2\alpha^{\frac{1}{2}} \beta^{\frac{1}{2}} \Gamma(\frac{1}{2})} \, e^{-\frac{t}{2\alpha}} \, \psi(\frac{1}{2}, 1; \frac{\alpha+\beta}{2\alpha\beta} t), & t \geq 0 \\ \\ \dfrac{1}{2\alpha^{\frac{1}{2}} \beta^{\frac{1}{2}} \Gamma(\frac{1}{2})} \, e^{\frac{t}{2\beta}} \, \psi(\frac{1}{2}, 1; -\frac{\alpha+\beta}{2\alpha\beta} t), & t \leq 0 \, . \end{cases} \tag{5.34}$$

Now one can relate (see, for example, Abramowitz and Stegun, 1965, p. 510) $\psi(\gamma+\frac{1}{2}, 2\gamma+1, 2z)$ to a K (modified) Bessel function of order γ by the relation

$$\psi(\gamma+\tfrac{1}{2},\ 2\gamma+1,\ 2Z) = \pi^{-\frac{1}{2}} e^{Z}(2Z)^{-\gamma} K_{\gamma}(Z), \tag{5.35}$$

where $K_{\gamma}(Z)$ denotes a K Bessel function with argument Z and order γ.

For our situation using (5.35) with $\gamma=0$, we can write (5.34) as

$$p_Z(t) = \begin{cases} \dfrac{e^{\frac{(\alpha-\beta)}{4\alpha\beta}t}}{2\alpha^{\frac{1}{2}}\beta^{\frac{1}{2}}\pi} K_0\left(\frac{\alpha+\beta}{4\alpha\beta}\,t\right), & t \geq 0 \\[2ex] \dfrac{e^{\frac{(\alpha-\beta)}{4\alpha\beta}t}}{2\alpha^{\frac{1}{2}}\beta^{\frac{1}{2}}\pi} K_0\left(-\,\frac{\alpha+\beta}{4\alpha\beta}\,t\right), & t \leq 0\ . \end{cases} \tag{5.36}$$

To illustrate, without loss of generality, how the probabilities of misclassification in (5.30) and (5.31) were determined, we will find

$$P[\alpha X-\beta Y \geq C] \tag{5.37}$$

where $\alpha>0$, $\beta>0$, $X\sim\chi_1^2$, $Y\sim\chi_1^2$, X and Y are independent, and C is a positive constant. From (5.36) and (5.37) we have

$$P[\alpha X-\beta Y \geq C] = \int_C^\infty p_Z(t)\,dt\ , \tag{5.38}$$

where

$$p_Z(t) = \frac{e^{\frac{(\alpha-\beta)}{4\alpha\beta}t}}{2\alpha^{\frac{1}{2}}\beta^{\frac{1}{2}}\pi} K_0\left(\frac{\alpha+\beta}{4\alpha\beta}t\right).$$

We can write (5.38) as

$$\int_C^\infty p_Z(t)dt = \int_C^B p_Z(t)dt + \int_B^\infty p_Z(t)dt . \qquad (5.39)$$

We evaluated the first term of the right-hand side of (5.39) numerically by a Newton-Cotes quadrature formula employing Romberg integration with the trapezoidal rule. For a discussion of these procedures, the reader is referred to (Ralston, 1965, pp. 114-124). The interval from C to B was first split into shorter intervals of equal length. For each of these intervals of integration we employed our numerical method of integration, each time halving the given interval of integration until two consecutive results agree within the input tolerance ε specified, in our case $\varepsilon = .0001$, or until the maximum number of specified halvings was accomplished. The K Bessel function of zero order, $K_0(\frac{\alpha+\beta}{4\alpha\beta}t)$, was evaluating using the IBM Scientific Subroutine Program BESK. To determine the second term of the right-hand side of (5.39) we used the fact that for large x,

$$\psi(a,b;x) \cong x^{-a} \quad \text{(Slater, 1960)} .$$

Hence,

$$\int_B^\infty p_Z(t)dt \cong \int_B^\infty \frac{1}{2\alpha^{\frac{1}{2}}\beta^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{t}{2\alpha}} \left(\frac{\alpha+\beta}{2\alpha\beta}t\right)^{-\frac{1}{2}} dt$$

$$= \left(\frac{\alpha}{\alpha+\beta}\right)^{\frac{1}{2}} [1-F_1(\frac{B}{\alpha})] , \qquad (5.40)$$

where F_1 is the c.d.f. of a central chi-square variate with 1 degree of freedom. The IBM Scientific Subroutine Program CDTR was used to evaluate (5.40).

If $\theta \neq 0$, then the probabilities of misclassification P(2/1) and P(1/2) are given by (5.28) and (5.29), respectively. It appears that none of the methods available, for example, Press (1966), Shah (1963), are easily applicable in determining these probabilities numerically. Since later the probabilities given in (5.28) and (5.29) were determined empirically by Monte Carlo procedures, we used a "rough" approximation suggested by Imhof (1961) to check these probabilities of misclassification P(2/1) and P(1/2) for some parameter values. Following Pearson (1959), Imhof (1961) equated the first three cumulants K_r^*, r=1,2,3, of a positive quadratic form Q where

$$Q = \sum_{i=1}^{m} \lambda_i \chi_1'^2(\delta_i^2), \tag{5.41}$$

and

$$K_r^* = \sum_{i=1}^{m} \lambda_i^r 2^{r-1}(r-1)!(1+r\delta_i^2), \tag{5.42}$$

with the first three cumulants K_r, r=1,2,3, of $V = c\chi_f^2 + b$ where

$$\begin{aligned} K_1 &= cf + b, \\ K_2 &= 2c^2 f, \\ K_3 &= 8c^3 f \ . \end{aligned} \tag{5.43}$$

Then

$$P[Q \geq x] \cong P[\chi_f^2 \geq y] , \qquad (5.44)$$

where

$$f = \frac{c_2^3}{c_3^2} , \; y = (x-c_1)\sqrt{\frac{f}{c_2}} + f, \quad c_r = \sum_{i=1}^{m} \lambda_i^r (1+r\delta_i^2),$$

$$r = 1,2,3.$$

In determining the probabilities in (5.28) and (5.29), our quadratic forms are non-positive. Imhof (1961) suggested the same approximation as used in (5.44) if Q is non-positive assuming $c_3 > 0$. If $c_3 < 0$, then approximate the distribution of -Q. Certainly, there could be an appreciable loss of accuracy for non-positive forms, but the approximation still gives useful indications which could be of value for practical considerations. As mentioned previously, for certain parameter values, Imhof's procedure was used to approximate P(2/1) and P(1/2) given by (5.28) and (5.29), respectively.

2. Monte Carlo procedures

As in the previous section, assume π_1 and π_2 are bivariate normal populations given by

$$\pi_1 : N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_0\rho_1' \\ \sigma_0\rho_1' & 1 \end{pmatrix} \right) \text{ and } \pi_2 : N_2 \left(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2' \\ \rho_2' & 1 \end{pmatrix} \right).$$

In this section, we will illustrate how to estimate the probabilities of misclassification by Monte Carlo procedures.

We will generate independent uniform pseudo-random variates U on (0,1) by a composite multiplicative congruential generator suggested by Marsaglia and Bray (1968). A pair of independent standard normal pseudo-random variates Y_1, Y_2 can be obtained from each pair of independent uniform variates U_1, U_2 by using the transformation of Box and Muller (1958), that is,

$$Y_1 = (-2 \log U_1)^{\frac{1}{2}} \cos(2\pi U_2)$$

and

$$Y_2 = (-2 \log U_1)^{\frac{1}{2}} \sin(2\pi U_2).$$

To obtain $\underline{X} = (X_1, X_2)' \varepsilon \pi_1$, we need to determine a matrix $\underline{G}$ such that $\underline{X} = \underline{G}\ \underline{Y}$ where $\underline{G}\ \underline{G}' = \begin{pmatrix} \sigma_0^2 & \sigma_0 \rho_1' \\ \sigma_0 \rho_1' & 1 \end{pmatrix}$ and to obtain $\underline{X} = (X_1, X_2)' \varepsilon \pi_2$, we need to determine a matrix $\underline{H}$ such that $\underline{X} = \underline{H}\ \underline{Y} + (\theta, 0)'$ where $\underline{H}\ \underline{H}' = \begin{pmatrix} 1 & \rho_2' \\ \rho_2' & 1 \end{pmatrix}$.

Dempster (1969) uses a sweep operation to obtain the matrices $\underline{G}$ and $\underline{H}$. For our covariance structure, the matrices $\underline{G}$ and $\underline{H}$ are given by

$$\underline{G} = \begin{pmatrix} \sigma_0 & 0 \\ \rho_1' & \sqrt{1-\rho_1'^2} \end{pmatrix} ,$$

$$\underline{H} = \begin{pmatrix} 1 & 0 \\ \rho_2' & \sqrt{1-\rho_2'^2} \end{pmatrix} .$$

Therefore, for π_1 we make the transformation

$$\begin{aligned} X_1 &= \sigma_0 Y_1, \\ X_2 &= \rho_1' Y_1 + \sqrt{1-\rho_1'^2}\, Y_2, \end{aligned} \tag{5.45}$$

and for π_2 we make the transformation

$$\begin{aligned} X_1 &= Y_1 + \theta , \\ X_2 &= \rho_2' Y_1 + \sqrt{1-\rho_2'^2}\, Y_2. \end{aligned} \tag{5.46}$$

The probabilities of misclassification were estimated empirically for each of the three situations. For the univariate unadjusted discriminator, we use the variate X_1 as our discriminator and the classification rule (5.1). For the univariate adjusted discriminator, we use the variate Z in (5.9) as our discriminator and the classification rule (5.12). For the bivariate unadjusted discriminator, we use the variates $(X_1, X_2)'$ as our discriminators and the rule of classification (5.17).

The Monte Carlo method involves substituting the values of the parameters into the respective classification rules, assigning normal random numbers generated from π_1 or π_2 to X_1, Z, and $\underline{X} = (X_1, X_2)'$ and then classifying into π_1

or π_2 according to (5.1), (5.12), and (5.17), respectively. A probability of misclassification is then estimated by the ratio of the number of individuals misclassified to the number tested for one or the other population.

C. Methods of Computing Probabilities of Misclassification for the Trivariate (p=1, q=2) Case

In this section, we shall consider the problem of classification of a trivariate observation $\underline{X} = (X_1, X_2, X_3)'$ into one of two trivariate normal populations π_1 and π_2 where

$$\pi_i : N_3\left(\begin{pmatrix} \mu^{(i)} \\ \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \sigma_i\sigma\rho_i' & \sigma_i\sigma\rho_i' \\ \sigma_i\sigma\rho_i' & \sigma^2 & \sigma^2\rho \\ \sigma_i\sigma\rho_i' & \sigma^2\rho & \sigma^2 \end{pmatrix}\right),$$

$i=1,2$. Hence X_2 and X_3 can be viewed as covariates. We shall assume all parameters are known. Therefore, we could make the transformation

$$X_1 \to \frac{1}{\sigma_2}(X_1 - \mu^{(1)})$$

$$X_2 \to \frac{1}{\sigma}(X_2 - \mu)$$

$$X_3 \to \frac{1}{\sigma}(X_3 - \mu)$$

and, consequently, the populations π_1 and π_2 could be given by

$$\pi_1:\ N_3\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\ \begin{pmatrix}\sigma_0^2 & \sigma_0\rho_1' & \sigma_0\rho_1'\\ \sigma_0\rho_1' & 1 & \rho\\ \sigma_0\rho_1' & \rho & 1\end{pmatrix}\right)$$

and

$$\pi_2:\ N_3\left(\begin{pmatrix}\theta\\0\\0\end{pmatrix},\ \begin{pmatrix}1 & \rho_2' & \rho_2'\\ \rho_2' & 1 & \rho\\ \rho_2' & \rho & 1\end{pmatrix}\right),$$

where

$$\theta = \frac{\mu^{(2)}-\mu^{(1)}}{\sigma_2} \quad \text{and} \quad \sigma_0^2 = \frac{\sigma_1^2}{\sigma_2^2} \quad .$$ We shall assume $\sigma_0^2 > 1$.

As in Section B, three procedures are still available for use in classification, namely

(a) univariate ($p=1$) unadjusted discriminator,

(b) univariate adjusted (for $q=2$ covariates) discriminator, and

(c) trivariate ($p+q = 3$) unadjusted discriminator.

Numerical integration procedures employing Pearson's approximation are used to obtain the probabilities of misclassification for the univariate unadjusted and univariate adjusted discriminator cases. A few approximate values for the probabilities of misclassification were determined by using Pearson's approximation for the trivariate unadjusted discriminator. For the sets of parameters specified, Monte Carlo procedures were used to calculate the probabilities of misclassification empirically for each of these three classification situations.

1. Numerical integration procedures

a. The univariate (p=1) unadjusted discriminator We consider only the variate X_1 as a discriminator ignoring the covariates X_2 and X_3. This situation is identical to that which is explained in Section B-1-a.

b. The univariate adjusted (for q=2 covariates) discriminator As in Section B-1-b, we form the variates

$$Z_1 = X_1 - \underline{\Sigma}_{12}^{(1)}\underline{\Sigma}_{22}^{-1}\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} = X_1 - \frac{1}{1+\rho}\,\sigma_0\rho_1'(X_2+X_3),$$

$$Z_2 = X_1 - \underline{\Sigma}_{12}^{(2)}\underline{\Sigma}_{22}^{-1}\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} = X_1 - \frac{1}{1+\rho}\,\rho_2'(X_2+X_3). \tag{5.47}$$

Then use the weighted variate Z where

$$Z = W_1Z_1 + W_2Z_2 \tag{5.48}$$

as the adjusted discriminator.

Following (5.8), we shall use the weights

$$W_1 = \frac{\sigma_{11.2}^{(2)}}{\sigma_{11.2}^{(1)}+\sigma_{11.2}^{(2)}} = \frac{(1-\frac{2\rho_2'^2}{1+\rho})}{\sigma_0^2(1-\frac{2\rho_1'^2}{1+\rho}) + (1-\frac{2\rho_2'^2}{1+\rho})},\quad W_2=1-W_1. \tag{5.49}$$

Simplifying (5.48) we obtain

$$Z = X_1 - \frac{1}{1+\rho}(W_1\sigma_0\rho_1' + W_2\rho_2')(X_2+X_3), \tag{5.50}$$

where W_i are given in (5.49). If $\underline{X}\varepsilon\pi_1$, then

$Z \sim N(0, V_1)$ where

$$V_1 = \sigma_0^2 + \frac{2}{1+\rho}[(W_1\sigma_0\rho_1' + W_2\rho_2')]^2 - \frac{4\sigma_0\rho_1'}{1+\rho}(W_1\sigma_0\rho_1' + W_2\rho_2'). \tag{5.51}$$

If $\underline{X}\varepsilon\pi_2$, then $Z \sim N(\theta, V_2)$ where

$$V_2 = 1 + \frac{2}{1+\rho}[(W_1\sigma_0\rho_1' + W_2\rho_2')]^2 - \frac{4\rho_2'}{1+\rho}(W_1\sigma_0\rho_1' + W_2\rho_2'). \tag{5.52}$$

Since the univariate adjusted (for q=2 covariates) discriminator Z is now in the same form as given in Section B-1-b, the probabilities of misclassification are given by (5.14) and (5.16).

c. <u>The trivariate (p+q=3) unadjusted discriminator</u>

The trivariate (p+q=3) unadjusted discriminator consists of the variates X_1, X_2, and X_3. The rule of classification for this procedure is given in (5.17). As mentioned previously, there exists a non-singular matrix $\underline{P}$ such that

$$\underline{P}'\underline{\Sigma}^{(1)}\underline{P} = \underline{I},\ \underline{P}'\underline{\Sigma}^{(2)}\underline{P} = \underline{\Lambda} = \text{diag}(\lambda_1,\lambda_2,\lambda_3),$$

where the λ_j's are the roots of $|\underline{\Sigma}^{(2)} - \lambda\underline{\Sigma}^{(1)}| = 0$. If we let $\underline{Y} = \underline{P}'\underline{X}$, then using (5.17), we classify $\underline{X}$ into π_1 if

$$(\underline{Y}-\underline{\nu}^{(2)})'\underline{\Lambda}^{-1}(\underline{Y}-\underline{\nu}^{(2)})-\underline{Y}'\underline{Y} + \log|\underline{\Lambda}| \geq 0, \qquad (5.53)$$

where $\underline{\nu}^{(2)} = \underline{P}'\underline{\mu}^{(2)}$. For the parameter values specified, the matrices $\underline{P}$ and $\underline{\Lambda}$ were determined by the method explained in Section B-1-c. One of the roots of $|\underline{\Sigma}^{(2)}-\lambda\underline{\Sigma}^{(1)}| = 0$ was greater than one, one root was identically equal to one (this can be shown analytically), and the other root was less than one. The roots depend on σ_0^2, ρ, and (ρ_1',ρ_2'). The roots obtained from $(\rho_1', -\rho_2')$ and from $(-\rho_1', \rho_2')$ are the same. Also, the roots from (ρ_1', ρ_2') and from $(-\rho_1', -\rho_2')$ are the same. Hence we only considered $(\rho_1', -\rho_2')$ and (ρ_1', ρ_2').

For the trivariate unadjusted discriminator situation, let λ_3, λ_2, λ_1 denote the roots greater than one, equal to one, less than one, respectively. For the parameter values specified in Tables 9-12, the first component of the eigenvector associated with the eigenvalue $\lambda_2=1$ is zero and so $\nu_2^{(2)} = 0$. The rule of classification given in (5.53) can be written as

$$\left(\frac{1}{\lambda_1} - 1\right)\left(Y_1 - \frac{\nu_1^{(2)}}{1-\lambda_1}\right)^2 - \left(1 - \frac{1}{\lambda_3}\right)\left(Y_3 - \frac{\nu_3^{(2)}}{1-\lambda_3}\right)^2 - \frac{(\nu_1^{(2)})^2}{1-\lambda_1}$$

$$- \frac{(\nu_3^{(2)})^2}{1-\lambda_3} + \log(\lambda_1\lambda_3) \geq 0 \ . \qquad (5.54)$$

Since (5.54) is in the same form as (5.27), we can refer to Section B-1-c to determine the probabilities of misclassification. We shall consider only the case where $\theta \neq 0$. Equation (5.44) could be used to approximate the probabilities of misclassification for non-positive quadratic forms by Pearson's approximation.

2. Monte Carlo procedures

Assume π_1 and π_2 are trivariate normal populations given by

$$\pi_1 : N_3\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_0\rho_1' & \sigma_0\rho_1' \\ \sigma_0\rho_1' & 1 & \rho \\ \sigma_0\rho_1' & \rho & 1 \end{pmatrix}\right) \quad \text{and}$$

$$\pi_2 : N_3\left(\begin{pmatrix} \theta \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2' & \rho_2' \\ \rho_2' & 1 & \rho \\ \rho_2' & \rho & 1 \end{pmatrix}\right).$$

In this section, we will estimate the probabilities of misclassification by Monte Carlo procedures. Following Section B-2, we obtain a triple of independent standard normal pseudorandom variates Y_1, Y_2, and Y_3. To obtain $\underline{X} = (X_1, X_2, X_3)' \epsilon \pi_1$, make the transformation

$$\underline{X} = \underline{G}\ \underline{Y} \quad \text{where} \quad \underline{Y} = (Y_1, Y_2, Y_3)' \quad \text{and}$$

$$\underline{G} = \begin{pmatrix} \sigma_0 & 0 & 0 \\ \rho_1' & \sqrt{1-\rho_1'^2} & 0 \\ \rho_1' & \dfrac{\rho-\rho_1'^2}{\sqrt{1-\rho_1'^2}} & \sqrt{\dfrac{1-2\rho_1'^2-\rho^2+2\rho_1'^2\rho}{1-\rho_1'^2}} \end{pmatrix}.$$

To obtain $\underline{X} = (X_1, X_2, X_3) \varepsilon \pi_2$, make the transformation

$$\underline{X} = \underline{H}\ \underline{Y} + (\theta, 0, 0)'$$

where

$$\underline{H} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_2' & \sqrt{1-\rho_2'^2} & 0 \\ \rho_2' & \dfrac{\rho-\rho_2'^2}{\sqrt{1-\rho_2'^2}} & \sqrt{\dfrac{1-2\rho_2'^2-\rho^2+2\rho_2'^2\rho}{1-\rho_2'^2}} \end{pmatrix} .$$

The probabilities of misclassification were estimated by the relative frequency explained in Section B-2 for each of the three situations. For the univariate unadjusted discriminator, we use the variate X_1 as our discriminator and the classification rule (5.1). For the univariate adjusted (for q=2 covariates) discriminator, we use the variate Z in (5.50) as our discriminator and the classification rule (5.12) where V_1 and V_2 are given by (5.51) and (5.52), respectively. For the trivariate unadjusted discriminator, we use the variates $(X_1, X_2, X_3)'$ as our discriminators and the rule of classification (5.17).

D. Discussion

1. Consideration of weights

In Chapter III, to form the weighted variate $\underline{Z} = W_1\underline{Z}_1+W_2\underline{Z}_2$ where

$$\underline{Z}_1 = \underline{X}_1-\underline{\Sigma}_{12}^{(1)}\underline{\Sigma}_{22}^{-1}\underline{X}_2,$$

$$\underline{Z}_2 = \underline{X}_2-\underline{\Sigma}_{12}^{(2)}\underline{\Sigma}_{22}^{-1}\underline{X}_2,$$

we used as weights

$$W_1 = \frac{|\underline{\Sigma}_{11.2}^{(2)}|}{|\Sigma_{11.2}^{(1)}|+|\underline{\Sigma}_{11.2}^{(2)}|}, \quad W_2 = \frac{|\underline{\Sigma}_{11.2}^{(1)}|}{|\underline{\Sigma}_{11.2}^{(1)}|+|\underline{\Sigma}_{11.2}^{(2)}|}. \qquad (5.55)$$

Note that $\underline{\Sigma}_{11.2}^{(i)}$ is the covariance matrix of $\underline{Z}_i$ when $\underline{X} = \begin{pmatrix}\underline{X}_1\\ \underline{X}_2\end{pmatrix} \epsilon \pi_i$, $i=1,2$. Hence, we are weighting each $\underline{Z}_i$ by the inverse of the determinant of its covariance matrix. These are convenient weights to use in computations since they are constant weights with $W_1+W_2 = 1$. However, they may not be "optimal" in the sense of minimizing $[P(2/1)+P(1/2)]$.

Analogous to the weights given in (5.55) we could use the covariance matrices of $\underline{Z}_i$ in place of the determinants of these covariance matrices of $\underline{Z}_i$. That is, our weights are now the matrices given by

$$\underline{W}_1 = [(\underline{\Sigma}_{11.2}^{(1)})^{-1} + (\underline{\Sigma}_{11.2}^{(2)})^{-1}]^{-1}[\underline{\Sigma}_{11.2}^{(1)}]^{-1},$$

$$\underline{W}_2 = [(\underline{\Sigma}_{11.2}^{(1)})^{-1}+(\underline{\Sigma}_{11.2}^{(2)})^{-1}]^{-1}[\underline{\Sigma}_{11.2}^{(2)}]^{-1}. \tag{5.56}$$

These weights are intuitively appealing based on procedures commonly used in statistics when combining variates. If these weights were used in Chapter III, then the covariance matrices $\underline{V}^{(i)}$ of $\underline{Z}$ given by (3.6) and (3.9) when $\underline{X}\varepsilon\pi_i$, i=1,2, respectively, are now

$$\begin{aligned}\underline{V}^{(i)} &= \underline{W}_1E(\underline{Z}_1-E(\underline{Z}_1))(\underline{Z}_1-E(\underline{Z}_1))'\underline{W}_1' \\ &+ 2\underline{W}_1E(\underline{Z}_1-E(\underline{Z}_1))(\underline{Z}_2-E(\underline{Z}_2))'\underline{W}_2' \\ &+ \underline{W}_2E(\underline{Z}_2-E(\underline{Z}_2))(\underline{Z}_2-E(\underline{Z}_2))'\underline{W}_2' \ . \end{aligned} \tag{5.57}$$

Since $\underline{W}_i$ have intraclass correlation structure and each of the variance and covariance matrices in (5.57) have intraclass correlation structure, it is easily seen that the covariance matrices $\underline{V}^{(i)}$ have intraclass correlation structure. Hence, we could obtain a (pxp) orthogonal matrix $\underline{\Gamma}$ such that

$$\underline{\Gamma}\ \underline{V}^{(i)}\underline{\Gamma}' = \underline{D}^{(i)} = \text{diag}(\alpha_i,\beta_i,\ldots,\beta_i)$$

as done previously. These matrix weights are certainly not as convenient computationally as the constant weights based on determinants. For the case when p=1, the weights given by (5.55) and (5.56) are identical.

A third choice would be weights which minimize [P(2/1)+P(1/2)]. For convenience, let these weights W_1 and

W_2 be constant weights such that $W_1+W_2 = 1$. Define $W = W_1$, then $1-W = W_2$.

For purposes of illustration, suppose $V_1>V_2$ and $\theta=0$, with Z given by (5.7). Hence, the probabilities of misclassification are given by (5.14) and (5.16). They are

$$P(2/1) = P[\chi_1^2 < \frac{V_2}{V_1-V_2} \log \frac{V_1}{V_2}]$$

and

$$P(1/2) = P[\chi_1^2 \geq \frac{V_1}{V_1-V_2} \log \frac{V_1}{V_2}],$$

where V_i are given by

$$V_1 = \sigma_0^2 + [W(\sigma_0\rho_1'-\rho_2')+\rho_2']^2-2[W(\sigma_0\rho_1'-\rho_2')]\sigma_0\rho_1' ,$$

$$V_2 = 1 + [W(\sigma_0\rho_1'-\rho_2')+\rho_2']^2-2[W(\sigma_0\rho_1'-\rho_2']\rho_2' .$$

Hence differentiating $[P(2/1)+P(1/2)]$ with respect to W and setting the expression equal to zero, we obtain

$$f(\frac{V_2}{V_1-V_2}\log\frac{V_1}{V_2}) \frac{d[\frac{V_2}{V_1-V_2}\log\frac{V_1}{V_2}]}{dW} - f(\frac{V_1}{V_1-V_2}\log\frac{V_1}{V_2}) \frac{d[\frac{V_1}{V_1-V_2}\log\frac{V_1}{V_2}]}{dW} = 0, \tag{5.58}$$

where f(x) denotes the probability density function for a central chi-square variate with 1 degree of freedom. Solving for W in Equation (5.58) does not appear feasible analytically. Possibly by computational procedures, one could solve for W.

In general, the expressions for the probabilities of misclassification involve densities of indefinite forms in non-central chi-square variates. If the forms are definite, one might use Pearson's approximation to obtain a central chi-square density and continue as outlined.

These weights are "optimal" in the sense of minimizing $[P(2/1)+P(1/2)]$. However, in practical situations, they may not be computationally convenient.

One interesting case occurs if $\underline{\Sigma}_{12}^{(1)}=\underline{\Sigma}_{12}^{(2)} = \underline{\Sigma}_{12}$, say. Then it is natural to use

$$\underline{Z} = \underline{X}_1 - \underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{X}_2 \tag{5.59}$$

as the adjusted discriminator. By making a linear transformation, assuming the parameters are known, the populations would be π_1: $N_{p+q}\left(\begin{pmatrix}\underline{0}\\ \underline{0}\end{pmatrix}, \underline{\Sigma}^{(1)}\right)$ and π_2: $N_{p+q}\left(\begin{pmatrix}\underline{\theta}\\ \underline{0}\end{pmatrix}, \underline{\Sigma}^{(2)}\right)$, where $\underline{\theta} = \underline{\mu}^{(2)}-\underline{\mu}^{(1)}$ and where $\underline{\Sigma}^{(i)}$ is given in (3.1) except we are assuming $\underline{\Sigma}_{12}^{(1)}=\underline{\Sigma}_{12}^{(2)}=\underline{\Sigma}_{12}$.

Forming the likelihood ratio λ for $\underline{Z}$ in (5.59) we have

$$\begin{aligned} 2\log\lambda = \log\frac{|\underline{\Sigma}_{11.2}^{(2)}|}{|\underline{\Sigma}_{11.2}^{(1)}|} &+ [(\underline{X}_1-\underline{\theta})-\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{X}_2]'[\underline{\Sigma}_{11.2}^{(2)}]^{-1} \\ &\times [(\underline{X}_1-\underline{\theta})-\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{X}_2]-[\underline{X}_1-\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{X}_2]'[\underline{\Sigma}_{11.2}^{(1)}]^{-1} \\ &\times [\underline{X}_1-\underline{\Sigma}_{12}\underline{\Sigma}_{22}^{-1}\underline{X}_2] \ . \end{aligned} \tag{5.60}$$

If all the variates in $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$ are discriminators, we have, using the likelihood ratio,

$$2 \log \lambda = \log \frac{|\underline{\Sigma}^{(2)}|}{|\underline{\Sigma}^{(1)}|} + \begin{pmatrix} \underline{X}_1 - \underline{\theta} \\ \underline{X}_2 \end{pmatrix}' [\underline{\Sigma}^{(2)}]^{-1} \begin{pmatrix} \underline{X}_1 - \underline{\theta} \\ \underline{X}_2 \end{pmatrix}$$

$$- \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}' [\underline{\Sigma}^{(1)}]^{-1} \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}. \qquad (5.61)$$

Using the facts (see, for example, Graybill, 1969) that

$$|\underline{\Sigma}^{(i)}| = |\underline{\Sigma}_{22}| \; |\underline{\Sigma}_{11.2}^{(i)}|,$$

$$[\underline{\Sigma}^{(i)}]^{-1} = \begin{bmatrix} [\underline{\Sigma}_{11.2}^{(i)}]^{-1} & -[\underline{\Sigma}_{11.2}^{(i)}]^{-1} \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1} \\ -\underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} [\underline{\Sigma}_{11.2}^{(i)}]^{-1} & [\underline{\Sigma}_{22.1}^{(i)}]^{-1} \end{bmatrix},$$

$$[\underline{\Sigma}_{22.1}^{(i)}]^{-1} = \underline{\Sigma}_{22}^{-1} + \underline{\Sigma}_{22}^{-1} \underline{\Sigma}_{21} [\underline{\Sigma}_{11.2}^{(i)}]^{-1} \underline{\Sigma}_{12} \underline{\Sigma}_{22}^{-1}, \qquad (5.62)$$

for i=1,2, we see that Equations (5.60) and (5.61) are the same. Therefore, when $\underline{\Sigma}_{12}^{(1)} = \underline{\Sigma}_{12}^{(2)}$, the discriminant functions for the adjusted (for q covariates) discriminator $\underline{Z}$ and for the unadjusted discriminator $\underline{X} = \begin{pmatrix} \underline{X}_1 \\ \underline{X}_2 \end{pmatrix}$ are identical.

2. Construction of tables

As outlined in Section B-2, Monte Carlo procedures were used to estimate the probabilities of misclassification for the three situations considered. To obtain the probabilities of misclassification by Monte Carlo methods for Tables 1-4 (bivariate case), 500 bivariate normal observations were generated from π_1 or π_2 for each parameter set specified. These 500 observations were then used by each of the three procedures given in Section B to estimate the probabilities of misclassification.

For Tables 5-8 (bivariate case), 1000 bivariate normal observations were generated from π_1 or π_2 and used in estimating the probabilities of misclassification for each of the three situations. Tables 1-8 also contain the probabilities of misclassification obtained by Pearson's approximation for the univariate unadjusted and univariate adjusted discriminator situations which were explained in Sections B-1-a and B-1-b, respectively.

For the bivariate unadjusted discriminator situation, Tables 1-4 contain probabilities of misclassification by the numerical integration method derived in Section B-1-c. When $\theta \neq 0$, Tables 5-8 contain some values for the probabilities of misclassification using Pearson's approximation for the bivariate unadjusted discriminator. Since this approximation for a non-positive quadratic form is not extremely accurate, these values are included primarily as a check on the values obtained by the Monte Carlo procedure.

In Tables 9-12 (trivariate case), 1000 trivariate normal observations were generated from π_1 or π_2 and used in estimating the probabilities of misclassification for each of the three situations. Tables 9-12 also contain the probabilities of misclassification obtained by Pearson's approximation for the univariate unadjusted and univariate adjusted discriminator situations.

For the trivariate unadjusted discriminator situation, Tables 9-12 contain a few values obtained by Pearson's approximation for non-positive quadratic forms. These values are included again as a check on the Monte Carlo values.

3. Findings and conclusion

We now examine the tables in the Appendix and compare the probabilities of misclassification for the three classification procedures. For convenience, in this section, we shall refer to the bivariate unadjusted discriminator and the univariate unadjusted discriminator as the bivariate discriminator and the univariate discriminator, respectively.

When $\theta = 0.0$ (Tables 1-4), the bivariate discriminator generally gives smaller probabilities of misclassification than the univariate adjusted discriminator. It is generally much better when $\rho_1' = 0.9$. The univariate adjusted discriminator is also generally superior to the univariate discriminator except when $\rho_1' = 0.9$; $\rho_2' = -0.5, -0.2, 0.0, 0.2, 0.5$, and $\rho_1' = 0.5$; $\rho_2' = -0.2, 0.0, 0.2$. If we look only at P(1/2), the bivariate discriminator is better than the univariate discriminator except when $\rho_1' = 0.9$; $\rho_2' = -0.5, -0.2, 0.0, 0.2, 0.5$. However, in these parameter situations, [P(2/1)+P(1/2)] is smaller for the bivariate discriminator. As expected, the probabilities of misclassification are usually less for all three classification procedures when $\sigma_0^2 = 4.0$, $\theta = 0.0$ rather than $\sigma_0^2 = 2.0$, $\theta = 0.0$.

When $\theta = 2.0$ (Tables 5-8), the bivariate discriminator and the univariate adjusted discriminator both have smaller P(2/1) than the univariate discriminator. For P(1/2), they are usually better than the univariate discriminator except when $\rho_1' = 0.9$; $\rho_2' = -0.5, -0.2, 0.0, 0.2, 0.5$. However, in these parameter situations, [P(2/1)+P(1/2)] is less for the bivariate discriminator. The bivariate discriminator and the univariate adjusted discriminator have probabilities of misclassification essentially the same except for P(1/2) when the bivariate discriminator is much better for $\rho_1' = 0.9$; $\rho_2' = -0.5, -0.2$.

Confining our attention to Tables 1-8, it is clear that the probabilities of misclassification for respective values of σ_0^2, (ρ_1', ρ_2'), for the bivariate and the univariate adjusted discriminators are less when $\theta = 2.0$ than for $\theta = 0.0$.

Again, this seems to be what one would expect.

Focusing on Tables 9-12, we observe that the probabilities of misclassification for the trivariate discriminator and the univariate adjusted discriminator are generally the same except for $\rho_1' = 0.7$; $\rho_2' = -0.7, -0.5$, when P(1/2) for the trivariate discriminator is much less than for the univariate adjusted discriminator. They are both generally better than the univariate discriminator except for P(1/2) when $\rho = 0.0$; $\rho_1' = 0.2, 0.5$; $\rho_2' = -0.2, 0.0, 0.2$, and when $\rho = 0.5$; $\rho_1' = 0.5, 0.7$; $\rho_2' = -0.2, 0.0, 0.2$.

Comparing Tables 5 vs. 9 and 6 vs. 10, we find that using two covariates when $\rho = 0.0$ is superior to using just one covariate when $\sigma_0^2 = 2.0$, $\theta = 2.0$, since more information is available for the former case.

In conclusion, it appears based on our study that if covariates are available it is advantageous to use them either as discriminators or as covariates under intraclass correlation models. Our study shows that, in general, the performances are the same whether we use them as discriminators or as covariates. For a few special cases, the former is better. This probably is due to the fact that in constructing the adjusted discriminators, we used the intuitive appealing and easily computed weights instead of "optimal weights." Since computation for "optimal weights" appears to be difficult, this problem needs further consideration.

We have assumed in this thesis that the population covariance matrices have intraclass correlation structure which occur in practice (see, for example, Bartlett and Please, 1963). In such a model, the number of parameters is not very large. If the investigator has arbitrary covariance matrices, then the number of parameters which need to be considered becomes extremely large. To compare probabilities of misclassification for the various classification procedures, useful approximations need to be found or one will have to use empirical results.

VI. REFERENCES

Abramowitz, M. and Stegun, I. A., eds. 1965. Handbook of mathematical functions. U.S. Department of Commerce, National Bureau of Standards Applied Mathematics Series.

Anderson, T. W. 1951. Classification by multivariate analysis. Psychometrika 16: 31-50.

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. John Wiley and Sons, Inc., New York.

Anderson, T. W. and Bahadur, R. R. 1962. Classification into two multivariate normal distributions with different covariance matrices. Annals of Mathematical Statistics 33: 420-431.

Bartlett, M. S. and Please, N. W. 1963. Discrimination in the case of zero mean differences. Biometrika 50: 17-21.

Bellman, R. 1960. Introduction to matrix analysis. McGraw-Hill, Inc., New York.

Box, G. E. P. and Muller, M. E. 1958. A note on the generation of random normal deviates. Annals of Mathematical Statistics 29: 610-611.

Cochran, W. G. 1964a. Comparison of two methods of handling covariates in discriminatory analysis. Annals of the Institute of Statistical Mathematics 16: 43-53.

Cochran, W. G. 1964b. On the performance of the linear discriminant function. Technometrics 6: 179-190.

Cochran, W. G. and Bliss, C. I. 1948. Discriminant functions with covariance. Annals of Mathematical Statistics 19: 151-176.

Cooley, W. W. and Lohnes, P. R. 1962. Multivariate procedures for the behavioral sciences. John Wiley and Sons, Inc., New York.

Cramér, H. 1946. Mathematical methods of statistics. Princeton University Press, Princeton, N.J.

Dempster, A. P. 1969. Elements of continuous multivariate analysis. Addison-Wesley Publishing Company, Reading, Massachusetts.

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

Gilbert, E. S. 1969. The effect of unequal variance-covariance matrices on Fisher's linear discriminant function. Biometrics 25: 505-515.

Graybill, F. A. 1969. Introduction to matrices with applications in statistics. Wadsworth Publishing Company, Inc., Belmont, California.

Gurland, J. 1955. Distribution of definite and of indefinite quadratic forms. Annals of Mathematical Statistics 26: 122-127.

Gurland, J. 1962. Correction to 'Distribution of definite and of indefinite quadratic forms.' Annals of Mathematical Statistics 33: 813.

Han, C. P. 1968. A note on discrimination in the case of unequal covariance matrices. Biometrika 55: 586-587.

Han, C. P. 1969. Distribution of discriminant function when covariance matrices are proportional. Annals of Mathematical Statistics 40: 979-985.

Han, C. P. 1970. Distribution of discriminant function in circular models. Annals of the Institute of Statistical Mathematics 22: 117-125.

Hartley, H. O. 1938. Studentization and large-sample theory. Journal of the Royal Statistical Society Supplement 5: 80-88.

Imhof, J. 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48: 419-426.

Ito, K. 1956. Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic. I. Annals of Mathematical Statistics 27: 1091-1105.

Ito, K. 1960. Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic. II. Annals of Mathematical Statistics 31: 1148-1153.

John, S. 1960. On some classification problem-I. Sankhyā, 22: 301-308.

Kshirsagar, A. M. 1972. Multivariate analysis. Marcel Dekker, Inc., New York.

Kudo, A. 1959. The classificatory problem viewed as a two-decision problem. Kyushu University, Japan. Memoirs of the Faculty of Science, Series A, 13: 96-125.

Mahalanobis, P. C. 1930. On tests and measures of group divergence. Journal of the Asiatic Society of Bengal 26: 541-588.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India 2: 49-55.

Marks, S. 1970. Discriminant functions when covariance matrices are unequal. Unpublished Ph.D. thesis. Library, The University of California, Los Angeles, California.

Marsaglia, G. and Bray, T. A. 1968. One-line random number generators and their use in combinations. Communications of the Association for Computing Machinery 11: 757-759.

Memon, A. Z. 1968. Z statistic in discriminant analysis. Unpublished Ph.D. thesis. Library, Iowa State University of Science and Technology, Ames, Iowa.

Memon, A. Z. and Okamoto, M. 1970. The classification statistic W* in covariate discriminant analysis. Annals of Mathematical Statistics 41: 1491-1499.

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

Okamoto, M. 1961. Discrimination for variance matrices. Osaka Mathematical Journal 13: 1-39.

Okamoto, M. 1963. An asymptotic expansion for the distribution of the linear discriminant function. Annals of Mathematical Statistics 34: 1286-1301.

Okamoto, M. 1968. Correction to 'An asymptotic expansion for the distribution of the linear discriminant function.' Annals of Mathematical Statistics 39: 1358-1359.

Pachares, J. 1952. The distribution of the difference of two independent chi-squares. Annals of Mathematical Statistics 23: 639.

Patnaik, P. B. 1949. The non-central χ^2- and F-distributions and their applications. Biometrika 36: 202-232.

Pearson, E. S. 1959. Note on an approximation to the distribution on non-central χ^2. Biometrika 46: 364.

Pearson, K. 1926. On the coefficient of racial likeness. Biometrika 18: 105-117.

Penrose, L. S. 1947. Some notes on discrimination. Annals of Eugenics, London 13: 228-237.

Press, S. J. 1964. Some hypothesis testing problems involving multivariate normal distributions with unequal and intraclass structured covariance matrices. Stanford University, Department of Statistics, Technical Report No. 12.

Press, S. J. 1966. Linear combinations of non-central chi-square variates. Annals of Mathematical Statistics 37: 480-487.

Press, S. J. 1972. Applied multivariate analysis. Holt, Rinehart, and Winston, Inc., New York.

Ralston, A. 1965. A first course in numerical analysis. McGraw-Hill, Inc., New York.

Rao, C. R. 1949. On some problem arising out of discrimination with multiple characters. Sankhyā 9: 343-366.

Rao, C. R. 1966. Covariance adjustment and related problems in multivariate analysis. In Krishnaiah, P. K., ed. Multivariate analysis. Pp. 87-103. Academic Press, New York.

Robbins, H. and Pitman, E. J. G. 1949. Application of the method of mixtures to quadratic forms in normal variates. Annals of Mathematical Statistics 20: 552-560.

Ruben, H. 1963. A new result on the distribution of quadratic forms. Annals of Mathematical Statistics 34: 1582-1584.

Shah, B. K. 1963. Distribution of definite and of indefinite quadratic forms from a non-central normal distribution. Annals of Mathematical Statistics 34: 186-190.

Shah, B. K. and Khatri, C. G. 1961. Distribution of a definite quadratic form for non-central normal variates. Annals of Mathematical Statistics 32: 883-887.

Shah, B. K. and Khatri, C. G. 1963. Correction to 'Distribution of a definite quadratic form for non-central normal variates.' Annals of Mathematical Statistics 34: 673.

Sitgreaves, R. 1952. On the distribution of two random matrices used in classification procedures. Annals of Mathematical Statistics 23: 263-270.

Slater, L. J. 1960. Confluent hypergeometric functions. Cambridge University Press, London.

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

Subrahmaniam, K. 1970. Discrimination in the presence of covariables. The University of Manitoba, Department of Statistics, Technical Report No. 3.

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

Wallace, D. L. 1958. Asymptotic approximations to distributions. Annals of Mathematical Statistics 29: 635-654.

Welch, B. L. 1939. Note on discriminant functions. Biometrika 31: 218-220.

Welch, B. L. 1947. The generalization of 'Student's' problem when several different population variances are involved. Biometrika 34: 28-35.

VII. ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. Chien-Pai Han for introducing this problem and for his valuable suggestions and guidance throughout the research and preparation of this thesis.

Financial support from the Public Health Service Grant No. 5T01 GM00034-13 is acknowledged.

Special thanks go to my wife, Linda, for her patience and encouragement during my graduate study.

VIII. APPENDIX

Table 1. Probability of misclassification P(2/1) for σ_0^2 = 2.0, θ = 0.0, (ρ_1', ρ_2')

ρ_1'	ρ_2'	Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
		Monte Carlo	Numerical Integration	Monte Carlo	Exact	Monte Carlo	Exact
0.0	-0.9	.3600	.3310	.3800	.3506		
0.0	-0.5	.5540	.5292	.5580	.5518		
0.0	-0.2	.6020	.5842	.6060	.5889		
0.0	0.0	.5830	.5949	.5830	.5949	.5830	.5949
0.0	0.2	.5920	.5842	.5860	.5889		
0.0	0.5	.5140	.5295	.5260	.5518		
0.0	0.9	.3080	.3310	.3340	.3506		
0.2	-0.9	.3120	.3060	.3340	.3378		
0.2	-0.5	.5200	.4963	.5780	.5504		
0.2	-0.2	.5820	.5599	.6040	.5915		
0.2	0.0	.6240	.5874	.6220	.5991		
0.2	0.2	.5900	.5973	.5820	.5948		
0.2	0.5	.5680	.5569	.5640	.5614		
0.2	0.9	.3360	.3588	.3460	.3696		
0.5	-0.9	.2720	.2685	.3400	.3336		
0.5	-0.5	.4140	.4291	.5700	.5734		
0.5	-0.2	.4880	.4836	.6040	.6188		
0.5	0.0	.5220	.5203	.6240	.6275		
0.5	0.2	.5980	.5603	.6500	.6241		
0.5	0.5	.6000	.5866	.6100	.5941		
0.5	0.9	.3900	.4121	.3940	.4138		
0.9	-0.9	.1920	.1736	.5260	.5222		
0.9	-0.5	.1480	.1355	.2000	.2023		
0.9	-0.2	.1380	.1388	.1940	.1797		
0.9	0.0	.1340	.1470	.1840	.1810		
0.9	0.2	.1980	.1604	.1960	.1911		
0.9	0.5	.1940	.1984	.2380	.2264		
0.9	0.9	.5220	.5070	.6120	.5854		

[a]These values do not depend on (ρ_1', ρ_2').

Table 2. Probability of misclassification P(1/2) for σ_0^2 = 2.0, θ = 0.0, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Numerical Integration	Monte Carlo	Exact	Monte Carlo	Exact
0.0	-0.9	.0720	.0895	.0680	.0931		
0.0	-0.5	.1860	.1976	.2000	.2065		
0.0	-0.2	.2320	.2314	.2300	.2343		
0.0	0.0	.2310	.2390	.2310	.2390	.2310	.2390
0.0	0.2	.2440	.2315	.2580	.2343		
0.0	0.5	.1920	.1976	.2220	.2065		
0.0	0.9	.0860	.0895	.0760	.0931		
0.2	-0.9	.0700	.0830	.0720	.0876		
0.2	-0.5	.1780	.1897	.1860	.2056		
0.2	-0.2	.2160	.2264	.2240	.2363		
0.2	0.0	.2180	.2371	.2200	.2424		
0.2	0.2	.2340	.2347	.2240	.2389		
0.2	0.5	.2020	.2108	.2200	.2135		
0.2	0.9	.0740	.0990	.0840	.1016		
0.5	-0.9	.0700	.0773	.0700	.0858		
0.5	-0.5	.2000	.1912	.2220	.2224		
0.5	-0.2	.2520	.2398	.2700	.2586		
0.5	0.0	.2660	.2555	.2820	.2660		
0.5	0.2	.2480	.2551	.2280	.2631		
0.5	0.5	.2360	.2344	.2240	.2384		
0.5	0.9	.1440	.1223	.1500	.1229		
0.9	-0.9	.1240	.0889	.2040	.1862		
0.9	-0.5	.3060	.2768	.5320	.5458		
0.9	-0.2	.3400	.3321	.5060	.5122		
0.9	0.0	.3500	.3641	.5280	.5143		
0.9	0.2	.3640	.3965	.5000	.5295		
0.9	0.5	.4760	.4478	.5860	.5786		
0.9	0.9	.1720	.2135	.2420	.2316		

[a]These values do not depend on (ρ_1', ρ_2').

Table 3. Probability of misclassification P(2/1) for $\sigma_0^2 = 4.0$, $\theta = 0.0$, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Numerical Integration	Monte Carlo	Exact	Monte Carlo	Exact
0.0	-0.9	.2520	.2808	.2600	.2869		
0.0	-0.5	.4560	.4551	.4700	.4612		
0.0	-0.2	.4700	.4961	.4660	.4973		
0.0	0.0	.4980	.5034	.4980	.5034	.4980	.5034
0.0	0.2	.5180	.4961	.5160	.4973		
0.0	0.5	.4380	.4551	.4500	.4612		
0.0	0.9	.2780	.2808	.2720	.2869		
0.2	-0.9	.2640	.2633	.2820	.2756		
0.2	-0.5	.3960	.4361	.4260	.4569		
0.2	-0.2	.4720	.4860	.5020	.4976		
0.2	0.0	.4860	.5008	.4900	.5062		
0.2	0.2	.4880	.5014	.4940	.5027		
0.2	0.5	.4640	.4602	.4720	.4709		
0.2	0.9	.2660	.2998	.2740	.3019		
0.5	-0.9	.2540	.2361	.2780	.2663		
0.5	-0.5	.3940	.3967	.4720	.4669		
0.5	-0.2	.4360	.4519	.5040	.5150		
0.5	0.0	.4340	.4775	.5140	.5267		
0.5	0.2	.4780	.4935	.5360	.5258		
0.5	0.5	.4780	.4878	.4980	.4983		
0.5	0.9	.3400	.3332	.3440	.3339		
0.9	-0.9	.1840	.1678	.4120	.3728		
0.9	-0.5	.2280	.2028	.6660	.6801		
0.9	-0.2	.1380	.1900	.2860	.2753		
0.9	0.0	.1800	.1979	.2960	.2701		
0.9	0.2	.2300	.2161	.2880	.2782		
0.9	0.5	.3120	.3004	.6660	.6805		
0.9	0.9	.3700	.3682	.4540	.4561		

[a]These values do not depend on (ρ_1', ρ_2').

Table 4. Probability of misclassification P(1/2) for σ_0^2 = 4.0, θ = 0.0, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Numerical Integration	Monte Carlo	Exact	Monte Carlo	Exact
0.0	-0.9	.0820	.0666	.0800	.0675		
0.0	-0.5	.1300	.1461	.1220	.1486		
0.0	-0.2	.2080	.1696	.2120	.1702		
0.0	0.0	.1700	.1740	.1700	.1740	.1700	.1740
0.0	0.2	.1620	.1696	.1700	.1702		
0.0	0.5	.1480	.1461	.1680	.1486		
0.0	0.9	.0780	.0666	.0820	.0675		
0.2	-0.9	.0720	.0621	.0600	.0634		
0.2	-0.5	.1420	.1395	.1480	.1461		
0.2	-0.2	.1900	.1655	.1940	.1704		
0.2	0.0	.1740	.1731	.1780	.1758		
0.2	0.2	.1920	.1728	.1960	.1736		
0.2	0.5	.1520	.1347	.1580	.1542		
0.2	0.9	.0820	.0727	.0800	.0731		
0.5	-0.9	.0400	.0577	.0360	.0601		
0.5	-0.5	.1380	.1357	.1760	.1519		
0.5	-0.2	.1700	.1646	.1860	.1814		
0.5	0.0	.2040	.1747	.2140	.1892		
0.5	0.2	.1560	.1776	.1680	.1886		
0.5	0.5	.1620	.1662	.1620	.1708		
0.5	0.9	.0800	.0913	.0800	.0860		
0.9	-0.9	.0600	.0625	.0820	.1031		
0.9	-0.5	.1900	.2016	.3260	.3147		
0.9	-0.2	.3120	.2753	.6320	.6381		
0.9	0.0	.2920	.3035	.6280	.6322		
0.9	0.2	.3200	.3214	.6220	.6414		
0.9	0.5	.2960	.2882	.3360	.3151		
0.9	0.9	.1140	.1271	.1640	.1457		

[a]These values do not depend on (ρ_1', ρ_2').

Table 5. Probability of misclassification P(2/1) for $\sigma_0^2 = 2.0$, $\theta = 2.0$, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.9	.2060		.2090	.1923		
0.0	-0.5	.2820		.2710	.2455		
0.0	-0.2	.2380	.2481	.2580	.2475		
0.0	0.0	.2530	.2474	.2530	.2474	.2530	.2474
0.0	0.2	.2530		.2520	.2475		
0.0	0.5	.2650		.2660	.2455		
0.0	0.9	.1830		.2070	.1923		
0.2	-0.9	.2110	.1950	.2240	.2020		
0.2	-0.5	.2490		.2400	.2524		
0.2	-0.2	.2710		.2430	.2485		
0.2	0.0	.2640		.2440	.2447		
0.2	0.2	.2430	.2416	.2430	.2415		
0.2	0.5	.2430		.2440	.2346		
0.2	0.9	.1810		.1850	.1790		
0.5	-0.9	.1650		.2070	.2139		
0.5	-0.5	.2640		.2740	.2531		
0.5	-0.2	.2510		.2500	.2358		
0.5	0.0	.2250		.2190	.2261		
0.5	0.2	.2190		.2280	.2184		
0.5	0.5	.2170	.2065	.2170	.2058		
0.5	0.9	.1480	.1462	.1640	.1464		
0.9	-0.9	.1270		.2750	.2573		
0.9	-0.5	.0880		.1130	.1080		
0.9	-0.2	.1000		.0940	.0877		
0.9	0.0	.0880		.0910	.0830		
0.9	0.2	.0910		.0940	.0812		
0.9	0.5	.0900	.1040	.0850	.0800		
0.9	0.9	.0530		.0410	.0445		

[a]These values do not depend on (ρ_1', ρ_2').

Table 6. Probability of misclassification P(1/2) for σ_0^2 = 2.0, θ = 2.0, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
ρ_1'	ρ_2'	Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
		Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.9	.0370		.0380	.0438		
0.0	-0.5	.1040		.1020	.1118		
0.0	-0.2	.1390	.1406	.1400	.1362		
0.0	0.0	.1440	.1407	.1440	.1407	.1440	.1407
0.0	0.2	.1410		.1410	.1362		
0.0	0.5	.1320		.1230	.1118		
0.0	0.9	.0510		.0480	.0438		
0.2	-0.9	.0530	.0554	.0450	.0448		
0.2	-0.5	.1400		.1300	.1135		
0.2	-0.2	.1500		.1500	.1385		
0.2	0.0	.1400		.1160	.1428		
0.2	0.2	.1540	.1387	.1540	.1379		
0.2	0.5	.1190		.1150	.1130		
0.2	0.9	.0410		.0360	.0426		
0.5	-0.9	.0550		.0540	.0475		
0.5	-0.5	.1190		.1270	.1276		
0.5	-0.2	.1340		.1550	.1552		
0.5	0.0	.1670		.1550	.1577		
0.5	0.2	.1580		.1600	.1500		
0.5	0.5	.1200	.1249	.1210	.1204		
0.5	0.9	.0460	.0419	.0440	.0398		
0.9	-0.9	.0650		.1010	.1016		
0.9	-0.5	.1750		.2630	.2437		
0.9	-0.2	.1830		.2350	.2268		
0.9	0.0	.2200		.2330	.2106		
0.9	0.2	.1850		.1900	.1894		
0.9	0.5	.1300	.1446	.1330	.1416		
0.9	0.9	.0260		.0280	.0277		

[a]These values do not depend on (ρ_1', ρ_2').

Table 7. Probability of misclassification P(2/1) for $\sigma_0^2 =$ 4.0, θ = 2.0, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.9	.1910		.2020	.2159		
0.0	-0.5	.3250		.3270	.2990		
0.0	-0.2	.3140	.3134	.3140	.3136		
0.0	0.0	.3100	.3158	.3100	.3158	.3100	.3158
0.0	0.2	.3230		.3160	.3136		
0.0	0.5	.3260		.3220	.2990		
0.0	0.9	.2050		.2190	.2159		
0.2	-0.9	.2170	.2135	.2110	.2183		
0.2	-0.5	.3210		.3140	.3029		
0.2	-0.2	.3400		.3520	.3152		
0.2	0.0	.3460		.3530	.3152		
0.2	0.2	.3270	.3108	.3280	.3110		
0.2	0.5	.3020		.3010	.2941		
0.2	0.9	.2030		.2150	.2116		
0.5	-0.9	.1800		.1990	.2211		
0.5	-0.5	.3010		.3420	.3094		
0.5	-0.2	.3040		.3240	.3155		
0.5	0.0	.3310		.3430	.3110		
0.5	0.2	.3100		.3190	.3029		
0.5	0.5	.3010	.2807	.3060	.2818		
0.5	0.9	.1980	.1979	.1970	.1980		
0.9	-0.9	.1580		.2930	.2740		
0.9	-0.5	.1590		.2560	.2463		
0.9	-0.2	.1330		.1960	.1920		
0.9	0.0	.1340		.1860	.1791		
0.9	0.2	.1330		.1740	.1760		
0.9	0.5	.1800	.2008	.1980	.1796		
0.9	0.9	.1420		.1290	.1360		

[a]These values do not depend on (ρ_1', ρ_2').

Table 8. Probability of misclassification P(1/2) for σ_0^2 = 4.0, θ = 2.0, (ρ_1', ρ_2')

		Bivariate (p=1, q=1) Case					
		Bivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.9	.0380		.0370	.0426		
0.0	-0.5	.0900		.0850	.0939		
0.0	-0.2	.1040	.1115	.1010	.1091		
0.0	0.0	.1160	.1120	.1160	.1120	.1160	.1120
0.0	0.2	.1090		.1180	.1091		
0.0	0.5	.1010		.1000	.0939		
0.0	0.9	.0440		.0490	.0426		
0.2	-0.9	.0410	.0496	.0350	.0421		
0.2	-0.5	.0880		.0780	.0940		
0.2	-0.2	.1160		.1260	.1097		
0.2	0.0	.1010		.1170	.1129		
0.2	0.2	.1040	.1128	.1030	.1103		
0.2	0.5	.0920		.0760	.0950		
0.2	0.9	.0340		.0460	.0431		
0.5	-0.9	.0370		.0350	.0421		
0.5	-0.5	.1070		.1070	.0986		
0.5	-0.2	.1290		.1220	.1165		
0.5	0.0	.1220		.1370	.1204		
0.5	0.2	.1100		.1080	.1175		
0.5	0.5	.1070	.1170	.1180	.0999		
0.5	0.9	.0530	.0435	.0590	.0433		
0.9	-0.9	.0500		.0710	.0697		
0.9	-0.5	.1630		.2350	.2414		
0.9	-0.2	.2000		.2610	.2641		
0.9	0.0	.2370		.2760	.2541		
0.9	0.2	.2240		.2500	.2328		
0.9	0.5	.1560	.1999	.1620	.1771		
0.9	0.9	.0500		.0460	.0437		

[a]These values do not depend on (ρ_1', ρ_2').

Table 9. Probability of misclassification P(2/1) for σ_0^2 = 2.0, θ = 2.0, ρ = 0.0, (ρ_1', ρ_2')

ρ_1'	ρ_2'	Trivariate (p=1, q=2) Case: Trivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
		Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.7	.0790		.0810	.1390		
0.0	-0.5	.2360		.2430	.2339		
0.0	-0.2	.2640	.2484	.2630	.2475		
0.0	0.0	.2580	.2474	.2580	.2474	.2580	.2474
0.0	0.2	.2600		.2590	.2475		
0.0	0.5	.2600		.2630	.2339		
0.0	0.7	.0900		.0920	.1390		
0.2	-0.7	.0740	.1479	.0890	.1490		
0.2	-0.5	.2370		.2580	.2474		
0.2	-0.2	.2640		.2630	.2495		
0.2	0.0	.2360		.2380	.2418		
0.2	0.2	.2540	.2356	.2540	.2353		
0.2	0.5	.2260		.2260	.2133		
0.2	0.7	.0730		.0770	.1207		
0.5	-0.7	.0700		.0810	.1562		
0.5	-0.5	.2160		.2810	.2565		
0.5	-0.2	.2090		.2220	.2051		
0.5	0.0	.1920		.1930	.2114		
0.5	0.2	.1600		.1570	.1748		
0.5	0.5	.1450	.1501	.1450	.1490		
0.5	0.7	.0440		.0440	.0610		
0.7	-0.7	.0500		.2710	.2562		
0.7	-0.5	.0130		.0190	.0232		
0.7	-0.2	.0290		.0230	.0190		
0.7	0.0	.0150		.0160	.0181		
0.7	0.2	.0180		.0150	.0164		
0.7	0.5	.0100	.0076	.0050	.0095		
0.7	0.7	.0000	.0006	.0000	.0001		

[a]These values do not depend on (ρ_1', ρ_2').

Table 10. Probability of misclassification P(1/2) for σ_0^2 = 2.0, θ = 2.0, ρ = 0.0, (ρ_1', ρ_2')

		Trivariate (p=1, q=2) Case					
		Trivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.7	.0110		.0120	.0118		
0.0	-0.5	.0760		.0750	.0821		
0.0	-0.2	.1350	.1401	.1340	.1316		
0.0	0.0	.1460	.1407	.1460	.1407	.1460	.1407
0.0	0.2	.1390		.1380	.1316		
0.0	0.5	.0910		.0850	.0821		
0.0	0.7	.0100		.0100	.0118		
0.2	-0.7	.0110	.0035	.0110	.0121		
0.2	-0.5	.0860		.0870	.0844		
0.2	-0.2	.1430		.1420	.1364		
0.2	0.0	.1470		.1480	.1451		
0.2	0.2	.1600	.1365	.1600	.1350		
0.2	0.5	.0870		.0890	.0829		
0.2	0.7	.0110		.0110	.0108		
0.5	-0.7	.0160		.0160	.0124		
0.5	-0.5	.1020		.1380	.1154		
0.5	-0.2	.1890		.2070	.1875		
0.5	0.0	.1790		.1760	.1927		
0.5	0.2	.1530		.1550	.1626		
0.5	0.5	.1040	.0964	.1050	.0899		
0.5	0.7	.0070		.0070	.0059		
0.7	-0.7	.0150		.0850	.0946		
0.7	-0.5	.0890		.1380	.1736		
0.7	-0.2	.0860		.1070	.1589		
0.7	0.0	.1090		.1200	.1472		
0.7	0.2	.0880		.1010	.1274		
0.7	0.5	.0450	.0631	.0520	.0621		
0.7	0.7	.0000	.0003	.0000	.0001		

[a]These values do not depend on (ρ_1', ρ_2').

Table 11. Probability of misclassification P(2/1) for σ_0^2 = 2.0, θ = 2.0, ρ = 0.5, (ρ_1', ρ_2')

		Trivariate (p=1, q=2) Case					
		Trivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.7	.2260		.2250	.2186		
0.0	-0.5	.2750		.2750	.2431		
0.0	-0.2	.2670	.2482	.2680	.2475		
0.0	0.0	.2540	.2474	.2540	.2474	.2540	.2474
0.0	0.2	.2490		.2490	.2475		
0.0	0.5	.2640		.2690	.2431		
0.0	0.7	.2200		.2260	.2186		
0.2	-0.7	.2370	.2211	.2660	.2305		
0.2	-0.5	.2660		.2760	.2523		
0.2	-0.2	.2540		.2520	.2488		
0.2	0.0	.2590		.2590	.2438		
0.2	0.2	.2390	.2396	.2400	.2395		
0.2	0.5	.2520		.2540	.2286		
0.2	0.7	.2420		.2430	.2021		
0.5	-0.7	.1920		.2530	.2465		
0.5	-0.5	.2550		.2730	.2545		
0.5	-0.2	.2390		.2440	.2288		
0.5	0.0	.2010		.1950	.2156		
0.5	0.2	.2010		.2030	.2060		
0.5	0.5	.1800	.1899	.1790	.1890		
0.5	0.7	.1580		.1580	.1595		
0.7	-0.7	.1530		.2620	.2574		
0.7	-0.5	.1790		.2290	.2056		
0.7	-0.2	.1330		.1620	.1543		
0.7	0.0	.1270		.1340	.1422		
0.7	0.2	.1240		.1310	.1368		
0.7	0.5	.1370	.1378	.1370	.1276		
0.7	0.7	.0940	.1035	.0950	.1025		

[a]These values do not depend on (ρ_1', ρ_2').

Table 12. Probability of misclassification P(1/2) for σ_0^2 = 2.0, θ = 2.0, ρ = 0.5, (ρ_1', ρ_2')

		Trivariate (p=1, q=2) Case					
		Trivariate Unadjusted Discriminator		Univariate Adjusted Discriminator		Univariate[a] Unadjusted Discriminator	
ρ_1'	ρ_2'	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation	Monte Carlo	Pearson's Approximation
0.0	-0.7	.0710		.0700	.0639		
0.0	-0.5	.0860		.0870	.1019		
0.0	-0.2	.1270	.1404	.1270	.1347		
0.0	0.0	.1330	.1407	.1330	.1407	.1330	.1407
0.0	0.2	.1330		.1320	.1347		
0.0	0.5	.1090		.1040	.1019		
0.0	0.7	.0850		.0780	.0639		
0.2	-0.7	.0800	.1008	.0660	.0653		
0.2	-0.5	.1060		.1120	.1039		
0.2	-0.2	.1340		.1340	.1378		
0.2	0.0	.1520		.1540	.1436		
0.2	0.2	.1260	.1380	.1260	.1370		
0.2	0.5	.1000		.1000	.1032		
0.2	0.7	.0680		.0690	.0632		
0.5	-0.7	.0690		.0710	.0733		
0.5	-0.5	.1060		.1140	.1234		
0.5	-0.2	.1460		.1560	.1634		
0.5	0.0	.1700		.1780	.1657		
0.5	0.2	.1650		.1640	.1539		
0.5	0.5	.1190	.1172	.1200	.1117		
0.5	0.7	.0790		.0800	.0635		
0.7	-0.7	.0830		.1050	.1084		
0.7	-0.5	.1250		.1980	.1956		
0.7	-0.2	.2190		.2280	.2172		
0.7	0.0	.2050		.2130	.2052		
0.7	0.2	.1830		.1780	.1820		
0.7	0.5	.1250	.1349	.1250	.1237		
0.7	0.7	.0730	.0677	.0730	.0631		

[a]These values do not depend on (ρ_1', ρ_2').