Analysis of Multivariate Data Using Kotz Type Distribution

Kusaya Plungpongpun

Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/mathstat_etds

Part of the Statistics and Probability Commons

Recommended Citation
Plungpongpun, Kusaya. "Analysis of Multivariate Data Using Kotz Type Distribution" (2003). Doctor of Philosophy (PhD), Dissertation, Mathematics & Statistics, Old Dominion University, DOI: 10.25777/6rcj-ch14
https://digitalcommons.odu.edu/mathstat_etds/51

This Dissertation is brought to you for free and open access by the Mathematics & Statistics at ODU Digital Commons. It has been accepted for inclusion in Mathematics & Statistics Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
ANALYSIS OF MULTIVARIATE DATA USING KOTZ TYPE DISTRIBUTION

by

Kusaya Plungpongpun
B.S. March 1983, Kasetsart University, Thailand
M.S. July 1986, National Institute of Development Administration, Thailand
M.S. December 1997, North Carolina State University

A Dissertation Submitted to the Faculty of Old Dominion University in Partial Fulfillment of the Requirement for the Degree of
DOCTOR OF PHILOSOPHY
COMPUTATIONAL AND APPLIED MATHEMATICS
OLD DOMINION UNIVERSITY
August 2003

Approved by:

Dayanand W. Naik (Director)

N. Rao Charanty (Member)

Larry D. Lee (Member)

Edward Markowski (Member)
ABSTRACT

ANALYSIS OF MULTIVARIATE DATA USING KOTZ TYPE DISTRIBUTION

Kusaya Plungpongpun
Old Dominion University, 2003
Director: Dr. Dayanand N. Naik

Most of the inferential statistical methods for multivariate data are developed under the fundamental assumption that the data are from a multivariate normal distribution. Unfortunately, one can never be sure a set of data is really from a multivariate normal distribution. There are numerous methods for checking (testing) multivariate normality, but based on many published and our own simulation studies, provided in the first chapter of this dissertation, we observe that these tests are generally not very powerful, especially for smaller sample sizes. Hence it is always beneficial to have alternative multivariate distributions available along with the methodology for using them.

In this dissertation, we focus on a probability distribution, called the Kotz type distribution, which has fatter tail regions than that of multivariate normal distribution and has its probability density function (pdf) in the form

\[ f(x, \mu, \Sigma) = c \left| \Sigma \right|^{-\frac{1}{2}} \exp \left\{ -[(x - \mu)^\top \Sigma^{-1} (x - \mu)]^{\frac{1}{2}} \right\}, \]

where \( \mu \in \mathbb{R}^p \), \( \Sigma \) is a positive definite matrix and \( c = \frac{\Gamma\left(\frac{p}{2}\right)}{2^\frac{p}{2} \Gamma(p)} \). Using this distribution as the basis we have developed statistical methods for performing various statistical inferences for multivariate data. Our main contributions in this dissertation are the following:

(i) Various characteristics of this distribution, such as, its moments, the marginal, and conditional distributions in specific forms, and a simulation algorithm for simulating samples from this distribution are provided.

(ii) Estimation of the parameters of this distribution using the maximum likelihood method under different assumptions of one and more populations, and under different covariance is performed. An interesting and important observation is
that the maximum likelihood estimators derived under this distribution are the

(iii) Using the asymptotic distribution of the estimates, multivariate analysis of vari­
ance (MANOVA) is performed and simultaneous confidence intervals for contrasts
are constructed and illustrated on data sets.

(iv) Finally, discrimination and classification rules under a Kotz type distribution are
derived and compared with the rules based on a multivariate normal distribution
using estimated expected error of misclassification. It is concluded that the ex­
pected error of misclassification can be reduced by using the methods developed
here when the underlying distributions are not multivariate normal, but are of
Kotz type distributions.
Dedicated

to

my parents, my husband, and my son

whose support and encouragement led me to pursue my dreams.
ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Dr. Dayanand N. Naik, for always giving his valuable time to help me with my dissertation. Without his constant encouragement and guidance, I could not have completed this dissertation.

I am also thankful for Dr. N. Rao Chaganty and Dr. John M. Dorrepaal, the Chairman of the Department of Mathematics and Statistics, for providing me financial support during the last two years of my graduate study. A deep sense of appreciation is extended to my dissertation committee, Dr. N. Rao Chaganty, Dr. Larry D. Lee and Dr. Edward Markowski for their generous suggestions and comments. Special thanks is extended to Dr. Ram C. Dahiya for his comments on an earlier draft of the thesis.

Many thanks are due to my friends, Nongnooch Deetae, Orasa Suriyaphan, and Deepak Mav for their friendship and support.

Finally, my heartfelt thanks to my father, who passed away recently and always wanted to see me in a Ph.D. gown. I am sure he is very proud of his daughter now. In addition, I am proud to acknowledge the grateful encouragement and support I received from my mother and my sister. Above all, I want to thank my husband and my son for their constant patience, encouragement and contribution. Without their support I would not have been able to achieve my dream of obtaining a Ph.D. degree.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Tests for Multivariate Normality</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Power Studies</td>
<td>6</td>
</tr>
<tr>
<td>1.1.2</td>
<td>General Comments and Concluding Remarks</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>Overview of Thesis</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>A Kotz Type Distribution and Estimation of Parameters</td>
<td>19</td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>19</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Definition</td>
<td>19</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Moments and Other Properties</td>
<td>21</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Marginal and Conditional Distributions</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulation Algorithm</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>Goodness-of-fit Tests for Kotz Type Distribution</td>
<td>26</td>
</tr>
<tr>
<td>2.4</td>
<td>Estimation of Parameters</td>
<td>27</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Generalized Spatial Median (GSM)</td>
<td>27</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Computation of GSM and $\hat{\Sigma}$</td>
<td>28</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Computation of GSM and $\hat{\Sigma}$ under Different Covariance Structures</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>Example</td>
<td>34</td>
</tr>
<tr>
<td>2.6</td>
<td>Comparisons of ML Estimates Between Kotz Type and Normal Populations</td>
<td>39</td>
</tr>
<tr>
<td>2.7</td>
<td>Concluding Remarks</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Statistical Inference Using Kotz Type Distribution</td>
<td>46</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>Testing of Hypothesis for One Population Under the General Covariance Structure</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>Testing of Hypothesis for One Population Under AR(1) Structure</td>
<td>48</td>
</tr>
</tbody>
</table>
3.4 Testing of Hypothesis for Several Populations Under the General Covariance Structure ...................................................... 50
3.5 Testing of Hypothesis for Several Populations Under AR(1) Structure . 54
3.6 Simultaneous Confidence Intervals for Linear Functions of a Single Mean Under the General Covariance Structure ..................... 57
3.7 Simultaneous Confidence Intervals Under AR(1) Structure .......... 60
3.8 Simultaneous Confidence Intervals for Several Population Means . . . 61
3.9 Simultaneous Confidence Intervals for Several Population Means Under AR(1) structure ..................................................... 63
3.10 Concluding Remarks .................................................................................................................. 65

4 Discrimination And Classification Analysis Using Kotz Type Distribution 66
4.1 Introduction ............................................................................................................................... 66
4.2 Separation and Classification for Two Populations ................................................................. 66
  4.2.1 Classification Rules Under Kotz Type Distributions and General Covariance Structure .......... 68
  4.2.2 Under AR(1) Structure ................................................................. 72
  4.2.3 Evaluating Classification Functions for Two Populations .......... 76
4.3 Separation and Classification for Several Populations ........................ 79
  4.3.1 Classification Rules Under Kotz Type Distributions and General Covariance Structure .......... 80
  4.3.2 Under AR(1) Structure ................................................................. 85
  4.3.3 Evaluating Classification Functions for Several Populations .... 87
4.4 Concluding Remarks .................................................................................................................. 90

References ................................................................................................................. 91

Appendix ...................................................................................................................... 97

Vita .............................................................................................................................. 118

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The approximate and empirical percentage points of $K_1$ and $K_2$ of size $1 - \alpha$ and sample sizes $n$ for $p = 3$.</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>The approximate and empirical percentage points of $T_{n,\beta}$ of size $1 - \alpha$, sample sizes $n$ and parameters $\beta$ for $p = 3$.</td>
<td>7</td>
</tr>
<tr>
<td>1.3</td>
<td>The approximate and empirical percentage points of $K_1$ and $K_2$ of size $1 - \alpha$ and sample sizes $n$ for $p = 5$.</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>The approximate and empirical percentage points of $T_{n,\beta}$ of size $1 - \alpha$, sample sizes $n$ and parameters $\beta$ for $p = 5$.</td>
<td>8</td>
</tr>
<tr>
<td>1.5</td>
<td>Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.01$ against multivariate normal, Cauchy, $t$, $\chi^2$ and gamma distributions.</td>
<td>11</td>
</tr>
<tr>
<td>1.6</td>
<td>Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.01$ against $Exp(1)$, contaminated normal and $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.</td>
<td>12</td>
</tr>
<tr>
<td>1.7</td>
<td>Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.05$ against multivariate normal, Cauchy, $t$, $\chi^2$ and gamma distributions.</td>
<td>13</td>
</tr>
<tr>
<td>1.8</td>
<td>Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.05$ against $Exp(1)$, contaminated normal and $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.</td>
<td>14</td>
</tr>
<tr>
<td>1.9</td>
<td>Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.10$ against multivariate normal, Cauchy, $t$, $\chi^2$ and gamma distributions.</td>
<td>15</td>
</tr>
</tbody>
</table>
Table

1.10 Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,3}$. Tests of sizes $\alpha = 0.10$ against $Exp(1)$, contaminated normal and $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions. ............................................. 16

2.1 Board data: measurements on stiffness. .......................................................... 35

2.2 The estimates of bias and root MSE for $\hat{\mu}$ and $\bar{x}$ and Pitman nearness probability of $\hat{\mu}$ Relative to $\bar{x}$ for Kotz type samples. .................. 43

2.3 The estimates of bias and root MSE for $\hat{\Sigma}$ and $\frac{1}{p+1}S_n$ and Pitman nearness probability of $\hat{\Sigma}$ Relative to $\frac{1}{p+1}S_n$ for Kotz type samples. ..... 44

2.4 The estimates of bias and root MSE for $\hat{\mu}$ and $\bar{x}$ and Pitman nearness probability of $\bar{x}$ Relative to $\hat{\mu}$ for normal samples. .................... 44

2.5 The estimates of bias and root MSE for $(p + 1)\hat{\Sigma}$ and $S_n$ and Pitman nearness probability of $S_n$ Relative to $(p + 1)\hat{\Sigma}$ for normal samples. 45

3.1 Helmet data for football players (HS: high school, C: college and NF: nonfootball) ................................................................. 53

4.1 Iris data: Iris Versicolor and Iris Virginica. ..................................................... 71

4.2 Data on Iris Setosa. ................................................................. 83
CHAPTER I

INTRODUCTION

Before doing any statistical modelling, it is crucial to verify if the data satisfy the underlying distributional assumptions. Multivariate normal distribution plays an important role in analysis of multivariate data. This is especially true in the analysis of mean vectors and multivariate regression where the multivariate analysis of variance (MANOVA) is used. Most of the techniques of multivariate statistical analysis are based on the assumption that the data are generated from a multivariate normal distribution. Further, the normal distribution is often a useful approximation to the true population distribution. Hence it is important for a practitioner to be able to determine whether the data that are being used for statistical inference are from multivariate normal distributions, if not exactly at least approximately.

The aim of this chapter is two-fold. First, if possible, to provide a practitioner, a single most powerful and easily usable test, or at least a small group of tests, for testing multivariate normality and secondly, to provide a brief overview of the chapters in the thesis. To achieve the first aim, we have reviewed state of the art literature on the topic of tests for multivariate normality and conducted our own extensive simulation studies. Based on published work and our studies we are able to provide a set of three tests: two tests based on already well known Mardia's measures of skewness and kurtosis and a third one which is recently proposed Henze-Zirkler test. However, we conclude that all the three tests are needed to definitely conclude that the multivariate normality holds. An overview of the chapters in the thesis is provided at the end of this chapter.

1.1 Tests for Multivariate Normality

In a series of recent articles Mecklin and Mundform (2002, 2003a, 2003b) and Henze (2002) reviewed and summarized the procedures and tests for multivariate normality, 

\[ \text{The model journal used for this dissertation is Statistica Sinica.} \]
developed by researchers over the years, into the following 5 categories.


(IV) Consistent procedures based on the empirical characteristic function (due to Csörgő (1986), Baringhaus and Henze (1988), Henze and Zirkler (1990), Ghosh and Ruymgaart (1992), Bowman and Foster (1993) and Naito (1996)).

(V) Miscellaneous procedures (due to Loh (1986), Hasofer and Stein (1990), Kuwana and Kariya (1991), Smith and Richardson (1993), Zhu et al. (1995), Zhu et al. (1997), Kariya et al. (1999), Slate (1999), and Liang et al. (2000)).

Based on analytical and several empirical and simulation studies, Henze and Zirkler (1990), Henze (2002) and Mecklin and Mundform (2002) identified only 13 of the following tests as useful tests. These are, tests based on Mardia’s multivariate skewness and kurtosis measures, based on the Mardia-Foster statistic and the Mardia-Kent statistic, Royston’s multivariate Shapiro-Wilk test, the Romeu-Ozturk test, the Mudholkar-Srivastava-Lin extension of Shapiro-Wilk test, the Henze-Zirkler empirical characteristic function test using asymptotic critical values, Hawkins’ extension of the
Anderson-Darling test, Koziol's extension of the Cramer-von Mises test, the Paulson-Roohan-Sullo version of the Anderson-Darling test, Singh's test of the correlation of the beta plot with classical estimates of mean and variance, and Singh's test of the correlation of the beta plot with robust M-estimates of mean and variance using the empirical critical values. Based on further simulation studies Mecklin and Mundform (2002, 2003a, 2003b) concluded that Henze-Zirkler test is the best. However, Henze and Zirkler (1990) and Henze (2002) concluded that Henze-Zirkler test is the best, but the descriptive use of multivariate measures of skewness and kurtosis derived by Mardia (1970) is recommended as omnibus invariant tests for multivariate normality (Henze, 2002). In the following, we provide some details about these three procedures and conduct a detailed simulation experiment to compare them.

However, many times multivariate normality of a set of multivariate data can be assessed using $Q-Q$ plots, and these plots along with the univariate tests for testing normality of the marginals should be routinely adopted. Univariate tests are discussed in details in many books and review articles. For example, see Shapiro, Wilk and Chen (1968), Lin and Mudholkar (1980), and D'Agostino (1986). A Q-Q plot can be drawn using the following procedure. See Khattree and Naik (1999) for details and computer programs in SAS.

Let $x_j$, $j = 1, ..., n$ be a random sample from a multivariate normal distribution $N_p(\mu, \Sigma)$. Then

$$z_j = \Sigma^{-1/2}(x_j - \mu), \quad j = 1, ..., n$$

are i.i.d. $N_p(0, I)$ and hence each

$$\delta_j^2 = z_j'z_j$$

$$= (x_j - \mu)'\Sigma^{-1}(x_j - \mu), \quad j = 1, ..., n$$

independently follows a chi-square distribution with $p$ degrees of freedom. The quantity $\delta_j^2$ is the squared Mahalanobis distance between $x_j$ and its expectation $\mu$. If the observations $x_j's$ are from a $N_p(\mu, \Sigma)$ then the distances (the sample versions of squared Mahalanobis distances)

$$d_j^2 = (x_j - \bar{x})'S^{-1}(x_j - \bar{x}), \quad j = 1, ..., n$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
will approximately be distributed as a chi-square on \( p \) degrees of freedom. Hence the ordered \( d_j^2 \) values are plotted against quantiles of \( p \) degrees of freedom of chi-square distribution.

When the plot approximately looks like a straight line passing through the origin at a 45° angle with the horizontal axis, it can be assumed that the observations come from a multivariate normal population.

Mardia (1970) proposed tests for multivariate normality based on generalizations of the univariate skewness and kurtosis measures respectively given by

\[
\beta_{1p} = E[(x - \mu)'\Sigma^{-1}(y - \mu)]^3, \\
\beta_{2p} = E[(x - \mu)'\Sigma^{-1}(x - \mu)]^2,
\]

where \( x \) and \( y \) are identically and independently distributed \( p \) variate random vectors with mean \( \mu \) and variance covariance (dispersion matrix) \( \Sigma \). For the multivariate normal distribution, \( \beta_{1p} = 0 \) and \( \beta_{2p} = p(p + 2) \).

For a sample \( x_1, ..., x_n \) of size \( n \), the estimates of \( \beta_{1p} \) and \( \beta_{2p} \) can be obtained as

\[
\hat{\beta}_{1p} = b_{1p} = \frac{1}{n^2} \sum_{j,k=1}^{n} (r_j' r_k)^3 = \frac{1}{n^2} \sum_{j,k=1}^{n} g_{jk}^3, \\
\hat{\beta}_{2p} = b_{2p} = \frac{1}{n} \sum_{j=1}^{n} g_{jj}^2 = \frac{1}{n} \sum_{j=1}^{n} d_j^4,
\]

where \( r_j = S^{\frac{1}{2}}_n(x_j - \bar{x}) \), \( g_{jk} = (x_j - \bar{x})'S_n^{-1}(x_k - \bar{x}) \), \( d_j = \sqrt{g_{jj}} \), \( j, k = 1, ..., n \), \( \bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j \) and \( S_n = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' \).

Then Mardia (1970)'s tests for multivariate normality are based on \( K_1 \) and \( K_2 \), where for large samples under the null hypothesis of multivariate normality,

\[
K_1 = \frac{nb_{1p}}{6} \sim \frac{\chi^2_{p(p+1)(p+2)}}{6} \quad \text{and} \\
K_2 = \frac{b_{2p} - p(p + 2)}{\left[\frac{8p(p+2)}{n}\right]^{\frac{3}{2}}} \sim N(0, 1).
\]

Henze and Zirkler (1990) generalized Baringhause-Henze procedures (Baringhaus and Henze (1988)) that extended Epps and Pulley test (Epps and Pulley, 1983) for univariate normality based on empirical characteristic function to the multivariate case. The Henze-Zirkler test statistic is based on a non-negative functional that measures
the distance between two distribution functions, the hypothesized function (which is the multivariate normal) and the observed function. In order to have the statistic be consistent, the functional, say $D_{n, \beta}$, must equal to zero if and only if the observed data is from the multivariate normal distribution. The non-negative functional considered is

$$D_{n, \beta} = \int_{\mathbb{R}^p} \left| \Psi_n(t) - \exp\left(-\frac{1}{2}||t||^2\right) \right|^2 \varphi_\beta(t) \, dt,$$

where $\beta$ denotes a positive number,

$$\Psi_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(it' \mathbf{r}_j) \quad (t \in \mathbb{R}^p)$$

is the empirical characteristic function of the scaled residuals $\mathbf{r}_j$ defined by $S_n^{-\frac{1}{2}}(x_j - \bar{x})$ and

$$\varphi_\beta(t) = \left(2\pi\beta^2\right)^{-\frac{p}{2}} \exp\left(-\frac{||t||^2}{2\beta^2}\right), \quad t \in \mathbb{R}^p,$$

stands for the weighted function to be the multivariate normal distribution $N_p(0, \beta^2 \mathbf{I}_p)$.

Henze and Zirkler (1990) suggested the smoothing parameter $\beta$ to be

$$\beta = \beta_p(n) = \frac{1}{\sqrt{2}} \left(\frac{2p + 1}{4}\right)^{\frac{1}{p+4}} n^{\frac{1}{p+4}},$$

so that it changes slowly with $n$.

Then Henze-Zirkler test statistic for testing multivariate normality is given by

$$T_{n, \beta} = nD_{n, \beta},$$

where

$$D_{n, \beta} = \frac{1}{n^2} \sum_{j,k=1}^{n} \exp\left(-\frac{\beta^2}{2} ||\mathbf{r}_j - \mathbf{r}_k||^2\right)$$

$$- 2(1 + \beta^2)^{-\frac{p}{2}} \frac{1}{n} \sum_{j=1}^{n} \exp\left(-\frac{\beta^2}{2(1 + \beta^2)} ||\mathbf{r}_j||^2\right) + (1 + 2\beta^2)^{-\frac{p}{2}},$$

with $||\mathbf{r}_j - \mathbf{r}_k||^2 = (x_j - x_k)'S_n^{-1}(x_j - x_k) = g_{jj} - 2g_{jk} + g_{kk}$ and $||\mathbf{r}_j||^2 = g_{jj}$.

For $p \geq 1$, and $\beta > 0$, Henze and Zirkler (1990) provide mean and variance of the asymptotic distribution, $T_\beta(p)$, of $T_{n, \beta}$ as

$$E[T_\beta(p)] = 1 - (1 + 2\beta^2)^{-\frac{p}{2}} \left[1 + \frac{p\beta^2}{1 + 2\beta^2} + \frac{p(p + 2)\beta^4}{2(1 + 2\beta^2)^2}\right]$$

and

$$Var[T_\beta(p)] = 2(1 + 4\beta^2)^{-\frac{p}{2}} + 2(1 + 2\beta^2)^{-p} \left[1 + \frac{2p\beta^4}{(1 + 2\beta^2)^2} + \frac{3p(p + 2)\beta^8}{4(1 + 2\beta^2)^4}\right].$$
\[-4 \omega(\beta)^{-\frac{2\beta}{2\omega(\beta)}} \left[ 1 + \frac{3p\beta^4}{2\omega(\beta)} + \frac{p(p+2)\beta^8}{2\omega(\beta)^2} \right],\]

where \(\omega(\beta) = (1 + \beta^2)(1 + 3\beta^2)\).

Further, Henze and Zirkler (1990) approximated the distribution of \(T_\beta(p)\) by the lognormal distribution with the parameters \(\mu_{\beta,p} = E[T_\beta(p)]\) and \(\sigma_{\beta,p}^2 = Var[T_\beta(p)]\). That is, the \((1 - \alpha)\) quantile of the asymptotic distribution of \(T_{n,\beta}\), that is, of \(T_\beta(p)\), is given by

\[q_{\beta,p}(1 - \alpha) = \mu_{\beta,p} \left(1 + \frac{\sigma_{\beta,p}^2}{\mu_{\beta,p}^2}\right)^{-\frac{1}{2}} \exp\left[\Phi^{-1}(1 - \alpha) \sqrt{\log\left(1 + \frac{\sigma_{\beta,p}^2}{\mu_{\beta,p}^2}\right)}\right],\]

where \(\Phi^{-1}(\cdot)\) is the inverse of standard normal distribution function.

1.1.1 Power Studies

As we have noted, large number of tests for evaluating univariate and multivariate normality are available in the literature. However, there is no one test that is optimal for all possible deviations from normality. The usual process adopted to detect the sensitivity of these tests is to perform power studies where the tests are applied to a wide range of non-normal populations for a variety of sample sizes, levels of significance etc. Hence for our simulation study, we first determine the approximate upper \(\alpha\)% cutoff points of the null distribution of Mardia's tests based on \(K_1\) and \(K_2\) and Henze and Zirkler test, \(T_{n,\beta}\) for \(\beta = 0.5, 1, 3\) and \(\beta\) depending on \(n\) and \(p\), using Monte Carlo simulation method based on 50,000 simulations on a variety of sample sizes, \(n = 20, 30, 50, 100\), the dimension \(p = 3, 5\), and the significance levels \(\alpha = 0.01, 0.05, 0.1\).

The results are given in Tables 1.1 - 1.4. In each table, the last row (except for the last three columns in which \(\beta\) depends on \(n\)) is the \(\chi^2\) with \(\frac{p(p+1)(p+2)}{6}\) degrees of freedom, \(N(0, 1)\) and lognormal approximation to the empirical quantiles of tests based on \(K_1\), \(K_2\) (Table 1.1, 1.3) and \(T_{n,\beta}\) (Table 1.2, 1.4) respectively. It shows that our approximate cutoff points are very close to the empirical quantiles.

Next, to compare the power of the Mardia's tests with Henze-Zirkler test, we perform the following simulation study. Five thousand samples from a variety of \(p\)-dimensional continuous distributions are used under the same above combinations of
Table 1.1: The approximate and empirical percentage points of $K_1$ and $K_2$ of size $1 - \alpha$ and sample sizes $n$ for $p = 3$.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>13.196</td>
<td>15.428</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>14.222</td>
<td>16.659</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>14.988</td>
<td>17.518</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>15.660</td>
<td>18.234</td>
</tr>
<tr>
<td>$q_{\beta,1}(1 - \alpha)$</td>
<td>15.987</td>
<td>18.307</td>
</tr>
</tbody>
</table>

Table 1.2: The approximate and empirical percentage points of $T_{n,\beta}$ of size $1 - \alpha$, sample sizes $n$ and parameters $\beta$ for $p = 3$.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 1$</th>
<th>$\beta = 3$</th>
<th>$\beta = \beta_3(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
<td>0.99</td>
<td>0.90</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>0.097</td>
<td>0.112</td>
<td>0.145</td>
<td>0.613</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>0.100</td>
<td>0.114</td>
<td>0.150</td>
<td>0.614</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.102</td>
<td>0.118</td>
<td>0.151</td>
<td>0.616</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.104</td>
<td>0.120</td>
<td>0.155</td>
<td>0.619</td>
</tr>
<tr>
<td>$q_{\beta,3}(1 - \alpha)$</td>
<td>0.106</td>
<td>0.121</td>
<td>0.156</td>
<td>0.620</td>
</tr>
</tbody>
</table>

$n$, $p$ and $\alpha$. We use the above obtained approximate cutoff points as an approximate critical value for a level $\alpha$ test based on $K_1$ and $K_2$ and $T_{n,\beta}$. The number of rejections are counted. We have implemented all the calculations using IML procedure of SAS. The SAS programs are provided in Program 1 in APPENDIX.

However, in Tables 1.5 - 1.10, power estimates of the six tests are presented. Interpretations of the entries in each of these tables are provided in the following categories.

**Group 1:** Multivariate normal distribution.

In this case, the null hypothesis is true, so each test should reject at about the nominal probability $\alpha$. We notice that four Henze-Kirkler tests based on different parameters $\beta$ have a rejection rate below or close to the $\alpha$ level but the Mardia’s skewness and kurtosis measures based on $K_1$ and $K_2$ tests have a rejection rate slightly above the $\alpha$ level for $p = 3$ and 5.
Table 1.3: The approximate and empirical percentage points of $K_1$ and $K_2$ of size $1 - \alpha$ and sample sizes $n$ for $p = 5$.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 20$</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>38.467</td>
<td>42.029</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>0.176</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>41.450</td>
<td>45.569</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.468</td>
<td>0.833</td>
</tr>
<tr>
<td></td>
<td>43.576</td>
<td>47.733</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.769</td>
<td>1.168</td>
</tr>
<tr>
<td></td>
<td>45.139</td>
<td>49.230</td>
</tr>
<tr>
<td>$p_{2,p}(1 - \alpha)$</td>
<td>46.059</td>
<td>49.802</td>
</tr>
<tr>
<td></td>
<td>1.282</td>
<td>1.645</td>
</tr>
</tbody>
</table>

Table 1.4: The approximate and empirical percentage points of $T_{n,\beta}$ of size $1 - \alpha$, sample sizes $n$ and parameters $\beta$ for $p = 5$.

<table>
<thead>
<tr>
<th>$1 - \alpha$</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 1$</th>
<th>$\beta = 3$</th>
<th>$\beta = \beta_0(n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.313</td>
<td>0.843</td>
<td>0.905</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.008</td>
<td>1.018</td>
<td>1.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.880</td>
<td>0.909</td>
<td>0.971</td>
<td></td>
</tr>
<tr>
<td>$n = 20$</td>
<td>0.195</td>
<td>0.211</td>
<td>0.249</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.814</td>
<td>0.847</td>
<td>0.919</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.014</td>
<td>1.021</td>
<td>1.037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.932</td>
<td>0.963</td>
<td>1.026</td>
<td></td>
</tr>
<tr>
<td>$n = 30$</td>
<td>0.197</td>
<td>0.214</td>
<td>0.249</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.814</td>
<td>0.848</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.015</td>
<td>1.021</td>
<td>1.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.961</td>
<td>0.991</td>
<td>1.053</td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.201</td>
<td>0.217</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.814</td>
<td>0.849</td>
<td>0.920</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.015</td>
<td>1.021</td>
<td>1.031</td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.201</td>
<td>0.217</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.814</td>
<td>0.849</td>
<td>0.920</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.015</td>
<td>1.021</td>
<td>1.031</td>
<td></td>
</tr>
<tr>
<td>$p_{2,p}(1 - \alpha)$</td>
<td>0.814</td>
<td>0.849</td>
<td>0.920</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.015</td>
<td>1.021</td>
<td>1.031</td>
<td></td>
</tr>
</tbody>
</table>

Group 2: $p$ variate $t$ distribution.

In group 2 we generate samples from multivariate $t$ distribution (as described in Johnson and Kotz, 1972) with 1, 2 and 5 degrees of freedom for $p = 3$ and 5.

The $t$ distributions are symmetric and have very high kurtosis compared to normal. However, the $t$ distributions are closer to normal when degree of freedom increases, especially, $t$ with degree of freedom $\geq 30$.

The Mardia’s kurtosis test based on $K_2$ performs very well compared to the other tests and the next best test is $T_{0.5}$. The Mardia’s skewness test based on $K_1$ performs better for large $n$.

Group 3: $p$ variate chi $- square$ distribution.

We generate samples from multivariate chi $- square$ distribution with 5 and 10 degrees of freedom for $p = 3$ and 5. Note that chi $- square$ distribution is skewed.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
to the right. We find that $T_{0.5}$ test is generally superior to $K_1$ test and most of the Henze-Zirkler tests except $T_3$ perform better than the $K_2$ test.

**Group 4: $p$ variate Gamma distribution.**

We generate samples from multivariate gamma distribution (as described in Ronning (1977)) with the scale parameter fixed at 1. These gamma distributions are skewed to the right and have high kurtosis. However, the gamma distributions, for fixed scale parameter, become closer to normal as shape parameter increases and hence it is expected that the power will decrease.

It shows that the test based on $T_{0.5}$ performs very well compared to the other tests. $K_1$ test is comparative to $T_{0.5}$ test when sample size increases. $K_2$ and $T_3$ tests perform poorly in this group.

**Group 5: $p$ variate Exponential distribution, $Exp(1)$.**

In this group we generate samples from multivariate Exponential distribution with the scale parameter = 1. The performance of the tests based on $T_{0.5}$ and $T_1$ are better than other tests and $K_1$ test is comparable to $T_{0.5}$ and $T_1$ when $n$ increases. $K_2$ test does not perform well for small $n$.

**Group 6: $p$ variate Contaminated normal and $Exp(1)$ distribution, $\frac{1}{2}N(0,1) + \frac{1}{2}Exp(1)$.**

In this group, $K_2$ test is superior to others but $T_{0.5}$ and $K_1$ tests perform better when $n$ is large.

**Group 7: $p$ variate Contaminated normal and $t(5)$ distribution, $\frac{1}{2}N(0,1) + \frac{1}{2}t(5)$.**

In this group the $K_2$ test performs well compared to other tests and more sensitive for detecting normality with $n$ for most cases. $K_1$ and $T_{0.5}$ tests perform better with large $n$ and $p$.

**Group 8: $p$ variate Contaminated normal and $\chi^2(5)$ distribution, $\frac{1}{2}N(0,1) + \frac{1}{2}\chi^2(5)$.**

It appears that the Henze-Zirkler tests except $T_3$ perform well compared to $K_1$ test. $K_1$ test is comparable to Henze-Zirkler tests when $n = 100$.

**Group 9: $p$ variate Kotz type distribution.**

We generate samples from Kotz type distribution (as described later in Chapter 2). Kotz type distribution is symmetric and heavy-tailed. Here the $K_2$ test performs well when compared to other tests and $T_{0.5}$ test performs better when $p$ increases, but $K_1$
test does not perform well.

1.1.2 General Comments and Concluding Remarks

Rejection rates for Mardia's skewness and kurtosis measures based on $K_1$ and $K_2$ that are above the $\alpha$ level for small $n$ indicate a problem with Type I error rate. The power comparisons among six tests show that the power performance of Henze-Zirkler test varies considerably with the parameter $\beta$. Choice $\beta = 0.5$ gives a very strong test against skewed distributions. $T_3$ appears inferior among Henze-Zirkler tests. Perhaps a smaller value of $\beta$ would do better, in general. Thus $T_3$ is not recommended.

The Mardia's skewness test based on $K_1$ test and $T_{0.5}$ test have much the same sensitivity with $K_1$ is slightly inferior for some cases. The power performance of $K_1$ test depends on the tail behavior of $X$, that is, $K_1$ test is consistent against all distribution satisfying $\beta_1 > 0$ (Henze and Zirkler, 1990).

The test based on $K_2$ is consistent against all alternative distributions satisfying $\beta_2 \neq p(p+2)$. Our simulation results show that $K_2$ test is generally inferior to $T_{0.5}$ test if the distribution is light-tailed, but performs better against heavy-tailed distributions, like Kotz type distributions.

In conclusion, it is clear that the Henze-Zirkler test based on $T_{0.5}$ performs reasonably well compared to the other tests for most cases. However, $K_1$ and $K_2$ tests are comparative to $T_{0.5}$ for some cases.

Thus, a practitioner may use a combination of the three tests, namely, Mardia's skewness and kurtosis measures based on $K_1$, $K_2$ and Henze-Zirkler tests and use the strategy that all these tests must declare multivariate normality for claiming that the data really come from multivariate normal distribution.
Table 1.5: Monte Carlo power estimates of $K_1, K_2, T_{0.5}, T_1, T_3$ and $T_{n,\theta}$. Tests of sizes $\alpha = 0.01$ against multivariate normal, Cauchy, t, $\chi^2$ and gamma distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>$n$</th>
<th>$p=3$</th>
<th>$p=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p(0, I)$</td>
<td>20</td>
<td>0.009</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.013</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td>Cauchy</td>
<td>20</td>
<td>0.961</td>
<td>0.979</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.992</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$t(2)$</td>
<td>20</td>
<td>0.715</td>
<td>0.782</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.672</td>
<td>0.697</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.966</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$t(5)$</td>
<td>20</td>
<td>0.231</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.300</td>
<td>0.360</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.550</td>
<td>0.657</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\chi^2(5)$</td>
<td>20</td>
<td>0.481</td>
<td>0.515</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.830</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\chi^2(10)$</td>
<td>20</td>
<td>0.107</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.202</td>
<td>0.259</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.400</td>
<td>0.575</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$G(2, 1)$</td>
<td>20</td>
<td>0.311</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.593</td>
<td>0.708</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.930</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$G(5, 1)$</td>
<td>20</td>
<td>0.095</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.207</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.482</td>
<td>0.575</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.920</td>
<td>0.951</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 1.6: Monte Carlo power estimates of $K_1, K_2, T_{0.5}, T_1, T_3$ and $T_{n,3}$. Tests of sizes $\alpha = 0.01$ against $Exp(1)$, contaminated normal and $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>$n$</th>
<th>$p=3$</th>
<th>$p=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.631</td>
<td>0.421</td>
</tr>
<tr>
<td>$Exp(1)$</td>
<td>30</td>
<td>0.804</td>
<td>0.609</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.998</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>0.990</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.169</td>
<td>0.200</td>
</tr>
<tr>
<td>$\frac{1}{2}N(0,1)$</td>
<td>30</td>
<td>0.283</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.433</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.689</td>
<td>0.846</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.105</td>
<td>0.104</td>
</tr>
<tr>
<td>$\frac{1}{2}N(0,1)$</td>
<td>30</td>
<td>0.165</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.277</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.423</td>
<td>0.602</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.299</td>
<td>0.168</td>
</tr>
<tr>
<td>$\frac{1}{2}N(0,1)$</td>
<td>30</td>
<td>0.655</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.929</td>
<td>0.441</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>0.728</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.164</td>
<td>0.210</td>
</tr>
<tr>
<td>Kotz Type</td>
<td>30</td>
<td>0.229</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.359</td>
<td>0.594</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.461</td>
<td>0.879</td>
</tr>
</tbody>
</table>
Table 1.7: Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\delta}$. Tests of sizes $\alpha = 0.05$ against multivariate normal, Cauchy, t, $\chi^2$ and gamma distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>n</th>
<th>p=3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
<td>$T_{0.5}$</td>
<td>$T_1$</td>
<td>$T_3$</td>
<td>$T_{n,\delta}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.046</td>
<td>0.053</td>
<td>0.048</td>
<td>0.047</td>
<td>0.045</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.055</td>
<td>0.056</td>
<td>0.050</td>
<td>0.050</td>
<td>0.051</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.052</td>
<td>0.054</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.050</td>
<td>0.052</td>
<td>0.046</td>
<td>0.043</td>
<td>0.048</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.050</td>
<td>0.054</td>
<td>0.050</td>
<td>0.050</td>
<td>0.045</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.053</td>
<td>0.055</td>
<td>0.052</td>
<td>0.051</td>
<td>0.053</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.053</td>
<td>0.055</td>
<td>0.052</td>
<td>0.051</td>
<td>0.053</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052</td>
<td>0.054</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.051</td>
<td>0.050</td>
<td>0.048</td>
<td>0.043</td>
<td>0.046</td>
<td>0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alternative</th>
<th>n</th>
<th>p=5</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
<td>$T_{0.5}$</td>
<td>$T_1$</td>
<td>$T_3$</td>
<td>$T_{n,\delta}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.046</td>
<td>0.053</td>
<td>0.048</td>
<td>0.047</td>
<td>0.045</td>
<td>0.048</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.055</td>
<td>0.056</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.052</td>
<td>0.054</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.050</td>
<td>0.052</td>
<td>0.046</td>
<td>0.043</td>
<td>0.048</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.050</td>
<td>0.054</td>
<td>0.050</td>
<td>0.050</td>
<td>0.045</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.053</td>
<td>0.055</td>
<td>0.052</td>
<td>0.051</td>
<td>0.053</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.053</td>
<td>0.055</td>
<td>0.052</td>
<td>0.051</td>
<td>0.053</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.052</td>
<td>0.054</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.051</td>
<td>0.050</td>
<td>0.048</td>
<td>0.043</td>
<td>0.046</td>
<td>0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.044</td>
<td>0.048</td>
<td>0.045</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 1.8: Monte Carlo power estimates of $K_1, K_2, T_{0.5}, T_1, T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.05$ against $Exp(1)$, contaminated normal and $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>$n$</th>
<th>$p=3$</th>
<th>$p=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Exp(1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.818</td>
<td>0.595</td>
<td>0.638</td>
</tr>
<tr>
<td>30</td>
<td>0.974</td>
<td>0.783</td>
<td>0.992</td>
</tr>
<tr>
<td>50</td>
<td>1.000</td>
<td>0.936</td>
<td>1.000</td>
</tr>
<tr>
<td>100</td>
<td>1.000</td>
<td>0.997</td>
<td>1.000</td>
</tr>
<tr>
<td>$p$-variate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.332</td>
<td>0.383</td>
<td>0.318</td>
</tr>
<tr>
<td>30</td>
<td>0.444</td>
<td>0.533</td>
<td>0.362</td>
</tr>
<tr>
<td>50</td>
<td>0.617</td>
<td>0.754</td>
<td>0.627</td>
</tr>
<tr>
<td>100</td>
<td>0.813</td>
<td>0.942</td>
<td>0.857</td>
</tr>
<tr>
<td>$\frac{1}{2}N(0,1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.215</td>
<td>0.234</td>
<td>0.195</td>
</tr>
<tr>
<td>30</td>
<td>0.305</td>
<td>0.347</td>
<td>0.273</td>
</tr>
<tr>
<td>50</td>
<td>0.408</td>
<td>0.508</td>
<td>0.376</td>
</tr>
<tr>
<td>100</td>
<td>0.561</td>
<td>0.752</td>
<td>0.549</td>
</tr>
<tr>
<td>$\chi^2(5)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.564</td>
<td>0.334</td>
<td>0.085</td>
</tr>
<tr>
<td>30</td>
<td>0.705</td>
<td>0.447</td>
<td>0.220</td>
</tr>
<tr>
<td>50</td>
<td>0.983</td>
<td>0.625</td>
<td>0.098</td>
</tr>
<tr>
<td>100</td>
<td>1.000</td>
<td>0.860</td>
<td>1.000</td>
</tr>
<tr>
<td>$\chi^2(5)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.334</td>
<td>0.417</td>
<td>0.300</td>
</tr>
<tr>
<td>30</td>
<td>0.425</td>
<td>0.580</td>
<td>0.416</td>
</tr>
<tr>
<td>50</td>
<td>0.542</td>
<td>0.705</td>
<td>0.582</td>
</tr>
<tr>
<td>100</td>
<td>0.662</td>
<td>0.905</td>
<td>0.819</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 1.9: Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.10$ against multivariate normal, Cauchy, t, $\chi^2$ and gamma distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>$n$</th>
<th>$p=3$</th>
<th>$p=5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
</tr>
<tr>
<td>$N_p(0, I)$</td>
<td>20</td>
<td>0.100</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.102</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.095</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.099</td>
<td>0.103</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.986</td>
<td>0.995</td>
</tr>
<tr>
<td>Cauchy</td>
<td>30</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.869</td>
<td>0.925</td>
</tr>
<tr>
<td>t(2)</td>
<td>30</td>
<td>0.956</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.566</td>
<td>0.408</td>
</tr>
<tr>
<td>t(5)</td>
<td>30</td>
<td>0.795</td>
<td>0.552</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.973</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.997</td>
<td>0.942</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.372</td>
<td>0.281</td>
</tr>
<tr>
<td>$\chi^2(5)$</td>
<td>30</td>
<td>0.542</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.812</td>
<td>0.472</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.992</td>
<td>0.679</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.668</td>
<td>0.476</td>
</tr>
<tr>
<td>G(2, 1)</td>
<td>30</td>
<td>0.864</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.992</td>
<td>0.801</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000</td>
<td>0.959</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.362</td>
<td>0.272</td>
</tr>
<tr>
<td>G(5, 1)</td>
<td>30</td>
<td>0.546</td>
<td>0.351</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.811</td>
<td>0.480</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.990</td>
<td>0.674</td>
</tr>
</tbody>
</table>
Table 1.10: Monte Carlo power estimates of $K_1$, $K_2$, $T_{0.5}$, $T_1$, $T_3$ and $T_{n,\beta}$. Tests of sizes $\alpha = 0.10$ against $Exp(1)$, contaminated normal and $Exp(l)$, contaminated normal and $f(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.

<table>
<thead>
<tr>
<th>Alternative</th>
<th>n</th>
<th>$p=3$</th>
<th></th>
<th></th>
<th></th>
<th>$p=5$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Exp(1)$</td>
<td></td>
<td></td>
<td>$K_1$</td>
<td>$K_2$</td>
<td>$T_{0.5}$</td>
<td>$T_1$</td>
<td>$T_3$</td>
<td>$T_{n,\beta}$</td>
</tr>
<tr>
<td>20</td>
<td>0.892</td>
<td>0.996</td>
<td>0.045</td>
<td>0.636</td>
<td>0.725</td>
<td>0.923</td>
<td>0.895</td>
<td>0.757</td>
</tr>
<tr>
<td>30</td>
<td>0.889</td>
<td>0.852</td>
<td>0.907</td>
<td>0.994</td>
<td>0.923</td>
<td>0.990</td>
<td>0.993</td>
<td>0.901</td>
</tr>
<tr>
<td>50</td>
<td>1.000</td>
<td>0.962</td>
<td>1.000</td>
<td>1.000</td>
<td>0.986</td>
<td>1.000</td>
<td>1.000</td>
<td>0.985</td>
</tr>
<tr>
<td>100</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.438</td>
<td>0.517</td>
<td>0.417</td>
<td>0.353</td>
<td>0.236</td>
<td>0.329</td>
<td>0.479</td>
</tr>
<tr>
<td>30</td>
<td>0.558</td>
<td>0.668</td>
<td>0.540</td>
<td>0.476</td>
<td>0.340</td>
<td>0.442</td>
<td>0.636</td>
<td>0.712</td>
</tr>
<tr>
<td>50</td>
<td>0.705</td>
<td>0.833</td>
<td>0.708</td>
<td>0.676</td>
<td>0.515</td>
<td>0.648</td>
<td>0.798</td>
<td>0.882</td>
</tr>
<tr>
<td>100</td>
<td>0.888</td>
<td>0.972</td>
<td>0.912</td>
<td>0.826</td>
<td>0.842</td>
<td>0.920</td>
<td>0.949</td>
<td>0.990</td>
</tr>
<tr>
<td>$\frac{1}{2}N(0,1)$</td>
<td>20</td>
<td>0.329</td>
<td>0.322</td>
<td>0.280</td>
<td>0.203</td>
<td>0.114</td>
<td>0.180</td>
<td>0.320</td>
</tr>
<tr>
<td>30</td>
<td>0.382</td>
<td>0.430</td>
<td>0.359</td>
<td>0.265</td>
<td>0.148</td>
<td>0.224</td>
<td>0.437</td>
<td>0.480</td>
</tr>
<tr>
<td>50</td>
<td>0.505</td>
<td>0.605</td>
<td>0.469</td>
<td>0.347</td>
<td>0.184</td>
<td>0.285</td>
<td>0.587</td>
<td>0.663</td>
</tr>
<tr>
<td>100</td>
<td>0.641</td>
<td>0.626</td>
<td>0.536</td>
<td>0.505</td>
<td>0.242</td>
<td>0.393</td>
<td>0.762</td>
<td>0.804</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.571</td>
<td>0.434</td>
<td>0.803</td>
<td>0.819</td>
<td>0.523</td>
<td>0.783</td>
<td>0.618</td>
</tr>
<tr>
<td>30</td>
<td>0.885</td>
<td>0.566</td>
<td>0.956</td>
<td>0.953</td>
<td>0.761</td>
<td>0.937</td>
<td>0.879</td>
<td>0.596</td>
</tr>
<tr>
<td>50</td>
<td>0.997</td>
<td>0.727</td>
<td>0.999</td>
<td>0.999</td>
<td>0.958</td>
<td>0.998</td>
<td>0.996</td>
<td>0.786</td>
</tr>
<tr>
<td>100</td>
<td>1.000</td>
<td>0.921</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.955</td>
</tr>
<tr>
<td>$p$-variate</td>
<td>20</td>
<td>0.448</td>
<td>0.552</td>
<td>0.415</td>
<td>0.336</td>
<td>0.204</td>
<td>0.303</td>
<td>0.460</td>
</tr>
<tr>
<td>Kotz Type</td>
<td>30</td>
<td>0.543</td>
<td>0.713</td>
<td>0.339</td>
<td>0.458</td>
<td>0.285</td>
<td>0.432</td>
<td>0.580</td>
</tr>
<tr>
<td>50</td>
<td>0.631</td>
<td>0.869</td>
<td>0.666</td>
<td>0.657</td>
<td>0.426</td>
<td>0.616</td>
<td>0.725</td>
<td>0.903</td>
</tr>
<tr>
<td>100</td>
<td>0.751</td>
<td>0.987</td>
<td>0.891</td>
<td>0.919</td>
<td>0.750</td>
<td>0.895</td>
<td>0.856</td>
<td>0.992</td>
</tr>
</tbody>
</table>
1.2 Overview of Thesis

As we noted, most of the inferential statistical methods for multivariate data are developed under the fundamental assumption that the data are from multivariate normal distribution. Unfortunately, as we speculated in this chapter, one can never be sure that a set of data is really from a multivariate normal distribution. As we observed, there are numerous methods for checking (testing) multivariate normality, but based on many published and our own simulation studies we conclude that these tests are generally not very powerful, especially for smaller sample sizes. Hence it is beneficial to have alternative multivariate distributions available along with the methodology for using them.

In this dissertation, we focus on a probability distribution, called Kotz type distribution, which has fatter tail regions than that of multivariate normal distribution and has its probability density function (pdf) in the form

$$f(x, \mu, \Sigma) = c \left| \Sigma \right|^{-\frac{1}{2}} \exp \left\{ -[(x - \mu)'\Sigma^{-1}(x - \mu)]^{\frac{1}{2}} \right\},$$

where \( \mu \in \mathbb{R}^p \), \( \Sigma \) is a positive definite matrix and \( c = \frac{\Gamma\left(\frac{p}{2}\right)}{2^{p/2}\Gamma(p)} \).

Since, for \( p = 1 \) this pdf reduces to that of a double exponential distribution we may consider this as a multivariate generalization of double exponential distribution. However, this is not a multivariate double exponential distribution, because its marginals are not double exponential distributions.

In Chapter 2, adopting the Newton-Raphson method for optimization we derive the maximum likelihood (ML) estimates of location parameter and covariance matrix under a variety of covariance structures, such as, AR(1), equicorrelation, and unstructured covariance structures. The optimization process gives unique ML estimates in the feasible regions under covariance structures. The ML estimate of the location parameter under the assumption of Kotz type distribution is same as the generalized spatial median (GSM) defined by C. R. Rao (1988). Goodness-of-fit tests using multivariate skewness and kurtosis measures for Kotz type distribution are proposed using the results of Baringhaus and Henze (1992) and Henze (1994). The ML estimates of the parameters under Kotz type and multivariate normal distributions are compared.
using different criteria. As expected, the ML estimates under Kotz type distribution do better than normal ML estimates when the data are simulated from Kotz type distribution and vice versa.

In Chapter 3, the likelihood ratio tests are applied for testing the hypothesis of the location parameters. Further, using the asymptotic distribution of these estimates, construction of the simultaneous confidence intervals for linear combinations of the location parameters is shown.

Finally, in Chapter 4, we use Kotz type densities for determining classification rules that minimize the expected cost of misclassification under equal prior probabilities and equal misclassification costs. The classification rule under a common variance covariance matrix is same as the normal based method except that the parameter estimates used are the ML estimates of the parameters under Kotz type distribution. The performance of sample classification functions are evaluated using a nearly unbiased estimate of the expected actual error rate calculated from Lachenbruch’s holdout procedure under the equal prior probabilities and equal misclassification costs. The classification rules under Kotz type distribution perform quite well as compared to those based on multivariate normal distribution.
CHAPTER II

A KOTZ TYPE DISTRIBUTION AND ESTIMATION OF PARAMETERS

2.1 Introduction

In this chapter we introduce a multivariate probability distribution, named Kotz type distribution, which will be focus of our study. In the following subsections we provide the definition and properties of this distribution.

2.1.1 Definition

Probability density function of Kotz type distribution of a $p \times 1$ random vector is given by

$$f(x, \mu, \Sigma) = c |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\}, \quad \mu \in \mathbb{R}^p, \Sigma \text{ p. d.} \quad (2.1)$$

where $c = \frac{\Gamma(\frac{1}{2})}{\sqrt{2\pi} \Gamma(p)}$.

We note that this distribution has fatter tail regions than that of multivariate normal distribution and hence can be an alternative model to the multivariate normal distribution.

The above pdf has appeared at many places in the literature in different forms. For example, the pdf is a special case of the following families of distributions:

(i) Multivariate distributions proposed by Simoni (1968)

Simoni (1968) has considered generalizations of the univariate Subbotin distribution. These have the pdf proportional to

$$\exp \left\{ -\frac{1}{r} [(x - \xi)'A(x - \xi)]^{\frac{r}{2}} \right\},$$

where $A$ is p. d. and $r \geq 1$. For $r = 1$ one obtains our multivariate distribution.
(ii) **Elliptically symmetric distributions (Johnson, 1987)**

**Definition.** Let \( \mathbf{x} \) be a \( p \times 1 \) random vector, \( \mu \) be a \( p \times 1 \) vector in \( \mathbb{R}^p \), and \( \Sigma \) be a \( p \times p \) non-negative definite matrix. Then \( \mathbf{x} \) has an elliptically contoured distribution, denoted by \( EC_p(\mu, \Sigma, g) \) if the characteristic function \( \phi_{\mathbf{x}-\mu}(t) \) of \( \mathbf{x} - \mu \) is a function of the quadratic form \( t'\Sigma t \) as \( \phi_{\mathbf{x}-\mu}(t) = \phi(t'\Sigma t) \). This form can be written as

\[
\phi_{\mathbf{x}-\mu}(t) = \exp(it'\mu)\psi(t'\Sigma t)
\]

for some function \( \psi \).

Therefore, the elliptically symmetric distributions denoted by \( EC_p(\mu, \Sigma, g) \), have the pdf in the form

\[
f(\mathbf{x}) = k_p|\Sigma|^{-\frac{1}{2}}g[(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)],
\]

where \( g \) is a one-dimensional real-valued function independent of \( p \) and

\[
k_p = \frac{p\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma(1 + \frac{p}{2\beta})2^{1 + \frac{p}{2\beta}}}.\]

For our distribution \( g(t) = \exp\{-t^\frac{1}{2}\} \).

(iii) **Power exponential distributions by Gómez, Gómez-Villegas and Marín (1998)**

**Definition.** A random vector \( \mathbf{x} \) has a \( p \)-dimensional power exponential distribution, denoted by \( PE_p(\mu, \Sigma, \beta) \), with \( \mu, \Sigma \) and \( \beta \), where \( \mu \in \mathbb{R}^p \), \( \Sigma \) is a \( p \times p \) positive definite symmetric matrix and \( \beta \in (0, \infty) \). Its density function is

\[
f(\mathbf{x}, \mu, \Sigma, \beta) = k|\Sigma|^{-\frac{1}{2}}\exp\{-\frac{1}{2}[(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)]^\beta\},
\]

where

\[
k = \frac{p\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p}{2}}\Gamma(1 + \frac{p}{2\beta})2^{1 + \frac{p}{2\beta}}}.\]

For \( \beta = \frac{1}{2} \) one obtains our distribution. This function is actually the pdf of an elliptically contoured random vector \( EC_p(\mu, \Sigma, g) \).

(iv) **Kotz type distributions proposed in Fang, Kotz and Ng (1990)**

**Definition.** If \( \mathbf{x} \sim EC_p(\mu, \Sigma, g) \) and the density generator \( g \) is of the form

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\( g(u) = c_p u^{N-1} \exp(-ru^s), \ r, s > 0, \ 2N + p > 2 \) then we say that \( x \) possesses a symmetric Kotz distribution. The \( pdf \) of \( x \) is given by

\[
f(x, \mu, \Sigma) = c_p \mid \Sigma \mid^{-\frac{1}{2}} \mid (x - \mu)\Sigma^{-1}(x - \mu) \mid^{N-1} \exp \{-r\mid (x - \mu)\Sigma^{-1}(x - \mu) \mid^s\},
\]

where \( c_p = \frac{s^\Gamma(\frac{p}{2})}{\pi^\frac{1}{2} \Gamma(\frac{N+p-2}{2})} r^{\frac{2N+p-2}{2}} \).

When \( N = 1, s = \frac{1}{2} \) and \( r = 1 \) the distribution reduces to our distribution.

Although this distribution can be called by any of these different names, for convenience of presentation, we have referred to this distribution in this thesis as a Kotz type distribution. Since for \( p = 1 \), this \( pdf \) reduces to that of a double exponential distribution we may consider this as a multivariate generalization of double exponential distribution. However, this is not a multivariate double exponential distribution because its marginals are not double exponential distributions.

Many researchers have discussed statistical inference using elliptical distributions (Fang and Anderson, 1990). This particular distribution however, is not considered in their studies. No general theory of elliptical distributions applies to this distribution because the joint \( pdf \) of independent samples, \( x_1, ..., x_n \) from this distribution cannot be written in the form of elliptical distributions.

In the following subsections we will provide various characteristics of Kotz type distribution, such as, moments, the marginal, and conditional distributions. A simulation algorithm and goodness-of-fit tests for Kotz type distribution are also provided. Estimation of parameters of this model using maximum likelihood method under a variety of covariance structures, such as, AR(1), equicorrelation and unstructured covariance is also discussed. We show that the MLE of the location parameter under the assumption of Kotz type distribution is same as the generalized spatial median (GSM) defined by Rao (1988). We provide computational algorithms and computer programs to compute the estimates.

### 2.1.2 Moments and Other Properties

In the following we provide the expected value, variance covariance matrix, and Mardia’s measures of skewness and kurtosis of the distribution given in (2.1) using some

\[ E(x) = \mu, \]
\[ Var(x) = (p + 1)\Sigma, \]
\[ \beta_{1p} = 0, \text{ Mardia's multivariate skewness measure}, \]
\[ \beta_{2p} = \frac{p(p + 2)(p + 3)}{(p + 1)}, \text{ Mardia's multivariate kurtosis measure}. \]

By Proposition 2.2 of Gómez, Gómez-Villegas and Marín (1998), if \( x \) has a pdf as in (2.1) then its characteristic function is

\[ \psi_{x}(t) = \frac{1}{\Gamma(p)} e^{it\mu} \int_{0}^{\infty} \Psi_{p}(r \sqrt{t'\Sigma t}) r^{p-1} e^{-r} dr, \]

where \( \Psi_{1}(x) = \cos x \), \( \Psi_{p}(x) = \frac{\Gamma\left(\frac{p}{2}\right)}{\pi^{\frac{p-1}{2}} r_{0}^{p}} \int_{0}^{\pi} \exp\{ix\cos\theta\} \sin^{p-2}\theta d\theta \), for every \( p > 1 \) and \( R \) is an absolutely continuous positive random variable whose density function is

\[ f_{R}(r) = \frac{1}{\Gamma(p)} r^{p-1} e^{-r} I_{0,\infty}(r). \]

### 2.1.3 Marginal and Conditional Distributions

Suppose \( x \) is partitioned as \( x = (x_{(1)}, x_{(2)})' \), where \( x_{(1)} = (x_{1}, \ldots, x_{k})' \), \( x_{(2)} = (x_{k+1}, \ldots, x_{p})' \) with \( k < p \) and similarly, \( \mu = (\mu_{(1)}, \mu_{(2)})' \) with \( \mu_{(1)} = (\mu_{1}, \ldots, \mu_{k})' \) and \( \mu_{(2)} = (\mu_{k+1}, \ldots, \mu_{p})' \). Further suppose \( \Sigma \) is accordingly partitioned as \( \Sigma = \left( \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right) \), where \( \Sigma_{11} \) is a \( k \times k \) p. d. matrix and \( \Sigma_{22} \) is a \( p - k \times p - k \) p. d. matrix and

\[ \Sigma_{12} = \Sigma_{21}. \]

1. \( x_{(1)} \) has an elliptically symmetric \( EC_{k}(\mu_{(1)}, \Sigma_{11}, g_{1}) \) distribution with

\[ g_{1}(t) = t^{\frac{p-k}{2}} \int_{0}^{1} (1 - \omega)^{\frac{p-k}{2} - 1} e^{-\frac{E}{2}} d\omega. \]

The marginal characteristics of \( x_{(1)} \) are

\[ E(x_{(1)}) = \mu_{(1)}, \]
\[ Var(x_{(1)}) = (p + 1)\Sigma_{11}, \]
\[ \beta_{1p}(x_{(1)}) = 0, \text{ and} \]
\[ \beta_{2p}(x_{(1)}) = \frac{k(k + 2)(p + 3)}{p + 1}. \]
2. The conditional distribution of \( x(2) \) given \( x(1) \) is elliptically contoured 
\[ EC_{p-k}(\mu_{2,1}, \Sigma_{22,1}, g_{2,1}) \], where 
\begin{align*}
\mu_{2,1} &= \mu_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(x_{(1)} - \mu_{(1)}), \\
\Sigma_{22,1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \quad \text{and} \\
g_{2,1}(t) &= \exp \{ -[t + (x_{(1)} - \mu_{(1)})'\Sigma_{11}^{-1}(x_{(1)} - \mu_{(1)})]^{\frac{1}{2}} \}.
\end{align*}

### 2.2 Simulation Algorithm

In this section we propose an algorithm for simulating data from Kotz type distribution. Naik and Patwardhan (1991) had used a similar method for simulating data from a bivariate Kotz type distribution. We shall use that method to generate a random sample from a \( p \)-variate Kotz type distribution. The proposed algorithm is given in the following steps.

**Step 1.** Simulate \( y' = (y_1, \ldots, y_p) \) having the density 
\[ f(y) = c \exp \left\{ -\sqrt{y'y} \right\}, \]
where \(-\infty < y_i < \infty\), and \( c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)} \). Note that \( f(y) \) is the standardized version of Kotz type distribution given in (2.1).

The simulation of \( y \) is achieved by using the polar coordinate transformation,
\begin{align*}
y_1 &= R \cos \theta_1 \\
y_2 &= R \sin \theta_1 \cos \theta_2 \\
&\vdots \\
y_{p-1} &= R \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \cos \theta_{p-1} \\
y_p &= R \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \sin \theta_{p-1},
\end{align*}
where \( \theta_j \in [0, \pi) \) for \( 1 \leq j \leq p - 2 \) and \( \theta_{p-1} \in [0, 2\pi) \). The Jacobian of the transformation is \( R^{p-1} \prod_{j=1}^{p-2} \sin^{p-j-1}(\theta_j) \).

For an odd number \( p \), \( R \) and \( \theta_j, \quad j = 1, \ldots, p - 1 \), are independently distributed with the probability density function given by 
\[ g(r) = \frac{1}{\Gamma(p)} r^{p-1} e^{-r}, \text{ that is, } R \sim G(p, 1) \text{ and} \]
For an even number $p$, $R$ and $\theta_j$, $j = 1, \ldots, p - 1$, are independently distributed with the probability density function given by

\[
g(r) = \frac{1}{\Gamma(p)} r^{p-1} e^{-r}, \text{ that is, } R \sim G(p, 1) \text{ and}
\]

\[
g(\theta_1) = \frac{2^{p-2}}{\pi} \left(\frac{p-2}{2}\right)! \sin^{p-2}(\theta_1)
\]

\[
g(\theta_2) = \frac{2^{p-4}}{\pi} \left(\frac{p-4}{2}\right)! \sin^{p-4}(\theta_2)
\]

\[
g(\theta_3) = \frac{2^{p-6}}{\pi} \left(\frac{p-6}{2}\right)! \sin^{p-6}(\theta_3)
\]

\[
g(\theta_4) = \frac{2^{p-8}}{\pi} \left(\frac{p-8}{2}\right)! \sin^{p-8}(\theta_4)
\]

\[
\vdots
\]

\[
g(\theta_{p-2}) = \frac{1}{2} \sin(\theta_{p-2})
\]

\[
g(\theta_{p-1}) = \frac{1}{2\pi}.
\]

We note that $E(y) = 0$ and $Var(y) = (p + 1)I$. Of course any uniform random number generating algorithm can be successfully used with the inverse cumulative distribution function to generate pseudo-random non-uniform distribution.
To simulate $\theta \sim g(\theta)$, we use the bisection method that is one of the three popular numerical inversion algorithms for $G(\theta) = U(0,1)$.

Algorithm: Find an initial interval $[a, b]$ to which the solution belongs.

REPEAT
\[
\theta \leftarrow \frac{(a+b)}{2}
\]
IF $G(\theta) \leq U$ THEN $a \leftarrow \theta$
ELSE $b \leftarrow \theta$
UNTIL $b - a \leq 2\delta$
RETURN $\theta$

Here $\delta$ is a small number where $\delta > 0$.

Step 2. Obtain $x' = (x_1, ..., x_p)$ having the distribution (2.1) by making the transformation $x = \Gamma y + \mu$, where $\mu' = (\mu_1, ..., \mu_p)$ and $\Gamma \Gamma = \Sigma$. Note that $E(x) = \mu$ and $V(x) = (p + 1)\Sigma$.

For example, to generate a 5-variate ($p = 5$) random vector $x' = (x_1, ..., x_5)$ having the distribution (2.1), first we simulate $y' = (y_1, ..., y_5)$ which has the density

$$f(y) = \frac{1}{64\pi^2} \exp\left\{-\sqrt{y'y}\right\}, \quad -\infty < y_i < \infty,$$

where
\[
\begin{align*}
y_1 &= R \cos \theta_1 \\
y_2 &= R \sin \theta_1 \cos \theta_2 \\
y_3 &= R \sin \theta_1 \sin \theta_2 \cos \theta_3 \\
y_4 &= R \sin \theta_1 \sin \theta_2 \sin \theta_3 \cos \theta_4 \\
y_5 &= R \sin \theta_1 \sin \theta_2 \sin \theta_3 \sin \theta_4
\end{align*}
\]

and $R$ and $\theta_j$ are independently distributed with
\[
\begin{align*}
g(r) &= \frac{1}{24} r^4 e^{-r}, \text{ that is, } R \sim G(5,1) \text{ and} \\
g(\theta_1) &= \frac{3}{4} \sin^3(\theta_1) \\
g(\theta_2) &= \frac{2}{\pi} \sin^2(\theta_2) \\
g(\theta_3) &= \frac{1}{2} \sin(\theta_3) \\
g(\theta_4) &= \frac{1}{2\pi}
\end{align*}
\]
where $\theta_j \in [0, \pi)$ for $j = 1, 2, 3$ and $\theta_4 \in [0, 2\pi)$. Then we obtain $x$ by making the transformation $x = \Gamma y + \mu$ for fixed $\mu$ and $\Gamma$.

In the later sections, we will use this simulation algorithm to generate data from the distribution (2.1) and discuss estimation of parameters.

### 2.3 Goodness-of-fit Tests for Kotz Type Distribution

We shall provide goodness-of-fit tests for Kotz type distribution using results by Baringhaus and Henze (1992) and Henze (1994).

Baringhaus and Henze (1992) have derived the asymptotic distribution of Mardia’s skewness measure, $b_{1p}$, under any elliptically symmetric distribution. The asymptotic distribution of Mardia’s skewness measure is a weighted sum of two independent $\chi^2$ variates. That is,

$$nb_{1p} \xrightarrow{D} \alpha_1 \chi^2_p + \alpha_2 \chi^2_{2(p-1)(p+1)},$$

where

$$\alpha_1 = \frac{3}{p} \left[ \frac{m_6}{p+2} - 2m_4 + p(p + 2) \right] \quad \text{and} \quad \alpha_2 = \frac{6m_6}{p(p+2)(p+4)}.$$

For the distribution given in (2.1), $m_4 = \frac{p(p+2)(p+3)}{p+1}$, and $m_6 = \frac{p(p+2)(p+3)(p+4)(p+5)}{(p+1)^2}$.

Henze (1994) has derived the asymptotic distribution of Mardia’s kurtosis measure, $b_{2p}$, under any elliptically symmetric distribution. The asymptotic distribution of Mardia’s kurtosis measure is

$$\sqrt{n} \left( b_{2p} - p(p + 1)(p + 2)(p + 3) \right) \xrightarrow{D} N(0, \tau^2),$$

where $\tau^2 = r_8 - r_4^2 + \frac{4}{p} r_4(r_4^2 - r_6)$.

For the distribution given in (2.1),

$$r_k = E[R^k] = p(p + 1)(p + 2)(p + 3)\ldots(p + (k - 1)), \quad k \geq 1$$

These facts can be used to test for this distribution.
2.4 Estimation of Parameters

The importance of the model as an alternative to the normal model can be seen because of the following. Reconsider the density given in (2.1)

\[ f(x, \mu, \Sigma) = c |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu)'\Sigma^{-1}(x - \mu) \right\}, \quad \mu \in \mathbb{R}^p, \Sigma \text{ p.d.} \]

First, we note that this distribution has fatter tail regions than that of multivariate normal distribution. Secondly, if we observe \( x_1, \ldots, x_n \) iid sample from this distribution then the joint pdf of \( x_1, \ldots, x_n \) does not belong to the class of so called elliptically symmetric distributions. Thus, when there is more than one sample from \( f(\cdot) \) above, the MLEs of \( \mu \) and \( \Sigma \) are not the same as those in the multivariate normal case. But for those elliptically symmetric distributions for which the joint pdf of \( x_1, \ldots, x_n \) is also elliptically symmetric, the MLEs of \( \mu \) and \( \Sigma \) (except for a constant) are the same as those for multivariate normal distribution. See Fang and Anderson (1990).

Let \( x_1, \ldots, x_n \) be a random sample from Kotz type distribution. Then the log-likelihood function is given by

\[ \log L(\mu, \Sigma) = n \ln c - \frac{n}{2} \ln |\Sigma| - \sum_{i=1}^{n} \sqrt{(x_i - \mu)'\Sigma^{-1}(x_i - \mu)} \]

The MLEs of \( \mu \) and \( \Sigma \) are obtained by minimizing

\[ \frac{n}{2} \ln |\Sigma| + \sum_{i=1}^{n} \sqrt{(x_i - \mu)'\Sigma^{-1}(x_i - \mu)} \]

w.r.t. \( \mu \) and \( \Sigma \) simultaneously.

When \( \Sigma = I \), the solution to the above problem or the MLE of \( \mu \) is the spatial median introduced by Haldane (1948) and for general \( \Sigma \) it is generalized spatial median introduced by Rao (1988) and studied by Naik (1993).

2.4.1 Generalized Spatial Median (GSM)

In this section we consider the estimation of the location parameter \( \mu \) of the Kotz type distribution given in (2.1). Haldane (1948) defined the spatial median of the points \( x_1, \ldots, x_n \) as a point \( \hat{\mu} \in \mathbb{R}^p \) which minimizes

\[ \sum_{i=1}^{n} \|x_i - \mu\| = \sum_{i=1}^{n} \sqrt{(x_i - \mu)'(x_i - \mu)} \]
with respect to $\mu$. For $p > 1$, the vector $\hat{\mu}$ is unique except when all the mass of the distribution is concentrated on a line (Haldane (1948) and Ducharme and Milasevic (1987)) and is invariant under orthogonal transformation but not under affine transformation (Brown (1983) and Ducharme and Milasevic (1987)).

Rao (1988) defined two generalized spatial medians which are invariant under affine transformation as follows

1. A vector $\hat{\mu}$ which minimizes
   \[ \sum_{i=1}^{n} \sqrt{(x_i - \mu)'S^{-1}(x_i - \mu)} \]  
   with respect to $\mu$, where $S$ is the usual sample variance covariance matrix.

2. A vector $\hat{\mu}$ which minimizes
   \[ \frac{n}{2} \ln |\Sigma| + \sum_{i=1}^{n} \sqrt{(x_i - \mu)'\Sigma^{-1}(x_i - \mu)} \]  
   simultaneously with respect to $\mu$ and $\Sigma$.

Thus, we note that the MLE of $\mu$ under the assumption of Kotz type distribution of $x_1, ..., x_n$ is same as the generalized spatial median defined by Rao (1988).

### 2.4.2 Computation of GSM and $\hat{\Sigma}$

Let $x_1, ..., x_n$ be a random sample from (2.1). Then the log-likelihood function can be written as

\[ \ln L(\mu, \Sigma) = n \ln c - \frac{n}{2} \ln |\Sigma| - \sum_{i=1}^{n} \sqrt{(x_i - \mu)'\Sigma^{-1}(x_i - \mu)}. \]  

For a fixed general covariance structure $\Sigma$, Naik (1993) first showed estimation of $\hat{\mu}$ which minimizes $\sum_{i=1}^{n} \sqrt{(x_i - \mu)'\Sigma^{-1}(x_i - \mu)}$ w.r.t. $\mu$ by making transformation

\[ y_i = G^{-1}x_i \text{ and } \nu = G^{-1}\mu, \]

where $\Sigma = GG'$. Thus, he obtained the generalized spatial median $\hat{\nu}$ and then obtained

\[ \hat{\mu} = G\hat{\nu}. \]
Next the MLE of $\Sigma$ is obtained as the matrix $\hat{\Sigma}$ which minimizes (2.3) with respect to $\Sigma$ as a solution to the non-linear equations given by

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})(x_i - \hat{\mu})'}{\sqrt{(x_i - \hat{\mu})'\Sigma^{-1}(x_i - \hat{\mu})}}.$$

Solving these equations generally requires computational algorithms. We have adopted SAS' IML procedure for writing the computer programs. The Newton–Raphson method was used for optimization and to get the estimates. The optimization gives unique ML estimates in the feasible regions under certain covariance structures. The GSM results are similar to that of Gower’s algorithm (Gower, 1974).

Note that maximizing (2.4) simultaneously with respect to all parameters $\mu = (\mu_1, ..., \mu_p)'$ and parameters that appear in $\Sigma$, is equivalent to minimizing (2.3) simultaneously with respect to $\mu$ and $\Sigma$. The steps involved in our optimization procedure are the following:

**Step 1.** Take the initial values for $\mu$ and $\Sigma$, say $\hat{\mu} = \bar{x}$ and $\hat{\Sigma} = S$ (for general covariance structure) or when $\Sigma = \Sigma(\theta)$ is a structured covariance matrix depending on the parameter vector $\theta$, take $\hat{\Sigma} = \Sigma(\hat{\theta})$, where $\hat{\theta}$ is the vector of initial estimators obtained using the sample. For example, if $\Sigma = \Sigma(\sigma, \delta)$, (equicorrelation structure with $\sigma$ as the common standard deviation and $\delta$ as the correlation coefficient) take $\hat{\sigma}^2$ = the p-sample pooled variance and $\hat{\delta}$ = the average of the p-sample correlation coefficients.

**Step 2.** Using initial values and the optimization procedure, minimize (2.3) to get $\hat{\mu}$ and $\hat{\Sigma}$.

**Step 3.** Taking $\hat{\mu}$ and $\hat{\Sigma}$ as the initial values repeat Step 2. Continue until termination rules for all the estimators are met. For example, for equicorrelation structure, continue until

$$||\hat{\mu}_{i+1} - \hat{\mu}_{i}|| \leq \epsilon_1,$$

$$|\hat{\sigma}_{i+1}\hat{\sigma}_i^{-1} - 1| \leq \epsilon_2,$$

$$|\hat{\delta}_{i+1} - \hat{\delta}_i| \leq \epsilon_3,$$

where $\epsilon_1 = \epsilon_2 = \epsilon_3 = 10^{-3}$ (say) and $\hat{\mu}_i, \hat{\sigma}_i,$ and $\hat{\delta}_i$ are the estimators of $\mu_i, \sigma_i,$ and $\delta_i$ respectively at the $i^{th}$ iteration.
2.4.3 Computation of GSM and \( \hat{\Sigma} \) under Different Covariance Structures

The computation of GSM can become easy under simpler structures of \( \Sigma \). In the following we present computation formulae for different structures.

The General Covariance Structure:

Suppose \( \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{pmatrix} \)

\( = (\sigma_{ij}) \),

where for \( i, j = 1, \ldots, p, \sigma_{ij} > 0 \) if \( i = j \) and \( \sigma_{ij} \in \mathbb{R} \) if \( i \neq j \).

Naik and Patwardhan (1991) showed computation of GSM for the bivariate case under the general covariance structure,

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

The ML estimators of \( \rho, \sigma_1^2, \sigma_2^2 \) can be obtained by solving the following three equations simultaneously.

\[
\begin{align*}
np\sqrt{1-\rho^2} + (1-\rho^2) \sum_{i=1}^{n} \frac{(x_{1i} - \hat{\mu}_1)(x_{2i} - \hat{\mu}_2)}{\sqrt{Q_i}} - \rho \sum_{i=1}^{n} \sqrt{Q_i} &= 0 \\
\sigma_1^2 - \frac{1}{n} \frac{1}{\sqrt{1-\rho^2}} \sum_{i=1}^{n} \frac{(x_{1i} - \hat{\mu}_1)^2}{\sqrt{Q_i}} - \rho^2 \frac{1}{\sigma_2^2} (x_{1i} - \hat{\mu}_1)(x_{2i} - \hat{\mu}_2) &= 0 \\
\sigma_2^2 - \frac{1}{n} \frac{1}{\sqrt{1-\rho^2}} \sum_{i=1}^{n} \frac{(x_{2i} - \hat{\mu}_2)^2}{\sqrt{Q_i}} - \rho^2 \frac{1}{\sigma_1^2} (x_{1i} - \hat{\mu}_1)(x_{2i} - \hat{\mu}_2) &= 0,
\end{align*}
\]

where

\[ Q_i = \left( \frac{x_{1i} - \hat{\mu}_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_{1i} - \hat{\mu}_1}{\sigma_1} \right) \left( \frac{x_{2i} - \hat{\mu}_2}{\sigma_2} \right) + \left( \frac{x_{2i} - \hat{\mu}_2}{\sigma_2} \right)^2, \quad i = 1, \ldots, n. \]

It was shown by Naik and Patwardhan (1991) that there is at least one real solution for each of these equations, when \( -1 < \rho < 1 \). For a \( p \)-variate case we have used
optimization algorithm and SAS software to find the estimates.

The Diagonal Covariance Structure:

The structure

\[ \Sigma = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{pmatrix}, \]

where \( \sigma_{ii} > 0 \) for \( i = 1, \ldots, p \) can be used and the ML estimates can be easily obtained.

The Equicorrelation Structure:

This structure, sometimes known as the compound symmetry structure or the intraclass correlation structure, has the same correlation coefficient, say \( \delta \). The covariance matrix has the form,

\[ \Sigma = \sigma^2 [(1 - \delta)I_p + \delta J_p] \]

\[ = \sigma^2 \begin{pmatrix} 1 & \delta & \cdots & \delta \\ \delta & 1 & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & 1 \end{pmatrix} = \sigma^2 V(\delta). \]

The determinant and inverse of \( \Sigma \) respectively are

\[ |\Sigma| = (\sigma^2)^p [(1 - \delta)^{p-1}(1 + (p - 1)\delta)] \quad \text{and} \]

\[ \Sigma^{-1} = \frac{1}{\sigma^2(1 - \delta)}[I_p - \frac{\delta}{(1 - (p - 1)\delta)}J_p], \]

where \( \sigma^2 > 0, \ -\frac{1}{p-1} < \delta < 1, \) that ensure \( \Sigma \) to be the positive definite matrix, and \( I_p \) is an identity matrix of order \( p \) and \( J_p \) is a \( p \times p \) matrix of all ones.

Naik (1993) derived the estimators of \( \mu \) and \( \Sigma \); that is, of \( \mu, \sigma \) and \( \delta \) by maximizing the log-likelihood function of the distribution with respect to \( \mu, \sigma \) and \( \delta \). The MLE of \( \sigma \) and \( \delta \) were obtained by solving the following non-linear equations:
\[ [1 - (p - 1)(1 - \delta)] \sum_{i=1}^{n} \sqrt{Q_1(\delta)} + \\
\frac{[1 + (p - 1)\delta](1 - \delta)}{\sqrt{Q_1(\delta)}} \sum_{i=1}^{n} \frac{(p - 1)(x_i - \mu)'(x_i - \mu) - [1'(x_i - \mu)]^2}{\sqrt{Q_1(\delta)}} = 0 \]

for \( \delta \), and

\[ \hat{\sigma} = \frac{1}{np\sqrt{[1 + (p - 1)\delta](1 - \delta)}} \sum_{i=1}^{n} \sqrt{Q_1(\delta)} \]

for \( \hat{\sigma} \).

Here \( Q_1(\delta) = [1 + (p - 1)\delta]x_i - \hat{\mu}'x_i - \hat{\mu} - \delta_1(x_i - \hat{\mu}')1^2. \)

However, in this thesis we directly use the optimization algorithm to compute these estimates. In addition, we have implemented the algorithm under the following structures.

**The First Order of Autoregressive Structure, AR(1):**

This covariance structure is often used in time series models. It also includes 2 parameters, \( \sigma \) and \( \rho \) and the structure is given by

\[ \Sigma = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{p-1} \\ \rho & 1 & \cdots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \cdots & 1 \end{pmatrix} \]

\[ = \frac{\sigma^2}{1 - \rho^2} V(\rho), \]

where \( V(\rho) = \rho^{i-j}, \ i, j = 1, \ldots, p, \ \sigma^2 > 0, \ -1 < \rho < 1. \)

The determinant and inverse respectively are

\[ |\Sigma| = \frac{(\sigma^2)^p}{1 - \rho^2} \quad \text{and} \quad \Sigma^{-1} = \frac{1}{\sigma^2} [I_p + \rho^2 C_1 - \rho C_2], \]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
where
\[
C_1 = \begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 1 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 1 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0
\end{pmatrix},
\]
(2.5)
\[
C_2 = \begin{pmatrix}
0 & 1 & 0 & \ldots & 0 & 0 & 0 \\
1 & 0 & 1 & \ldots & 0 & 0 & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 1 & 0 \\
0 & 0 & 0 & \ldots & 1 & 0 & 1 \\
0 & 0 & 0 & \ldots & 0 & 1 & 0
\end{pmatrix},
\]
(2.6)

The Circular Covariance (1):

These structures are sometimes useful for modelling certain types of physical phenomenon. For example, see Khattree and Naik (1994). The form of the structure, for example, for \( p = 4 \) and \( p = 5 \) are:
\[
\Sigma = \sigma^2 \begin{pmatrix}
1 & \rho_1 & \rho_2 & \rho_3 \\
\rho_1 & 1 & \rho_2 & \rho_3 \\
\rho_2 & \rho_1 & 1 & \rho_3 \\
\rho_3 & \rho_2 & \rho_1 & 1
\end{pmatrix}
\]
and
\[
\Sigma = \sigma^2 \begin{pmatrix}
1 & \rho_1 & \rho_2 & \rho_3 & \rho_4 \\
\rho_1 & 1 & \rho_2 & \rho_3 & \rho_4 \\
\rho_2 & \rho_1 & 1 & \rho_3 & \rho_4 \\
\rho_3 & \rho_2 & \rho_1 & 1 & \rho_4 \\
\rho_4 & \rho_3 & \rho_2 & \rho_1 & 1
\end{pmatrix},
\]
where \( \sigma > 0 \) and \(-1 < \rho_i < 1, i = 1, 2, \ldots, p\).

The Circular Covariance (2):

Another form of Circular Covariance structure that was used by Hartley and Naik (2001) has the form for \( p = 4 \) and \( p = 5 \):
\[
\Sigma = \sigma^2 \begin{pmatrix}
1 & \rho & \rho^2 & \rho \\
\rho & 1 & \rho & \rho^2 \\
\rho^2 & \rho & 1 & \rho \\
\rho & \rho^2 & \rho & 1
\end{pmatrix}
\]
and
\[
\Sigma = \sigma^2 \begin{pmatrix}
1 & \rho & \rho^2 & \rho & \rho^2 \\
\rho & 1 & \rho & \rho^2 & \rho \\
\rho^2 & \rho & 1 & \rho & \rho^2 \\
\rho & \rho^2 & \rho & 1 & \rho \\
\rho^2 & \rho & \rho & \rho & 1
\end{pmatrix},
\]
In this case, \( \sigma > 0 \) and \(-1 < \rho < 1\).
A Special Case General Structure:

Sometimes the interest may be in fitting the general correlation structure, that is, 
\( \Sigma = R \), where

\[
R = \begin{pmatrix}
1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\
\rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{1p} & \rho_{2p} & \cdots & 1
\end{pmatrix}
\]

and \(-1 < \rho_{ij} < 1, \ i, j = 1, \ldots, p\).

2.5 Example

To illustrate the estimation of \( \mu \) and \( \Sigma \) under a variety of covariance structures, we select the Board data consisting of four different measures of stiffness on each of \( n = 30 \) boards. The first measurement \( x_1 \) involves sending a shock wave down the board, the second measurement \( x_2 \) is determined while vibrating the board and the last two measurements \( x_3 \) and \( x_4 \) are obtained from static tests. These data are taken from Table 4.3 in Johnson and Wichern (1998, p. 198) and displayed in Table 2.1. First, using the methods discussed in Chapter 1, we tested for normality of these data and we found that the data are not multivariate normal. Then we used goodness-of-fit tests for Kotz type distribution and found that the Board data are from Kotz type distribution with \( p\)-value = 0.0897 for Mardia’s skewness measure and \( p\)-value = 0.6627 for Mardia’s kurtosis measure. Next, we provide sample statistics, namely, sample mean, \( \bar{x} \), covariance matrix, \( S \), and correlation matrix, \( R \), and then provide ML estimates of the parameters involved.
Table 2.1: Board data: measurements on stiffness.

<table>
<thead>
<tr>
<th>Observation no.</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1889</td>
<td>1651</td>
<td>1561</td>
<td>1778</td>
</tr>
<tr>
<td>2</td>
<td>2403</td>
<td>2048</td>
<td>2087</td>
<td>2197</td>
</tr>
<tr>
<td>3</td>
<td>2119</td>
<td>1700</td>
<td>1815</td>
<td>2222</td>
</tr>
<tr>
<td>4</td>
<td>1645</td>
<td>1627</td>
<td>1110</td>
<td>1533</td>
</tr>
<tr>
<td>5</td>
<td>1976</td>
<td>1916</td>
<td>1614</td>
<td>1883</td>
</tr>
<tr>
<td>6</td>
<td>1712</td>
<td>1712</td>
<td>1439</td>
<td>1546</td>
</tr>
<tr>
<td>7</td>
<td>1943</td>
<td>1685</td>
<td>1271</td>
<td>1671</td>
</tr>
<tr>
<td>8</td>
<td>2104</td>
<td>1820</td>
<td>1717</td>
<td>1874</td>
</tr>
<tr>
<td>9</td>
<td>2983</td>
<td>2794</td>
<td>2412</td>
<td>2581</td>
</tr>
<tr>
<td>10</td>
<td>1745</td>
<td>1600</td>
<td>1384</td>
<td>1508</td>
</tr>
<tr>
<td>11</td>
<td>1710</td>
<td>1591</td>
<td>1518</td>
<td>1667</td>
</tr>
<tr>
<td>12</td>
<td>2046</td>
<td>1907</td>
<td>1627</td>
<td>1898</td>
</tr>
<tr>
<td>13</td>
<td>1840</td>
<td>1841</td>
<td>1595</td>
<td>1741</td>
</tr>
<tr>
<td>14</td>
<td>1867</td>
<td>1685</td>
<td>1493</td>
<td>1678</td>
</tr>
<tr>
<td>15</td>
<td>1859</td>
<td>1649</td>
<td>1389</td>
<td>1714</td>
</tr>
<tr>
<td>16</td>
<td>1954</td>
<td>2149</td>
<td>1180</td>
<td>1281</td>
</tr>
<tr>
<td>17</td>
<td>1325</td>
<td>1170</td>
<td>1002</td>
<td>1176</td>
</tr>
<tr>
<td>18</td>
<td>1419</td>
<td>1371</td>
<td>1252</td>
<td>1308</td>
</tr>
<tr>
<td>19</td>
<td>1828</td>
<td>1634</td>
<td>1602</td>
<td>1755</td>
</tr>
<tr>
<td>20</td>
<td>1725</td>
<td>1594</td>
<td>1313</td>
<td>1646</td>
</tr>
<tr>
<td>21</td>
<td>2276</td>
<td>2189</td>
<td>1547</td>
<td>2111</td>
</tr>
<tr>
<td>22</td>
<td>1899</td>
<td>1614</td>
<td>1422</td>
<td>1477</td>
</tr>
<tr>
<td>23</td>
<td>1633</td>
<td>1513</td>
<td>1290</td>
<td>1516</td>
</tr>
<tr>
<td>24</td>
<td>2061</td>
<td>1867</td>
<td>1646</td>
<td>2037</td>
</tr>
<tr>
<td>25</td>
<td>1856</td>
<td>1493</td>
<td>1356</td>
<td>1533</td>
</tr>
<tr>
<td>26</td>
<td>1727</td>
<td>1412</td>
<td>1238</td>
<td>1469</td>
</tr>
<tr>
<td>27</td>
<td>2168</td>
<td>1896</td>
<td>1701</td>
<td>1834</td>
</tr>
<tr>
<td>28</td>
<td>1655</td>
<td>1675</td>
<td>1414</td>
<td>1597</td>
</tr>
<tr>
<td>29</td>
<td>2326</td>
<td>2301</td>
<td>2065</td>
<td>2234</td>
</tr>
<tr>
<td>30</td>
<td>1490</td>
<td>1382</td>
<td>1214</td>
<td>1284</td>
</tr>
</tbody>
</table>
The sample mean, covariance matrix, and correlation matrix, respectively are

\[ \bar{x} = \begin{pmatrix} 1906.1 & 1749.5333 & 1509.1333 & 1724.966 \end{pmatrix}', \]

\[ S = \begin{pmatrix} 105616.30 & 94613.531 & 87289.710 & 94230.728 \\ 94613.531 & 101510.12 & 76137.099 & 81064.363 \\ 87289.710 & 76137.099 & 91917.085 & 90352.384 \\ 94230.728 & 81064.363 & 90352.384 & 104227.96 \end{pmatrix}, \]

\[ R = \begin{pmatrix} 1 & 0.9137620 & 0.8859301 & 0.8981212 \\ 0.9137620 & 1 & 0.7882129 & 0.7881034 \\ 0.8859301 & 0.7882129 & 1 & 0.9231013 \\ 0.8981212 & 0.7881034 & 0.9231013 & 1 \end{pmatrix}. \]

In the following we provide ML estimates of \( \mu \) and parameters of \( \Sigma \) under various structures for \( \Sigma \) for the example.

**Under General Covariance Structure for \( \Sigma \), we have \( p + \frac{(p+1)p}{2} \) parameters to estimate.**

\[ \hat{\mu} = \begin{pmatrix} 1850.4159 & 1691.1202 & 1481.2977 & 1685.424 \end{pmatrix}', \]

\[ \hat{\Sigma} = \begin{pmatrix} 91485.308 & 81017.783 & 77644.475 & 86380.137 \\ 81017.783 & 75470.307 & 69361.802 & 77060.054 \\ 77644.475 & 69361.802 & 70111.928 & 75515.426 \\ 86380.137 & 77060.054 & 75515.426 & 85505.102 \end{pmatrix}. \]

**Under Diagonal Covariance Structure for \( \Sigma \), we have \( 2p \) parameters to estimate.** The estimation results are

\[ \hat{\mu} = \begin{pmatrix} 1852.7454 & 1685.772 & 1474.3039 & 1681.4894 \end{pmatrix}', \]

\[ \hat{\Sigma} = \begin{pmatrix} 14678.514 & 0 & 0 & 0 \\ 0 & 13766.555 & 0 & 0 \\ 0 & 0 & 14022.531 & 0 \\ 0 & 0 & 0 & 16301.33 \end{pmatrix}. \]
Under Equicorrelation Structure for $\Sigma$, we have $p + 2$ parameters to estimate. The estimation results are

\[
\hat{\mu} = \left(1859.9366, 1697.3872, 1484.3861, 1691.7187\right)^T,
\]

\[
\hat{\sigma} = 139.29078,
\]

\[
\hat{\delta} = 0.8898684,
\]

so

\[
\hat{\Sigma} = \begin{pmatrix}
19401.922 & 17265.158 & 17265.158 & 17265.158 \\
17265.158 & 19401.922 & 17265.158 & 17265.158 \\
17265.158 & 17265.158 & 19401.922 & 17265.158 \\
17265.158 & 17265.158 & 17265.158 & 19401.922 \\
\end{pmatrix}
\]

Under the First Order of Autoregressive Structure, AR(1) for $\Sigma$, we have $p + 2$ parameters to estimate. The estimation results are

\[
\hat{\mu} = \left(1859.3134, 1694.2681, 1485.1081, 1688.5484\right)^T,
\]

\[
\hat{\sigma} = 62.804267,
\]

\[
\hat{\rho} = 0.8946798,
\]

so

\[
\hat{\Sigma} = \begin{pmatrix}
19766.542 & 17684.726 & 15822.166 & 14155.772 \\
17684.726 & 19766.542 & 17684.726 & 15822.166 \\
15822.166 & 17684.726 & 19766.542 & 17684.726 \\
14155.772 & 15822.166 & 17684.726 & 19766.542 \\
\end{pmatrix}
\]

Under the Circular Covariance (1) Structure for $\Sigma$, we have $p + 3$ parameters corresponding to $\mu$ and $\sigma > 0$ and $-1 < \rho_i < 1, i = 1, 2$. The estimation results are

\[
\hat{\mu} = \left(1858.0636, 1695.4335, 1482.6518, 1690.9761\right)^T,
\]

\[
\hat{\sigma} = 112.65213,
\]

\[
\hat{\rho} = \left(0.7999, 0.7693483\right)^T,
\]
Under the Circular Covariance (2) Structure for $\Sigma$, we have $p + 2$ parameters to estimate. The estimation results are

$$\hat{\mu} = \begin{pmatrix} 1850.1956, 1692.4255, 1480.7527, 1689.6937 \end{pmatrix},$$

$$\hat{\sigma} = 115.34328,$$

$$\hat{\rho} = 0.7292725,$$

so $\hat{\Sigma} = \begin{pmatrix} 12690.502 & 10151.132 & 9763.4154 & 10151.132 \\ 10151.132 & 12690.502 & 10151.132 & 9763.4154 \\ 9763.4154 & 10151.132 & 12690.502 & 10151.132 \\ 10151.132 & 9763.4154 & 10151.132 & 12690.502 \end{pmatrix}$.$$

Finally, under a Special Case General Structure for $\Sigma$, we have $p + \frac{p(p-1)}{2}$ parameters to estimate. The estimation results are

$$\hat{\mu} = \begin{pmatrix} 13304.072, 9702.2945, 7075.6168, 9702.2945 \end{pmatrix},$$

$$\hat{\Sigma} = \begin{pmatrix} 1 & 0.5246036 & 0.4518808 & 0.5346612 \\ 0.5246036 & 1 & 0.371565 & 0.4128659 \\ 0.4518808 & 0.371565 & 1 & 0.550590 \\ 0.5346612 & 0.4128659 & 0.550590 & 1 \end{pmatrix}$.$$

We have used this example to illustrate the computation of the MLE's under various covariance structures. All the computations are done using programs written in SAS/IML software. The SAS program for computing the MLEs for $\mu$ and the general covariance $\Sigma$ is provided in Program 2 in APPENDIX.
2.6 Comparisons of ML Estimates Between Kotz Type and Normal Populations

For non-normal distribution, theoretical expressions for the bias, square root of mean squared error and Pitman Nearness probability of the ML estimates are not generally available but they can be estimated using simulation.

In this section, we want to compare the performance of the ML estimators of \( \mu \) and \( \Sigma \) under the Kotz type distribution with those under normal distribution when the data are simulated from normal and Kotz type distributions. We can expect the ML estimators under Kotz type distribution to do better than normal ML estimators when the data are simulated from Kotz type distribution and vice versa.

To compare the ML estimates under Kotz type population with those under normal population, it is desired to estimate the bias, square root of mean squared error and Pitman Nearness probability of estimates of \( \mu \) and \( \Sigma \). We shall focus on the comparisons of ML estimates of \( \mu \) and \( \Sigma \) using these three criteria based on 1000 random samples of various sample sizes \( n \) and variates \( p \) generated from Kotz type and normal populations.

As we have discussed earlier, the ML estimates of \( \mu \) and \( \Sigma \) under the general covariance structure are the GSM \( \hat{\mu} \) and \( \hat{\Sigma} \) for Kotz type population and they are \( \bar{x} \) and \( S_n \) for normal population. When the simulation of data is from Kotz type population, we compute \( \hat{\mu}_j = (\hat{\mu}_{1j}, ..., \hat{\mu}_{pj})' \) and \( \hat{\Sigma}_j \) as Kotz type ML estimates of \( \mu \) and \( \Sigma \) and \( \bar{x}_j = (\bar{x}_{1j}, ..., \bar{x}_{pj})' \) and \( \frac{1}{p+1}S_{nj} \) as normal ML estimates of \( \mu \) and \( \Sigma \) for the \( j^{th} \) simulation, \( j = 1, ..., 1000 \). However, when the simulation of data is from normal population, we compute \( \bar{x}_j, S_{nj} \) as normal ML estimates of \( \mu \) and \( \Sigma \) and \( \hat{\mu}_j, (p+1)\hat{\Sigma}_j \) as Kotz type ML estimates of \( \mu \) and \( \Sigma \) for the \( j^{th} \) simulation, \( j = 1, ..., 1000 \). A reason for doing these adjustments is that \( Var(x) = \Sigma \) if \( x \sim N_p(\mu, \Sigma) \) and \( Var(x) = (p + 1)\Sigma \) if \( x \) is distributed as Kotz type.

Also for convenience of presentation, suppose all the \( \nu = \frac{p(p+1)}{2} \) unknown parameters in \( \Sigma \) are placed in a vector \( (\sigma_1, ..., \sigma_\nu) \) (or \( s_1, ..., s_\nu \) are the corresponding elements of \( S_n \)) and the estimates of \( \sigma_i \) for the \( j^{th} \) simulation is denoted by \( \hat{\sigma}_{ij} \) (or \( s_{ij} \)). Then
we have the following formulae for computation.

(i) **Bias:**

The estimated bias for $\mu$:

$$bias(\hat{\mu}) = \sqrt{\sum_{i=1}^{p} (\hat{\mu}_i - \mu_i)^2},$$

$$bias(\bar{\mu}) = \sqrt{\sum_{i=1}^{p} (\bar{\mu}_i - \mu_i)^2}.$$

The estimated bias for $\Sigma$:

For Kotz type distribution,

$$bias(\hat{\Sigma}) = \sqrt{\sum_{i=1}^{\nu} (\hat{\Sigma}_i - \Sigma_i)^2},$$

$$bias(\frac{1}{p+1}S_n) = \sqrt{\sum_{i=1}^{\nu} \left(\frac{1}{p+1} \hat{s}_i - \sigma_i\right)^2}.$$

For Normal distribution,

$$bias((p+1)\hat{\Sigma}) = \sqrt{\sum_{i=1}^{\nu} \left((p+1)\hat{\Sigma}_i - \Sigma_i\right)^2},$$

$$bias(S_n) = \sqrt{\sum_{i=1}^{\nu} (\hat{s}_i - \sigma_i)^2},$$

where for any estimate, $\hat{\theta} = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\theta}_j$, is the average value of the estimate from all 1000 simulations.

(ii) **Square root of mean squared error (root MSE):**

The estimated root MSE for $\mu$:

$$\text{root MSE}(\hat{\mu}) = \sqrt{\frac{\sum_{i=1}^{p} \sum_{j=1}^{1000} (\hat{\mu}_{ij} - \mu_i)^2}{1000}},$$
\[
\text{root MSE}(\bar{x}) = \sqrt{\frac{\sum_{i=1}^{p} \sum_{j=1}^{1000} (\bar{x}_{ij} - \mu_i)^2}{1000}}.
\]

The estimated root MSE for \( \Sigma \):

For Kotz type distribution,
\[
\text{root MSE}(\hat{\Sigma}) = \sqrt{\frac{\sum_{i=1}^{\nu} \sum_{j=1}^{1000} (\hat{\sigma}_{ij} - \sigma_i)^2}{1000}},
\]
and
\[
\text{root MSE}(\frac{1}{p+1}S_n) = \sqrt{\frac{\sum_{i=1}^{\nu} \sum_{j=1}^{1000} \left(\frac{1}{p+1}s_{ij} - \sigma_i\right)^2}{1000}}.
\]

For Normal distribution,
\[
\text{root MSE}\left(p+1\hat{\Sigma}\right) = \sqrt{\frac{\sum_{i=1}^{\nu} \sum_{j=1}^{1000} \left((p+1)\hat{\sigma}_{ij} - \sigma_i\right)^2}{1000}},
\]
and
\[
\text{root MSE}(S_n) = \sqrt{\frac{\sum_{i=1}^{\nu} \sum_{j=1}^{1000} (s_{ij} - \sigma_i)^2}{1000}}.
\]

(iii) Pitman Nearness (PN) Probability:

For Kotz type distribution, the estimate of Pitman Nearness probability of \( \hat{\mu} \) relative to \( \bar{x} \) is computed as
\[
PN(\hat{\mu}, \bar{x}) = \frac{1}{1000} \left( \# \text{ of times } \|\hat{\mu}_j - \mu\| < \|\bar{x}_j - \mu\| \right) \text{ in 1000 samples},
\]
where
\[
\|\hat{\mu} - \mu\| = \sqrt{\sum_{i=1}^{p} (\hat{\mu}_i - \mu_i)^2}
\]
and
\[
\|\bar{x} - \mu\| = \sqrt{\sum_{i=1}^{p} (\bar{x}_i - \mu_i)^2}
\]
and the estimate of Pitman Nearness probability of \( \hat{\Sigma} \) relative to \( \frac{1}{p+1}S_n \) is computed as
\[
PN(\hat{\Sigma}, \frac{1}{p+1}S_n) = \frac{1}{1000} \left( \# \text{ of times } \|\hat{\Sigma}_j - \Sigma\| < \|\frac{1}{p+1}S_{nj} - \Sigma\| \right) \text{ in 1000 samples},
\]
where
\[ \| \hat{\Sigma} - \Sigma \| = \sqrt{\sum_{i=1}^{\nu} (\hat{\sigma}_i - \sigma_i)^2} \quad \text{and} \]
\[ \| \frac{1}{p+1} \mathbf{S}_n - \Sigma \| = \sqrt{\sum_{i=1}^{\nu} \left( \frac{1}{p+1} s_i - \sigma_i \right)^2} \]

For normal distribution, the estimate of Pitman Nearness probability of \( \bar{x} \) relative to \( \hat{\mu} \) is computed as
\[ PN(\bar{x}, \hat{\mu}) = \frac{1}{1000} \left( \text{# of times } \| \bar{x}_j - \mu \| < \| \hat{\mu}_j - \mu \| \right) \text{ in 1000 samples}, \]
where
\[ \| \bar{x} - \mu \| = \sqrt{\sum_{i=1}^{p} (\bar{x}_i - \mu_i)^2} \quad \text{and} \]
\[ \| \hat{\mu} - \mu \| = \sqrt{\sum_{i=1}^{p} (\hat{\mu}_i - \mu_i)^2} \]

and the estimate of Pitman Nearness probability of \( \mathbf{S}_n \) relative to \( (p+1)\hat{\Sigma} \) is computed as
\[ PN\left( \mathbf{S}_n, (p+1)\hat{\Sigma} \right) = \frac{1}{1000} \left( \text{# of times } \| \mathbf{S}_{nj} - \Sigma \| < \| (p+1)\hat{\Sigma}_j - \Sigma \| \right) \text{ in 1000 samples}, \]
where
\[ \| \mathbf{S}_n - \Sigma \| = \sqrt{\sum_{i=1}^{\nu} (s_i - \sigma_i)^2} \quad \text{and} \]
\[ \| (p+1)\hat{\Sigma} - \Sigma \| = \sqrt{\sum_{i=1}^{\nu} \left( (p+1)\hat{\sigma}_i - \sigma_i \right)^2} \]

The comparison results for all three measures are provided in the following tables. When the data are simulated from Kotz type population, the estimates of bias for MLEs under Kotz type population appear to be smaller than those for MLEs under normal population with the following relationship (see Tables 2.2 - 2.3):
\[ \text{bias}(\hat{\mu}) < \text{bias}(\bar{x}), \]
\[ \text{bias}(\hat{\Sigma}) < \text{bias}\left(\frac{1}{p+1}S_n\right). \]

The estimates of root MSE for MLEs under Kotz type population also appear to be smaller than those for MLEs under normal population with following relationship:

\[ \text{root MSE}(\hat{\mu}) < \text{root MSE}(\bar{x}), \]
\[ \text{root MSE}(\hat{\Sigma}) < \text{root MSE}\left(\frac{1}{p+1}S_n\right). \]

Moreover, the estimates of Pitman Nearness probability of MLEs under Kotz type population relative to MLEs under normal population are greater than 0.5. Thus, we can conclude that the ML estimators under Kotz type distribution do better than normal ML estimators when the data are simulated from Kotz type distribution.

Table 2.2: The estimates of bias and root MSE for \( \hat{\mu} \) and \( \bar{x} \) and Pitman nearness probability of \( \hat{\mu} \) relative to \( \bar{x} \) for Kotz type samples.

<table>
<thead>
<tr>
<th>p</th>
<th>n</th>
<th>Bias(( \hat{\mu} ))</th>
<th>Bias(( \bar{x} ))</th>
<th>root MSE(( \hat{\mu} ))</th>
<th>root MSE(( \bar{x} ))</th>
<th>PN(( \hat{\mu}, \bar{x} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.0327292</td>
<td>0.0429589</td>
<td>1.2266231</td>
<td>1.4047429</td>
<td>0.629</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.0264755</td>
<td>0.0479547</td>
<td>0.9710988</td>
<td>1.1197025</td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0263113</td>
<td>0.0357640</td>
<td>0.7372502</td>
<td>0.8777592</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0063597</td>
<td>0.0139617</td>
<td>0.5297203</td>
<td>0.6464303</td>
<td>0.663</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.1504333</td>
<td>0.1609594</td>
<td>7.4264586</td>
<td>7.9686126</td>
<td>0.618</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.3864212</td>
<td>0.4112645</td>
<td>6.2005286</td>
<td>6.6546604</td>
<td>0.623</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.1438450</td>
<td>0.1533291</td>
<td>4.7531360</td>
<td>5.1423876</td>
<td>0.611</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.1152852</td>
<td>0.1342547</td>
<td>3.3385742</td>
<td>3.6247224</td>
<td>0.647</td>
</tr>
</tbody>
</table>

When the simulated data are from normal population, the estimates of bias for MLEs under Kotz type population appear to be greater than those for MLEs under normal population with the following relationship (see Tables 2.4 - 2.5):

\[ \text{bias}(\bar{x}) < \text{bias}(\hat{\mu}), \]
\[ \text{bias}(S_n) < \text{bias}\left((p + 1)\hat{\Sigma}\right). \]

The estimates of root MSE for MLEs under Kotz type population also appear to be greater than that for MLEs under normal population with the following relationship:

\[ \text{root MSE}(\bar{x}) < \text{root MSE}(\hat{\mu}), \]
Table 2.3: The estimates of bias and root MSE for $\hat{\Sigma}$ and $\frac{1}{p+1}S_n$ and Pitman nearness probability of $\hat{\Sigma}$ relative to $\frac{1}{p+1}S_n$ for Kotz type samples.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$Bia\sigma(\hat{\Sigma})$</th>
<th>$Bia\sigma(\frac{1}{p+1}S_n)$</th>
<th>root MSE($\hat{\Sigma}$)</th>
<th>root MSE($\frac{1}{p+1}S_n$)</th>
<th>$PN(\hat{\Sigma}, \frac{1}{p+1}S_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.2173286</td>
<td>0.6136360</td>
<td>4.5729671</td>
<td>4.6180806</td>
<td>0.580</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.0594558</td>
<td>0.2972934</td>
<td>3.6564364</td>
<td>3.8758817</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0397885</td>
<td>0.2024529</td>
<td>2.8092322</td>
<td>3.0098940</td>
<td>0.595</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0423176</td>
<td>0.0496837</td>
<td>1.9725625</td>
<td>2.1130978</td>
<td>0.571</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>2.5227599</td>
<td>7.5841423</td>
<td>59.32383</td>
<td>59.885291</td>
<td>0.558</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1.3631267</td>
<td>4.8771157</td>
<td>47.940337</td>
<td>48.826692</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.8710543</td>
<td>1.4684184</td>
<td>37.83151</td>
<td>39.230765</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.6455298</td>
<td>1.4271476</td>
<td>25.851964</td>
<td>27.179129</td>
<td>0.616</td>
</tr>
</tbody>
</table>

Moreover, the estimates of Pitman Nearness probability of ML estimators under normal population relative to ML estimators under Kotz type population are greater than 0.5. Thus, we can conclude that the ML estimators under normal distribution do better than Kotz type ML estimators when the data are simulated from normal distribution.

Table 2.4: The estimates of bias and root MSE for $\hat{\mu}$ and $\bar{x}$ and Pitman nearness probability of $\bar{x}$ relative to $\hat{\mu}$ for normal samples.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$n$</th>
<th>$Bia\sigma(\hat{\mu})$</th>
<th>$Bia\sigma(\bar{x})$</th>
<th>root MSE($\hat{\mu}$)</th>
<th>root MSE($\bar{x}$)</th>
<th>$PN(\bar{x}, \hat{\mu})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>0.0419632</td>
<td>0.0250159</td>
<td>0.9201970</td>
<td>0.8175815</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.0235707</td>
<td>0.0146184</td>
<td>0.7436633</td>
<td>0.6521578</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0103899</td>
<td>0.0067315</td>
<td>0.5750868</td>
<td>0.5067936</td>
<td>0.620</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0036084</td>
<td>0.0033503</td>
<td>0.4029865</td>
<td>0.3635094</td>
<td>0.600</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>0.0685258</td>
<td>0.0337121</td>
<td>3.5255275</td>
<td>3.3563840</td>
<td>0.589</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.0653346</td>
<td>0.0331537</td>
<td>2.8187549</td>
<td>2.6684190</td>
<td>0.614</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0612400</td>
<td>0.0431314</td>
<td>2.1983321</td>
<td>2.0978606</td>
<td>0.580</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0313821</td>
<td>0.0227773</td>
<td>1.5524146</td>
<td>1.4771821</td>
<td>0.600</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 2.5: The estimates of bias and root MSE for \((p + 1) \Sigma\) and \(S_n\) and Pitman nearness probability of \(S_n\) relative to \((p + 1) \Sigma\) for normal samples.

<table>
<thead>
<tr>
<th>(p)</th>
<th>(n)</th>
<th>(\text{Bias}((p + 1) \Sigma))</th>
<th>(\text{Bias}(S_n))</th>
<th>(\text{root MSE}((p + 1) \Sigma))</th>
<th>(\text{root MSE}(S_n))</th>
<th>(PN(S_n, (p + 1) \Sigma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>1.5617102</td>
<td>0.4167722</td>
<td>4.3717427</td>
<td>3.2530058</td>
<td>0.599</td>
</tr>
<tr>
<td>30</td>
<td>1.4783830</td>
<td>0.3349251</td>
<td>3.7999676</td>
<td>2.8374782</td>
<td>0.602</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.7676229</td>
<td>0.1653396</td>
<td>3.2272657</td>
<td>2.1500709</td>
<td>0.654</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.7092333</td>
<td>0.1373975</td>
<td>2.5760868</td>
<td>1.5487331</td>
<td>0.711</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>7.6239726</td>
<td>6.9183341</td>
<td>56.9054100</td>
<td>49.532877</td>
<td>0.594</td>
</tr>
<tr>
<td>30</td>
<td>9.595013</td>
<td>4.6867674</td>
<td>48.4888900</td>
<td>41.584469</td>
<td>0.584</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>8.5435641</td>
<td>4.2156948</td>
<td>35.7407750</td>
<td>31.135813</td>
<td>0.589</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10.708864</td>
<td>1.7581084</td>
<td>28.4005060</td>
<td>23.347784</td>
<td>0.632</td>
<td></td>
</tr>
</tbody>
</table>

2.7 Concluding Remarks

We have considered one of the multivariate heavy-tailed distributions which has fatter tail regions than that of multivariate Normal distribution and the family of these distributions is called the Kotz type family. We have shown that the estimators of \(\mu\) obtained by maximizing the log-likelihood function of Kotz type distribution is same as the generalized spatial median of \(\mu\) introduced by Rao (1988). The estimators of \(\mu\) and \(\Sigma\) under a variety of covariance structures, namely the general covariance, AR(1), the equicorrelation and unstructured covariance structures are provided. Also we have provided a simulation algorithm for generating samples from Kotz type distribution by using the polar coordinate transformation and bisection methods. Next, we performed a simulation experiment to compare the ML estimates of \(\mu\) and \(\Sigma\) by using three measures, bias, root MSE and Pitman Nearness probability under Kotz type and Normal samples. Based on all the three criteria and using the results provided in Tables 2.2 - 2.5 we conclude that if we use normal ML estimates when the data are from Kotz type population we will lose out on efficiency, estimators will be more biased and further they will be inferior according to Pitman Nearness probability criteria as well.
CHAPTER III

STATISTICAL INFERENCE USING KOTZ TYPE DISTRIBUTION

3.1 Introduction

In this chapter we shall use the concepts and results from Chapter 2 to develop techniques for analyzing multivariate data. We consider statistical inferences about location parameter of the Kotz type distribution(s) under variety of covariance structures, including the general and AR(1) structured covariances. To save space however, here and in the subsequent chapter, we have presented the results only for the general and AR(1) structured covariances. In the following sections we shall discuss testing of hypotheses about one or more population parameters of the Kotz type distribution. We use the likelihood ratio method for the large samples to test the hypotheses. We also apply the asymptotic distribution of GSM for finding simultaneous confidence intervals. The results are illustrated using numerical examples.

3.2 Testing of Hypothesis for One Population Under the General Covariance Structure

We will provide the likelihood ratio test for testing of hypothesis on location parameter of one multivariate population under a general covariance structure. Likelihood ratio tests have several optimal properties for reasonably large samples and they are particularly convenient for hypothesis formulated in terms of multivariate parameters.

Let \( x_1, \ldots, x_n \) be a random sample from Kotz type distribution with the parameters \( \mu \) and \( \Sigma \). Let \( \theta = (\mu_1, \ldots, \mu_p, \sigma_{11}, \ldots, \sigma_{1p}, \ldots, \sigma_{p-1p}, \sigma_{pp})' \) be the vector of all unknown parameters and let \( L(\theta) \) be the likelihood function given the samples \( x_1, \ldots, x_n \). Suppose under the null hypothesis \( H_0 : \theta \in \Theta_0 \), where \( \Theta_0 \subset \Theta \) and \( \Theta \) is the parameter space.
Under $H_0$, let $\hat{\theta} = (\hat{\mu}_1, ..., \hat{\mu}_p, \hat{\sigma}_{11}, ..., \hat{\sigma}_{1p}, ..., \hat{\sigma}_{p-1,p}, \hat{\sigma}_{pp})'$ be the MLE of $\theta$. Then the maximum likelihood function under $H_0$ is given by

$$L(\hat{\theta}) = c^n |\hat{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\sum_{i=1}^{n} \sqrt{(x_i - \hat{\mu})' \hat{\Sigma}^{-1} (x_i - \hat{\mu})} \right\}.$$ 

Under $\theta \in \Theta$ (that is, under no restrictions on $\theta$), let $\hat{\theta} = (\hat{\mu}_1, ..., \hat{\mu}_p, \hat{\sigma}_{11}, ..., \hat{\sigma}_{1p}, ..., \hat{\sigma}_{p-1,p}, \hat{\sigma}_{pp})'$ be the MLE of $\theta$. Then the maximum likelihood function under no restriction is given by

$$L(\hat{\theta}) = c^n |\hat{\Sigma}|^{-\frac{n}{2}} \exp \left\{ -\sum_{i=1}^{n} \sqrt{(x_i - \hat{\mu})' \hat{\Sigma}^{-1} (x_i - \hat{\mu})} \right\}.$$ 

Then the likelihood ratio test of $H_0 : \theta \in \Theta_0$ rejects $H_0$ in favor of $H_1 : \theta \not\in \Theta_0$ if

$$\Lambda = \frac{L(\hat{\theta})}{L(\hat{\theta})} < c,$$

that is, if

$$\Lambda = \left[ \frac{|\hat{\Sigma}|}{|\Sigma|} \right]^{\frac{n}{2}} \exp \left\{ -\sum_{i=1}^{n} \sqrt{(x_i - \hat{\mu})' \hat{\Sigma}^{-1} (x_i - \hat{\mu})} - \sqrt{(x_i - \mu)' \Sigma^{-1} (x_i - \mu)} \right\} < c,$$

where $c$ is a suitably chosen constant.

When the sample size $n$ is large,

$$-2 \ln \Lambda = -2 \ln \left( \frac{L(\hat{\theta})}{L(\hat{\theta})} \right)$$

is approximately distributed as $\chi^2_r$ random variable, where the degrees of freedom, $r = (\text{dimension of } \Theta) - (\text{dimension of } \Theta_0)$.

**Example 3.1.** In the following, we illustrate this procedure for testing three specific hypotheses by taking the Board data considered in Chapter 2. The considered hypotheses are just for an illustration that various linear hypotheses like these can be easily tested using these methods and not for any specific relevance for the example in hand. For this data set $p = 4$, $n = 30$. The hypotheses and the results from testing are as follows:
(i) Test $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$.

Under $H_0$, the maximum likelihood estimate of $\mu$ is $\tilde{\mu} = 0$. The asymptotic distribution of $-2 \ln \Lambda \sim \chi^2_4$.

The test statistic $= -2 \ln \Lambda = 110.14185$. The P-value $= P[\chi^2_4 > 110.14185] < 0.0001$. Hence we reject $H_0$ and conclude that $\mu \neq 0$.

(ii) Suppose we want to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_4 = \gamma$ (say) vs. $H_1 : \mu$ is arbitrary. That is, test $H_0 : \mu = \gamma 1$ vs. $H_1 : \mu \neq \gamma 1$, where $1 = (1,\ldots,1)'$.

Under $H_0$, $\tilde{\mu} = \tilde{\gamma} 1$, where $\tilde{\gamma}$ is the MLE of $\gamma$.

The test statistic $= -2 \ln \Lambda = 54.197907$. Using $-2 \ln \Lambda \sim \chi^2_3$, we find that the P-value is $< 0.0001$. Hence we reject $H_0$.

(iii) Next for illustration, suppose we want to test that the last two components of $\mu$ are zero. That is, $H_0 : \mu = \left[ \begin{array}{c} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{array} \right]$ vs. $H_1 : \mu$ is arbitrary.

Under $H_0$, $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, 0, 0)'$ and the test statistic $= -2 \ln \Lambda = 178.72292$.

Using $-2 \ln \Lambda \sim \chi^2_2$, we note P-value $< 0.0001$. Hence once again we reject $H_0$.

3.3 Testing of Hypothesis for One Population Under AR(1) Structure

Let $x_1,\ldots,x_n$ be a random sample from Kotz type distribution with the parameters $\mu$ and $\Sigma$ and $\theta = (\mu_1,\ldots,\mu_p, \sigma, \rho)'$ be the vector of all unknown parameters and let $L(\theta)$ be the likelihood function. Suppose under the null hypothesis $H_0 : \theta \in \Theta_0$, where $\Theta_0 \subset \Theta$.

Under $H_0$, let $\tilde{\theta} = (\tilde{\mu}', \tilde{\sigma}, \tilde{\rho})'$ be the MLE of $\theta$. Then the maximum likelihood function is given by

$$L(\tilde{\theta}) = c^n \left[ \frac{(\tilde{\sigma}^2)^p}{1 - \tilde{\rho}^2} \right]^{\frac{n}{2}} \exp \left\{ -\frac{1}{\tilde{\sigma}^2} \sum_{i=1}^n \sqrt{(x_i - \tilde{\mu})' \left[ I_p + \tilde{\sigma}^2 C_1 - \tilde{\rho} C_2 \right] (x_i - \tilde{\mu})} \right\}$$

$$= c^n \left[ \frac{(\tilde{\sigma}^2)^p}{1 - \tilde{\rho}^2} \right]^{\frac{n}{2}} \exp \left\{ -\frac{1}{\tilde{\sigma}^2} \sum_{i=1}^n \sqrt{A_{ii} + \tilde{\rho}^2 A_{2i} - \tilde{\rho} A_{3i}} \right\},$$
where $C_1$ and $C_2$ are as defined in (2.5), (2.6) and

$$
A_{1i} = (x_i - \tilde{\mu})' (x_i - \tilde{\mu}),
$$

$$
A_{2i} = (x_i - \tilde{\mu})' C_1 (x_i - \tilde{\mu}),
$$

$$
A_{3i} = (x_i - \tilde{\mu})' C_2 (x_i - \tilde{\mu}).
$$

Under $\theta \in \Theta$, let $\tilde{\theta} = (\tilde{\mu}', \tilde{\sigma}, \tilde{\rho})'$ be the MLE of $\theta \in \Theta$. Then the maximum likelihood function is given by

$$
L(\tilde{\theta}) = c^n \left[ \frac{(\tilde{\sigma}^2)^p}{1 - \tilde{\rho}^2} \right]^{-\frac{n}{2}} \exp \left\{ -\frac{1}{\tilde{\sigma}^2} \sum_{i=1}^{n} \sqrt{(x_i - \tilde{\mu})' [I_p + \tilde{\rho}^2 C_1 - \tilde{\rho} C_2] (x_i - \tilde{\mu})} \right\}
$$

$$
= c^n \left[ \frac{(\tilde{\sigma}^2)^p}{1 - \tilde{\rho}^2} \right]^{-\frac{n}{2}} \exp \left\{ -\frac{1}{\tilde{\sigma}^2} \sum_{i=1}^{n} \sqrt{B_{1i} + \tilde{\rho}^2 B_{2i} - \tilde{\rho} B_{3i}} \right\},
$$

where

$$
B_{1i} = (x_i - \tilde{\mu})' (x_i - \tilde{\mu}),
$$

$$
B_{2i} = (x_i - \tilde{\mu})' C_1 (x_i - \tilde{\mu}),
$$

$$
B_{3i} = (x_i - \tilde{\mu})' C_2 (x_i - \tilde{\mu}).
$$

Then the likelihood ratio test of $H_0 : \theta \in \Theta_0$ rejects $H_0$ in favor of $H_1 : \theta \notin \Theta_0$ if

$$
\Lambda = \frac{L(\tilde{\theta})}{L(\theta)} < c,
$$

that is, if

$$
\Lambda = \left( \frac{\tilde{\sigma}^2}{\sigma^2} \right)^{\frac{n}{2}} \left( \frac{1 - \tilde{\rho}^2}{1 - \rho^2} \right)^{\frac{n}{2}} \exp \left\{ -\sum_{i=1}^{n} \left[ \frac{1}{2} \sqrt{A_{1i} + \tilde{\rho}^2 A_{2i} - \tilde{\rho} A_{3i}} - \frac{1}{2} \sqrt{B_{1i} + \tilde{\rho}^2 B_{2i} - \tilde{\rho} B_{3i}} \right] \right\} < c,
$$

where $c$ is a suitable constant.

When the sample size $n$ is large,

$$
-2 \ln \Lambda = -2 \ln \left( \frac{L(\tilde{\theta})}{L(\theta)} \right)
$$

is approximately distributed as $\chi^2_r$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
random variable, where the degrees of freedom, \( r = (\text{dimension of } \Theta) - (\text{dimension of } \Theta_0) \).

**Example 3.2.** In the following once again we illustrate this procedure for testing three specific hypotheses by taking the Board data. The hypotheses and the results from testing are as follows:

(i) Test \( H_0 : \mu = 0 \ vs. \ H_1 : \mu \neq 0 \).

Under \( H_0 \), the maximum likelihood estimate of \( \mu \) is \( \tilde{\mu} = 0 \) and the asymptotic distribution of \( -2 \ln \Lambda \sim \chi^2_r \).

The test statistic \( = -2 \ln \Lambda = 228.5621 \). The P-value \( = P[\chi^2_r > 228.5621] < 0.0001 \). Hence we reject \( H_0 \) and conclude that \( \mu \neq 0 \).

(ii) Next, we want to test \( H_0 : \mu_1 = \mu_2 = \ldots = \mu_4 = \gamma \) (say) vs. \( H_1 : \mu \) is arbitrary. That is, \( H_0 : \mu = \gamma \mathbf{1} \ vs. \ H_1 : \mu \neq \gamma \mathbf{1}, \) where \( \mathbf{1} = (1, \ldots, 1)' \).

Under \( H_0 \), \( \tilde{\mu} = \tilde{\gamma} \mathbf{1} \), where \( \tilde{\gamma} \) is the MLE of \( \gamma \).

The test statistic \( = -2 \ln \Lambda = 108.22776 \). Using \( -2 \ln \Lambda \sim \chi^2_3 \), we find that the P-value is \(< 0.0001 \) leading to rejection of \( H_0 \).

(iii) Next, suppose we want to test that the last two components of \( \mu \) are zero. That is, \( H_0 : \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix} \) vs. \( H_1 : \mu \) is arbitrary.

Under \( H_0 \), \( \tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, 0, 0)' \) and the test statistic \( = -2 \ln \Lambda = 209.20829 \).

Using \( -2 \ln \Lambda \sim \chi^2_2 \), we note P-value \(< 0.0001 \). Hence once again we reject \( H_0 \).

### 3.4 Testing of Hypothesis for Several Populations Under the General Covariance Structure

Suppose \( x_{i1}, x_{i2}, \ldots, x_{in_i} \) is a random sample of size \( n_i \) from Kotz type population with the parameters \( \mu_i \) and \( \Sigma_i, \ i = 1, \ldots, g \). The random samples from different \( g \) populations are assumed to be independent.

Let \( \mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})' \) and let \( \sigma_i = (\sigma_{i11}, \ldots, \sigma_{i1p}, \ldots, \sigma_{ip-1,p}, \sigma_{ipp})', \ i = 1, \ldots, g \).
Let $\theta = (\mu_1', ..., \mu_g', \sigma_1', ..., \sigma_g')'$ be the vector of all unknown parameters.

(i) Test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_g = \mu$ under the assumption $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$, that is, $\sigma_1 = \cdots = \sigma_g = \sigma$, where $\sigma = (\sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1, p}, \sigma_{pp})'$.

Under $H_0$, let $\tilde{\theta} = (\tilde{\mu}', \tilde{\sigma}')'$ be the MLE of $\theta$. Then the maximum likelihood function is given by

$$L(\hat{\theta}) = c^{(\Sigma_i^g n_i)}|\Sigma|^{-\frac{g n_i}{2}} \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \mu_i)' \Sigma^{-1} (x_{ij} - \mu_i)} \right\}.$$

Under $\theta \in \Theta$, let $\hat{\theta} = (\hat{\mu}_1', ..., \hat{\mu}_g', \hat{\Sigma}_i')'$ be the MLE of $\theta \in \Theta$, that is, under no restrictions. Then the maximum likelihood function is given by

$$L(\hat{\theta}) = c^{(\Sigma_i^g n_i)}|\Sigma|^{-\frac{g n_i}{2}} \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (x_{ij} - \hat{\mu}_i)} \right\}.$$

Then the likelihood ratio test of $H_0 : \theta \in \Theta_0$ rejects $H_0$ in favor of $H_1 : \theta \not\in \Theta_0$ if

$$\Lambda = \frac{L(\hat{\theta})}{L(\tilde{\theta})} < c,$$

that is, if

$$\Lambda = \left[ \frac{\sum_{i=1}^{g} \Sigma_i^{-\frac{g n_i}{2}}}{\Sigma} \right] \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left[ \sqrt{(x_{ij} - \mu_i)' \Sigma_i^{-1} (x_{ij} - \mu_i)} - \sqrt{(x_{ij} - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (x_{ij} - \hat{\mu}_i)} \right] \right\} < c.$$

(ii) Test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_g = \mu$ when $\Sigma_i$'s are different, that is, when $\sigma_i$'s are different.

Under $H_0$, let $\tilde{\theta} = (\tilde{\mu}', \tilde{\sigma}_1', ..., \tilde{\sigma}_g')'$ be the MLE of $\theta$. Then the maximum likelihood function is given by

$$L(\tilde{\theta}) = c^{(\Sigma_i^g n_i)} \prod_{i=1}^{g} |\Sigma_i|^{-\frac{n_i}{2}} \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \tilde{\mu}_i)' \Sigma_i^{-1} (x_{ij} - \tilde{\mu}_i)} \right\}.$$

Under no restrictions, that is, $\theta \in \Theta$, let $\hat{\theta} = (\tilde{\mu}', ..., \tilde{\mu}_g', \hat{\Sigma}_1', ..., \hat{\Sigma}_g')'$ be the MLE of $\theta$. Then the maximum likelihood function is given by

$$L(\hat{\theta}) = c^{(\Sigma_i^g n_i)} \prod_{i=1}^{g} |\Sigma_i|^{-\frac{n_i}{2}} \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (x_{ij} - \hat{\mu}_i)} \right\}.$$
Then the likelihood ratio test of \( H_0 : \theta \in \Theta_0 \) rejects \( H_0 \) in favor of \( H_1 : \theta \notin \Theta_0 \) if

\[
\Lambda = \frac{L(\tilde{\theta})}{L(\theta)} < c,
\]

that is, if

\[
\Lambda = \prod_{i=1}^{g} \left[ \frac{\hat{\Sigma}_i}{\hat{\Sigma}_i} \right]^{n_i} \exp \left\{ -\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left[ \sqrt{(x_{ij} - \tilde{\mu}_i)\hat{\Sigma}_i^{-1}(x_{ij} - \tilde{\mu}_i)} - \sqrt{(x_{ij} - \tilde{\mu})\hat{\Sigma}_i^{-1}(x_{ij} - \tilde{\mu})} \right] \right\} < c,
\]

where \( c \) is a suitably chosen constant.

When the sample size \( n \) is large,

\[
-2 \ln \Lambda = -2 \ln \left( \frac{L(\tilde{\theta})}{L(\theta)} \right)
\]

is approximately distributed as \( \chi_r^2 \) random variable, where the degrees of freedom, \( r = (\text{dimension of } \Theta) - (\text{dimension of } \Theta_0) \).

In the following, we illustrate the procedure for testing the equality of several population means using the Football helmet data given in Example 3.3. Before testing the equality of the means, we first test the equality of the variance covariance matrices. The hypotheses and the results from testing are provided in the example below.

**Example 3.3.** Data on three variables, \( x_1 = \) eye-to-top-of-head measurement, \( x_2 = \) ear-to-top-of-head measurement, and \( x_3 = \) jaw width are given for three groups of players, namely, high school football players, college football players, and nonfootball players. There are 30 observations in each group. The data are displayed in the following Table 3.1. The helmet data collected as part of a preliminary study of a possible link between football helmet design and neck injuries (see Rencher, 2002, Table 8.3, pp. 280 - 281). The hypotheses and the results from testing are as follows:

(i) Test \( H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma \).

The test statistic \( = -2 \ln \Lambda = 13.054878 \). Using \( -2 \ln \Lambda \sim \chi_{12}^2 \), the P-value = 0.3650646. Hence, we do not reject \( H_0 \) and conclude that the variance covariance matrices are the same for the three groups.
Table 3.1: Helmet data for football players (HS: high school, C: college and NF: non-football).

<table>
<thead>
<tr>
<th>HS</th>
<th>C</th>
<th>NF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x₁</td>
<td>x₂</td>
</tr>
<tr>
<td>12.5</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>13.5</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>15.5</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>13.5</td>
<td>15.5</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>13.5</td>
<td>14.5</td>
<td>12.5</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>14.5</td>
<td>11.5</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>12.5</td>
</tr>
<tr>
<td>14</td>
<td>14.5</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>13.5</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>14.5</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>12.5</td>
</tr>
<tr>
<td>12.5</td>
<td>14</td>
<td>12.5</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>14.5</td>
<td>14.5</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15.5</td>
<td>13.5</td>
</tr>
<tr>
<td>13</td>
<td>15.5</td>
<td>12.5</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>12.5</td>
</tr>
<tr>
<td>14.5</td>
<td>15.5</td>
<td>12.5</td>
</tr>
<tr>
<td>14.5</td>
<td>16.5</td>
<td>13.5</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>10.5</td>
</tr>
<tr>
<td>13.5</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>12.5</td>
<td>14.5</td>
<td>12.5</td>
</tr>
</tbody>
</table>
(ii) Next, we want to test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu$, given $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma$, that is, test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu$, given $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$,

where $\sigma = (\sigma_{11}, \ldots, \sigma_{1p}, \ldots, \sigma_{p-1,p}, \sigma_{pp})'$.

Under $H_0$, let $\hat{\theta} = (\hat{\mu}', \hat{\sigma}')'$ be the MLE of $\theta$. Then the maximum likelihood function is given by

$$ L(\hat{\theta}) = e^{(\Sigma_1^{-1} n_1)|\Sigma|^{-\frac{3}{2} n_1}} \exp \left\{ -\sum_{i=1}^{3} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \mu)^\prime \hat{\Sigma}^{-1} (x_{ij} - \mu)} \right\}. $$

Under $\theta \in \Theta$, let $\hat{\theta} = (\hat{\mu}_1', \hat{\mu}_2', \hat{\mu}_3', \hat{\Sigma}')'$ be the MLE of $\theta \in \Theta$. Then the maximum likelihood function is given by

$$ L(\hat{\theta}) = e^{(\Sigma_1^{-1} n_1)|\Sigma|^{-\frac{3}{2} n_1}} \exp \left\{ -\sum_{i=1}^{3} \sum_{j=1}^{n_i} \sqrt{(x_{ij} - \hat{\mu}_i)^\prime \hat{\Sigma}^{-1} (x_{ij} - \hat{\mu}_i)} \right\}. $$

Therefore, the resulting likelihood ratio test is to reject $H_0$ if

$$ \Lambda = \left[ \frac{\Sigma_1^{-1} n_1}{\Sigma} \right]^{\frac{3}{2} n_1} \exp \left\{ -\sum_{i=1}^{3} \sum_{j=1}^{n_i} \left[ \sqrt{(x_{ij} - \hat{\mu})^\prime \hat{\Sigma}^{-1} (x_{ij} - \hat{\mu})} - \sqrt{(x_{ij} - \hat{\mu}_i)^\prime \hat{\Sigma}^{-1} (x_{ij} - \hat{\mu}_i)} \right] \right\} < c, $$

where $c$ is a suitable constant.

When the sample size $n$ is large,

$$ -2 \ln \Lambda = -2 \ln \left( \frac{L(\hat{\theta})}{L(\hat{\theta})} \right) $$

is approximately distributed as $\chi^2_n$.

The test statistic $-2 \ln \Lambda = 91.70311$. The P-value $= P[\chi^2 > 91.70311] < 0.0001$.

Hence we reject $H_0$ and conclude that at least some $\mu_i's$ are different.

### 3.5 Testing of Hypothesis for Several Populations Under AR(1) Structure

Suppose $x_{i1}, x_{i2}, \ldots, x_{in_i}, i = 1, \ldots, g$ are random samples of size $n_i$ from $g$ independent Kotz type distributions with the parameters $\mu_i$ and $\Sigma_i, i = 1, \ldots, g$. We assume that $\Sigma_i's$ have AR(1) structures with different variance and correlation parameters.
\( \mu_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ig})', \ i = 1, \ldots, g. \) Let \( \theta = (\mu'_1, \ldots, \mu'_g, \sigma_1, \ldots, \sigma_g, \rho_1, \ldots, \rho_g)' \) be the vector of all unknown parameters.

(i) Test \( H_0 : \mu_1 = \mu_2 = \cdots = \mu_g = \mu \) when \( \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma, \) that is \( \sigma^2_1 = \cdots = \sigma^2_g = \sigma^2 \) and \( \rho_1 = \cdots = \rho_g = \rho. \)

Let \( \hat{\theta} = (\hat{\mu}'_1, \ldots, \hat{\mu}'_g, \hat{\sigma}, \hat{\rho})' \) be the unrestricted MLE of \( \theta \in \Theta. \) Then the maximum likelihood function is given by

\[
L(\hat{\theta}) = c(\Sigma^2 \cdot n_i) \left[ \frac{(\sigma^2)^p}{1 - \rho^2} \right]^{-\frac{\Sigma^2 n_i}{2}} \exp \left\{ -\frac{1}{\delta} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{D_{1,ij} + \rho^2 D_{2,ij} - \hat{\rho} D_{3,ij}} \right\},
\]

where

\[
D_{1,ij} = (x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i),
D_{2,ij} = (x_{ij} - \hat{\mu}_i)' C_1 (x_{ij} - \hat{\mu}_i),
D_{3,ij} = (x_{ij} - \hat{\mu}_i)' C_2 (x_{ij} - \hat{\mu}_i),
\]

and \( C_1 \) and \( C_2 \) are defined in (2.5), (2.6).

Under \( H_0, \) let \( \hat{\theta} = (\hat{\mu}', \hat{\sigma}, \hat{\rho})' \) be the MLE of \( \theta. \) Then the maximum likelihood function is given by

\[
L(\hat{\theta}) = c(\Sigma^2 \cdot n_i) \left[ \frac{(\sigma^2)^p}{1 - \rho^2} \right]^{-\frac{\Sigma^2 n_i}{2}} \exp \left\{ -\frac{1}{\delta} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{E_{1,ij} + \rho^2 E_{2,ij} - \hat{\rho} E_{3,ij}} \right\},
\]

where

\[
E_{1,ij} = (x_{ij} - \hat{\mu})(x_{ij} - \hat{\mu}),
E_{2,ij} = (x_{ij} - \hat{\mu}' C_1 (x_{ij} - \hat{\mu}),
E_{3,ij} = (x_{ij} - \hat{\mu}' C_2 (x_{ij} - \hat{\mu}).
\]

The resulting likelihood ratio test rejects \( H_0 \) if
\[
\Lambda = \left[ \frac{\sigma^2}{\rho^2} \right]^{\frac{n_i}{2}} \left[ \frac{1 - \rho^2}{1 - \rho_i^2} \right]^{\frac{n_i}{2}} \exp \left\{ - \frac{1}{\sigma^2} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_i} \sqrt{E_{1,ij} + \rho_i^2 E_{2,ij} - \rho E_{3,ij}} - \frac{1}{\sigma_i} \sqrt{D_{1,ij} + \rho_i^2 D_{2,ij} - \rho D_{3,ij}} \right] \right\} < c,
\]

where \( c \) is a suitable constant.

(ii) Test \( H_0 : \mu_1 = \mu_2 = \cdots = \mu_g = \mu \) when \( \sigma_i^2 \)'s and \( \rho_i \)'s are different.

Under \( \theta \in \Theta \), let \( \hat{\theta} = (\hat{\mu}_1, \ldots, \hat{\mu}_g, \hat{\sigma}_1, \ldots, \hat{\sigma}_g, \hat{\rho}_1, \ldots, \hat{\rho}_g)' \) be the MLE of \( \theta \). Then the maximum likelihood function is given by

\[
L(\hat{\theta}) = c^{\left( \sum_{i=1}^{g} n_i \right)} \prod_{i=1}^{g} \left[ \frac{\sigma^2}{\rho^2} \right]^{\frac{n_i}{2}} \exp \left\{ - \frac{1}{\sigma^2} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{E_{1,ij} + \rho_i^2 E_{2,ij} - \rho E_{3,ij}} \right\}.
\]

Under \( H_0 \), let \( \tilde{\theta} = (\tilde{\mu}_i, \tilde{\sigma}_1, \ldots, \tilde{\sigma}_g, \tilde{\rho}_1, \ldots, \tilde{\rho}_g)' \) be the MLE of \( \theta \). Then the maximum likelihood function is given by

\[
L(\hat{\theta}) = c^{\left( \sum_{i=1}^{g} n_i \right)} \prod_{i=1}^{g} \left[ \frac{\sigma^2}{\rho^2} \right]^{\frac{n_i}{2}} \exp \left\{ - \frac{1}{\sigma^2} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \sqrt{E_{1,ij} + \rho_i^2 E_{2,ij} - \rho_i E_{3,ij}} \right\}.
\]

Now the resulting likelihood ratio test rejects \( H_0 \) if

\[
\Lambda = \prod_{i=1}^{g} \left[ \left( \frac{\sigma_i^2}{\rho_i^2} \right) \left( \frac{1 - \rho_i^2}{1 - \rho^2} \right) \right]^{\frac{n_i}{2}} \exp \left\{ - \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left[ \frac{1}{\sigma_i^2} \sqrt{E_{1,ij} + \rho_i^2 E_{2,ij} - \rho E_{3,ij}} - \frac{1}{\sigma_i^2} \sqrt{D_{1,ij} + \rho_i^2 D_{2,ij} - \rho D_{3,ij}} \right] \right\} < c,
\]

where \( c \) is an appropriate constant.

When the sample size \( n \) is large,

\[
-2 \ln \Lambda = -2 \ln \left( \frac{L(\hat{\theta})}{L(\tilde{\theta})} \right) \text{ is approximately distributed as } \chi^2,
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
random variable, where the degrees of freedom, \( r = (\text{dimension of } \Theta) - (\text{dimension of } \Theta_0) \).

**Example 3.4.** In the following, we illustrate this procedure for testing the equality of several population means using the Football helmet data considered in Example 3.3. Before testing the equality of means, we first test the equality of the variance covariance matrices. The hypotheses and the results from testing are as follows:

(i) \( H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma \). That is, \( H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2 \) and \( \rho_1 = \rho_2 = \rho_3 = \rho \).

The asymptotic distribution of \( -2 \ln \Lambda \sim \chi^2_4 \). The test statistic \( -2 \ln \Lambda = 2.3775286 \). The P-value = 0.666692. Hence, we do not reject \( H_0 \) and conclude that a common AR(1) structure can be used for the three groups.

(ii) Next, we want to test \( H_0 : \mu_1 = \mu_2 = \mu_3 = \mu \), given \( \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma \).

The test statistic \( -2 \ln \Lambda = 153.46125 \). Using \( -2 \ln \Lambda \sim \chi^2_6 \), we note that P-value < 0.0001. Hence once again we reject \( H_0 \) and conclude that at least one \( \mu_i \) is different from the others.

### 3.6 Simultaneous Confidence Intervals for Linear Functions of a Single Mean Under the General Covariance Structure

If \( H_0 : \mu = 0 \) is rejected, it may be of interest to provide confidence intervals for the individual components or for certain linear functions of \( \mu \). Using the same arguments and derivations as in Huber (1981), Ducharme and Milasevic (1987), and Naik (1993), the asymptotic distribution of GSM, \( \hat{\mu} \), can be summarized in the following theorem.

**Theorem 3.1** Let \( x_1, \ldots, x_n \) be a random sample from Kotz type distribution with parameters \( \mu \) and \( \Sigma \) and \( \hat{\mu} \) and \( \hat{\Sigma} \) be the ML estimates of \( \mu \) and \( \Sigma \). Then

\[
\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{D} N(0, \hat{\Sigma}A^{-1}BA^{-1}\hat{\Sigma}),
\]
where

\[
B = E \left[ \frac{(x - \mu)(x - \mu)'}{(x - \mu)\Sigma^{-1}(x - \mu)} \right] \quad \text{and} \quad A = E \left[ \frac{1}{\sqrt{(x - \mu)\Sigma^{-1}(x - \mu)}} \left( \Sigma - \frac{(x - \mu)(x - \mu)'}{(x - \mu)\Sigma^{-1}(x - \mu)} \right) \right].
\]

Further \( B \) and \( A \) can be estimated by

\[
\hat{B} = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})(x_i - \hat{\mu})'}{(x_i - \hat{\mu})\hat{\Sigma}^{-1}(x_i - \hat{\mu})}, \quad \hat{A} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{\sqrt{(x_i - \hat{\mu})\hat{\Sigma}^{-1}(x_i - \hat{\mu})}} \left[ \hat{\Sigma} - \frac{(x_i - \hat{\mu})(x_i - \hat{\mu})'}{(x_i - \hat{\mu})\hat{\Sigma}^{-1}(x_i - \hat{\mu})} \right] \right\}.
\]

Using Theorem 3.1 we can summarize 100(1 - \( \alpha \))% confidence interval for any \( \mu_i \), where \( \mu = (\mu_1, ..., \mu_p)' \) as follows:

**Proposition 3.1** If \( \hat{\mu} \) and \( \hat{\Sigma} \) are the ML estimates of \( \mu \) and \( \Sigma \), then 100(1 - \( \alpha \))% confidence interval for \( \mu_i \) is given by

\[
\left( \hat{\mu}_i - z_{\alpha/2}\sqrt{\frac{\hat{\tau}_i}{n}}, \ \hat{\mu}_i + z_{\alpha/2}\sqrt{\frac{\hat{\tau}_i}{n}} \right), \quad i = 1, ..., p,
\]

where \( \hat{\tau}_i \) is the \( i \)th diagonal element of \( \hat{\Sigma} = \hat{\Sigma}^{-1}\hat{B}\hat{A}^{-1}\hat{\Sigma} \), and \( z_{\alpha/2} \) is the upper 100(1 - \( \alpha/2 \))th percentile of the standard normal distribution.

The following proposition provides simultaneous confidence intervals for a set of \( m \) linear functions of \( \mu \).

**Proposition 3.2** Using Theorem 3.1, the 100(1 - \( \alpha \))% Bonferroni simultaneous confidence intervals for \( m \) linear functions of \( \mu_i \)'s, are given by

\[
\left( a'_i\hat{\mu} - z_{\alpha/2m}\sqrt{\frac{a'_i\hat{\tau}a_i}{n}}, \ a'_i\hat{\mu} + z_{\alpha/2m}\sqrt{\frac{a'_i\hat{\tau}a_i}{n}} \right), \quad i = 1, ..., m
\]

where \( a_i \)'s are vectors of known constants and \( z_{\alpha/2m} \) is the upper 100(1 - \( \alpha/2m \))th percentile of a standard normal distribution.
Proof. For a fixed \( a \), \( a' \mu \) is a linear function of the \( \mu_i \)'s and using Theorem 3.1 we have the large sample distribution of \( a' \hat{\mu} \) as:

\[
a' \hat{\mu} \xrightarrow{D} N(a' \mu, a' \tau a).
\]

Then 100(1 - \( \alpha \))% Bonferroni simultaneous confidence intervals for the linear functions \( a_1' \mu, \ldots, a_m' \mu \) are clearly given by

\[
\left( a_i' \hat{\mu} - z_{\alpha/2m} \sqrt{\frac{\hat{\tau}_{ii}}{n}}, a_i' \hat{\mu} + z_{\alpha/2m} \sqrt{\frac{\hat{\tau}_{ii}}{n}} \right), \quad i = 1, \ldots, m.
\]

In the following we construct simultaneous confidence intervals for the components of the mean vector in Board data of Example 3.1.

(i) The 95% Bonferroni simultaneous confidence intervals for the individual \( \mu_i, i = 1, \ldots, 4 \) as given by (3.1) are:

\[
\left( \hat{\mu}_i - z_{0.05/4} \sqrt{\frac{\hat{\tau}_{ii}}{n}}, \hat{\mu}_i + z_{0.05/4} \sqrt{\frac{\hat{\tau}_{ii}}{n}} \right),
\]

where \( \hat{\tau}_{ii} \) is the \( i \)th diagonal element of \( \tau \) defined in Proposition 3.1. For our data, the 95% Bonferroni simultaneous confidence intervals for \( \mu_i, i = 1, \ldots, 4 \) are:

\[
\begin{align*}
\mu_1 &\in (1715.813, 1985.0188), \\
\mu_2 &\in (1565.0283, 1817.212), \\
\mu_3 &\in (1361.8197, 1600.7757), \\
\mu_4 &\in (1553.4839, 1817.3641).
\end{align*}
\]

(ii) Using Proposition 3.2, the 95% Bonferroni simultaneous confidence intervals for \( \mu_l - \mu_k, l < k = 1, \ldots, 4 \) are given with the choices \( a_1' = (1, -1, 0, 0), a_2' = (0, 1, -1, 0), \) and \( a_3' = (0, 0, 1, -1) \) as

\[
\begin{align*}
\mu_1 - \mu_2 &\in (100.17677, 218.41464), \\
\mu_2 - \mu_3 &\in (148.81812, 270.82676), \\
\mu_3 - \mu_4 &\in (-265.3922, -142.8604).
\end{align*}
\]
(iii) The 95% Bonferroni simultaneous confidence intervals for $\mu_3, \mu_4$ are given with the choices $\mathbf{a}'_1 = (0,0,1,0)$, and $\mathbf{a}'_2 = (0,0,0,1)$ as

$$
(1370.2547, 1592.3408),
(1562.7987, 1808.0494).
$$

### 3.7 Simultaneous Confidence Intervals Under AR(1) Structure

Using Proposition 3.2, the $100(1 - \alpha)\%$ Bonferroni simultaneous confidence intervals for $m$ linear functions of $\mu_i's$ are given by

$$
\left( \mathbf{a}'_i \hat{\mu} - z_{\alpha/2m} \sqrt{\frac{\mathbf{a}'_i \hat{\Sigma} \mathbf{a}_i}{n}}, \mathbf{a}'_i \hat{\mu} + z_{1-\alpha/2m} \sqrt{\frac{\mathbf{a}'_i \hat{\Sigma} \mathbf{a}_i}{n}} \right), \quad i = 1, \ldots, m,
$$

where $\mathbf{a}_i's$ are vectors of known constants, $z_{\alpha/2m}$ is the upper $100(1 - \alpha/2m)$ th percentile of a standard normal distribution, and $\hat{\Sigma} = \hat{\Sigma}^{-1} \hat{\Sigma}$ with

$$
\hat{\Sigma} = \frac{\hat{\sigma}^2}{1-\hat{\rho}^2} \mathbf{V}(\hat{\rho}) \quad \text{and} \quad \mathbf{V}(\hat{\rho}) = \hat{\rho}^{j-k}, \quad j,k = 1,\ldots,p.
$$

In the following examples, we illustrate finding the simultaneous confidence intervals for linear functions of $\mu_i's$ on the Board data considered in Example 3.2.

(i) The 95% Bonferroni simultaneous confidence intervals for the individual $\mu_i$, $i = 1, \ldots, 4$ using (3.2) with the successive choices $\mathbf{a}'_1 = (1,0,0,0)$, $\mathbf{a}'_2 = (0,1,0,0)$, $\mathbf{a}'_3 = (0,0,1,0)$, and $\mathbf{a}'_4 = (0,0,0,1)$ are given by

$$
\mu_1 \in (1740.2172, 1978.4096),
\mu_2 \in (1581.5249, 1807.0114),
\mu_3 \in (1379.4298, 1590.7865),
\mu_4 \in (1567.8883, 1809.2085).
$$

(ii) The 95% Bonferroni simultaneous confidence intervals for $\mu_l - \mu_k$, $l < k = 1, \ldots, 4$ are given with the choices $\mathbf{a}'_1 = (1,-1,0,0)$, $\mathbf{a}'_2 = (0,1,-1,0)$, and $\mathbf{a}'_3 = (0,0,1,-1)$ as

$$
\mu_1 - \mu_2 \in (99.01871, 231.07178),
$$
\[
\mu_2 - \mu_3 \in (139.92105, 278.39899),
\]
\[
\mu_3 - \mu_4 \in (-265.6623, -141.2182).
\]

(iii) The 95% Bonferroni simultaneous confidence intervals for \(\mu_3, \mu_4\) are given with the choices \(a'_1 = (0, 0, 1, 0)\), and \(a'_2 = (0, 0, 0, 1)\) as

\[
(1386.8905, 1583.3258),
\]
\[
(1576.4067, 1800.6901).
\]

### 3.8 Simultaneous Confidence Intervals for Several Population Means

Let \(x_{i1}, x_{i2}, \ldots, x_{im_i}, i = 1, \ldots, g\), be \(g\) independent random samples of size \(n_i\) each from Kotz type distributions with parameters \(\mu_i\) and \(\Sigma_i, i = 1, \ldots, g\). Suppose the tests have revealed that a significant difference exists between the population means. In order to pinpoint the differences we construct simultaneous confidence intervals on various contrasts of difference between any two mean vectors. The following proposition provides distributional results that enable constructing simultaneous confidence intervals for linear combinations of \(\mu_j\)'s.

(i) Assume that \(\Sigma_1 = \cdots = \Sigma_g = \Sigma\), that is, \(\sigma_1 = \cdots = \sigma_g = \sigma\), where 

\[
\sigma = (\sigma_{11}, \ldots, \sigma_{1p}, \ldots, \sigma_{p-1,p}, \sigma_{pp})'.
\]

**Proposition 3.3.** Let \(x_{i1}, x_{i2}, \ldots, x_{im_i}\) be a random sample of size \(n_i\) from Kotz type distribution with parameters \(\mu_i\) and \(\Sigma_i, i = 1, \ldots, g\). The random samples from different \(g\) populations are independent. We assume that \(\Sigma_1 = \cdots = \Sigma_g = \Sigma\). Using Theorem 3.1 and Proposition 3.2, the 100(1 - \(\alpha\))% Bonferroni simultaneous confidence intervals for \(m\) linear combinations of \(\mu_l - \mu_{l'}\), \(l < l' = 1, \ldots, g\) are given by

\[
a'_k (\hat{\mu}_l - \hat{\mu}_{l'}) \pm z_{\alpha/2m} \sqrt{\alpha'_k \left(\frac{1}{n_l} \hat{\tau}_l + \frac{1}{n_{l'}} \hat{\tau}_{l'}\right) a'_k}, \quad k = 1, \ldots, m \tag{3.3}
\]

where \(a'_k\)'s are vectors of known constants, \(z_{\alpha/2m}\) is the upper 100(1 - \(\alpha/2m\))th percentile of the standard normal distribution, \(\hat{\tau}_i = \hat{\Sigma} \hat{A}_i^{-1} \hat{B}_i \hat{A}_i^{-1} \hat{\Sigma}, i = l, l' = 1, \ldots, g\).
and

\[ \hat{B}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{(x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i)'}{\hat{\Sigma}^{-1}(x_{ij} - \hat{\mu}_i)}, \]

\[ \hat{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ \frac{1}{\sqrt{(x_{ij} - \hat{\mu}_i)'\hat{\Sigma}^{-1}(x_{ij} - \hat{\mu}_i)}} \left[ \hat{\Sigma} - \frac{(x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i)'}{(x_{ij} - \hat{\mu}_i)'\hat{\Sigma}^{-1}(x_{ij} - \hat{\mu}_i)} \right] \right\}. \]

**Proof.** For a fixed \( a, a'\mu_i \) is a linear function of the \( \mu_{ij} \) and using Theorem 3.1 and Proposition 3.2 we have

\[ a'(\hat{\mu}_i - \hat{\mu}_{i'}) \xrightarrow{D} N \left( a'(\mu_i - \mu_{i'}), a' \left( \frac{1}{n_i} \hat{\tau}_i + \frac{1}{n_{i'}} \hat{\tau}_{i'} \right) a \right) \]

Then 100(1 - \( \alpha \))% Bonferroni simultaneous confidence intervals for the linear functions \( a_k'(\mu_i - \mu_{i'}) \) are clearly given by

\[ a_k'(\hat{\mu}_i - \hat{\mu}_{i'}) \pm z_{\alpha / 2m} \sqrt{a_k'(\frac{1}{n_i} \hat{\tau}_i + \frac{1}{n_{i'}} \hat{\tau}_{i'}) a_k}, \quad k = 1, \ldots, m, \quad l < l' = 1, \ldots, g. \]

(ii) Assume that \( \Sigma_i \)s are different, that is \( \sigma_i \)s are different.

**Proposition 3.4.** Let \( x_{i1}, x_{i2}, \ldots, x_{in_i} \) be a random sample of size \( n_i \) from Kotz type distribution with \( \mu_i \) and \( \Sigma_i \), \( i = 1, \ldots, g \). The random samples from different \( g \) populations are independent. We assume that \( \Sigma_i \)s are different. Using Theorem 3.1 and Proposition 3.2, the 100(1 - \( \alpha \))% Bonferroni simultaneous confidence intervals for \( m \) linear combinations of \( \mu_i - \mu_{i'} \), \( l < l' = 1, \ldots, g \) are given by

\[ a_k'(\hat{\mu}_i - \hat{\mu}_{i'}) \pm z_{\alpha / 2m} \sqrt{a_k'(\frac{1}{n_i} \hat{\tau}_i + \frac{1}{n_{i'}} \hat{\tau}_{i'}) a_k}, \quad k = 1, \ldots, m \]

(3.4)

where \( a_k \)s are vectors of known constants, \( z_{\alpha / 2m} \) is the upper \( 100(1 - \alpha / 2m) \)th percentile of the standard normal distribution, \( \hat{\tau}_i = \hat{\Sigma}_i \hat{A}_i^{-1} \hat{B}_i \hat{A}_i^{-1} \hat{\Sigma}_i, \quad i = l, l' = 1, \ldots, g \) and

\[ \hat{B}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{(x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i)'}{\hat{\Sigma}_i^{-1}(x_{ij} - \hat{\mu}_i)}, \]

\[ \hat{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ \frac{1}{\sqrt{(x_{ij} - \hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x_{ij} - \hat{\mu}_i)}} \left[ \hat{\Sigma}_i - \frac{(x_{ij} - \hat{\mu}_i)(x_{ij} - \hat{\mu}_i)'}{(x_{ij} - \hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x_{ij} - \hat{\mu}_i)} \right] \right\}. \]
Since we reject the hypothesis of equality of means in Example 3.3, Section 3.4, we want to find simultaneous confidence intervals for linear functions of \( \mu_i - \mu_{i'} \), \( l < l' = 1, ..., 3 \). Let \( \mu_i = (\mu_{i1}, ..., \mu_{ip})' \), \( i = 1, ..., 3 \). Then with the choices, \( a_1' = (1, 0, 0) \), \( a_2' = (0, 1, 0) \), and \( a_3' = (0, 0, 1) \) using (3.3), the 95% Bonferroni simultaneous confidence intervals for \( \mu_{ij} - \mu_{i j} \), \( j = 1, ..., 3 \) are:

\[
\begin{align*}
\mu_{11} - \mu_{21} & \in (2.2932274, 3.9341077), \\
\mu_{12} - \mu_{22} & \in (0.4294995, 2.0009086), \\
\mu_{13} - \mu_{23} & \in (-0.212608, 0.923095). \\
\end{align*}
\]

The 95% Bonferroni simultaneous confidence intervals for \( \mu_{1j} - \mu_{3j} \), \( j = 1, ..., 3 \) are:

\[
\begin{align*}
\mu_{11} - \mu_{31} & \in (1.3875155, 3.1273591), \\
\mu_{12} - \mu_{32} & \in (0.3040446, 1.6128665), \\
\mu_{13} - \mu_{33} & \in (-0.000264, 1.0659512). \\
\end{align*}
\]

The 95% Bonferroni simultaneous confidence intervals for \( \mu_{2j} - \mu_{3j} \), \( j = 1, ..., 3 \) are:

\[
\begin{align*}
\mu_{21} - \mu_{31} & \in (-1.709631, -0.002829), \\
\mu_{22} - \mu_{32} & \in (-0.835963, 0.3224655), \\
\mu_{23} - \mu_{33} & \in (-0.374984, 0.7301842). \\
\end{align*}
\]

3.9 Simultaneous Confidence Intervals for Several Population Means Under AR(1) structure

As before, suppose we have independent random samples from \( g \) Kotz type distributions with parameters \( \mu_i \) and \( \Sigma_i \). We assume that \( \Sigma_i \)'s have AR(1) structures.

(i) Assume that \( \Sigma_1 = \cdots = \Sigma_g = \Sigma \), that is, \( \sigma_1^2 = \cdots = \sigma_g^2 = \sigma^2 \) and \( \rho_1 = \cdots = \rho_g = \rho \).

Using Proposition 3.3, the 100\((1 - \alpha)\)% Bonferroni simultaneous confidence intervals for the linear functions of \( a_k' (\mu_l - \mu_{i'}) \), \( l < l' = 1, ..., g \), are given by

\[
a_k' (\hat{\mu}_l - \hat{\mu}_{i'}) \pm z_{\alpha/2m} \sqrt{a_k' \left( \frac{1}{n_l} \tilde{\tau}_l + \frac{1}{n_{i'}} \tilde{\tau}_{i'} \right) a_k}, \quad k = 1, ..., m \tag{3.5}
\]
where \( \mathbf{a}_k \)s are vectors of known constants, \( z_{\alpha/2m} \) is the upper \( 100(1 - \alpha/2m) \)th percentile of the standard normal distribution, \( \hat{\tau}_i = \hat{\Sigma}_i^{-1} \hat{\mathbf{B}}_i \hat{\mathbf{A}}_i^{-1} \hat{\Sigma}_i \), \( i = l, l' = 1, ..., g \) and

\[
\hat{\mathbf{B}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle \langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle'}{(\mathbf{x}_{ij} - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)},
\]

\[
\hat{\mathbf{A}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ \frac{1}{\sqrt{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)}} \left[ \hat{\Sigma}_i - \frac{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle \langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle'}{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)} \right] \right\}.
\]

(ii) **Assume that \( \Sigma_i \)'s are different**, that is \( \sigma_i^2 \)'s and \( \rho_i \)'s are different.

Using Proposition 3.4, the \( 100(1 - \alpha) \)% Bonferroni simultaneous confidence intervals for the linear functions of \( \mathbf{a}_k' \langle \mu_l - \mu_{l'} \rangle \), \( l < l' = 1, ..., g \), are given by

\[
a_k' \langle \hat{\mu}_l - \hat{\mu}_{l'} \rangle \pm z_{\alpha/2m} \sqrt{a_k' \left( \frac{1}{n_l} \hat{\tau}_i + \frac{1}{n_{l'}} \hat{\tau}_i \right) a_k}, \quad k = 1, ..., m \tag{3.6}
\]

where \( \mathbf{a}_k \)'s are vectors of known constants, \( z_{\alpha/2m} \) is the upper \( 100(1 - \alpha/2m) \)th percentile of the standard normal distribution, \( \hat{\tau}_i = \hat{\Sigma}_i \hat{\mathbf{A}}_i^{-1} \hat{\mathbf{B}}_i \hat{\mathbf{A}}_i^{-1} \hat{\Sigma}_i \), \( i = l, l' = 1, ..., g \) and

\[
\hat{\mathbf{B}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle \langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle'}{(\mathbf{x}_{ij} - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)},
\]

\[
\hat{\mathbf{A}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ \frac{1}{\sqrt{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)}} \left[ \hat{\Sigma}_i - \frac{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle \langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle'}{\langle \mathbf{x}_{ij} - \hat{\mu}_i \rangle' \hat{\Sigma}_i^{-1} (\mathbf{x}_{ij} - \hat{\mu}_i)} \right] \right\}.
\]

In Example 3.4, we rejected the hypothesis of equality of means. In the following example, we want to find simultaneous confidence interval for linear functions of \( \mu_{i} - \mu_{l'} \), \( l < l' = 1, ..., 3 \). With the choices \( \mathbf{a}_1' = (1, 0, 0) \), \( \mathbf{a}_2' = (0, 1, 0) \), and \( \mathbf{a}_3' = (0, 0, 1) \) using (3.5), the \( 95\% \) Bonferroni simultaneous confidence intervals for \( \mu_{ij} - \mu_{ij} \), \( j = 1, ..., 3 \) are:

\[
\mu_{11} - \mu_{21} \in (2.3546585, 3.9408238),
\]
\[
\mu_{12} - \mu_{22} \in (0.4765961, 2.0264683),
\]
\[
\mu_{13} - \mu_{23} \in (-0.216871, 0.9422181).
\]
The 95% Bonferroni simultaneous confidence intervals for $\mu_{1j} - \mu_{3j}$, $j = 1, ..., 3$ are:

$$\begin{align*}
\mu_{11} - \mu_{31} & \in (1.3531563, 3.2021812), \\
\mu_{12} - \mu_{32} & \in (0.3352931, 1.6503519), \\
\mu_{13} - \mu_{33} & \in (0.0230982, 1.0618578).
\end{align*}$$

The 95% Bonferroni simultaneous confidence intervals for $\mu_{2j} - \mu_{3j}$, $j = 1, ..., 3$ are:

$$\begin{align*}
\mu_{21} - \mu_{31} & \in (-1.756473, 0.0163278), \\
\mu_{22} - \mu_{32} & \in (-0.808331, 0.2909115), \\
\mu_{23} - \mu_{33} & \in (-0.377271, 0.7368803).
\end{align*}$$

### 3.10 Concluding Remarks

We have considered inferences about location parameters of the Kotz type distribution(s) under the general covariance and AR(1) structures. We have discussed testing of hypotheses about one or more population parameters of the Kotz type distribution using the likelihood ratio test for the large samples. We have also applied the asymptotic distribution of GSM for finding simultaneous confidence intervals for linear combinations of $\mu_{ij}$'s.
CHAPTER IV

DISCRIMINATION AND CLASSIFICATION ANALYSIS
USING KOTZ TYPE DISTRIBUTION

4.1 Introduction

The statistical methods for estimation, hypothesis testing, and simultaneous confidence intervals under Kotz type distribution have been shown in the preceding chapters. In this chapter, we shall consider this distribution for determining classification rules which minimize the expected cost of misclassification under equal prior probabilities and equal misclassification costs. The procedures we develop will be same, in spirit, as the multivariate normal distribution based rules. In fact, the rule under common variance covariance matrix is same as the normal based method except that the parameter estimates used are the ML estimates of mean and variance under the Kotz type distribution. We will show sample classification procedures under the general covariance and AR(1) structures. Computation of misclassification probabilities using cross validation are also provided.

4.2 Separation and Classification for Two Populations

Let $f_1(x)$ and $f_2(x)$ be the probability density functions of a $p \times 1$ random vector $X$ under populations $\Pi_1$ and $\Pi_2$ respectively. Let $\Omega$ be the sample space and $R_1 \subset \Omega$ be the set of $x$ values for which we classify objects into $\Pi_1$ and $R_2 = \Omega - R_1$ be the remaining $x$ values for which we classify objects into $\Pi_2$.

The conditional probability $P(j/i)$ of classifying an object into $\Pi_j$ when it is from $\Pi_i$ is
\[ P(j/i) = P(\mathbf{X} \in R_j/\Pi_i) = \int_{R_j} f_i(x) dx, \quad \text{with } P(i/i) = 1 - \sum_{j=1}^{q} P(j/i), \quad i, j = 1, 2 \text{ and } i \neq j. \]

Let \( p_i \) be the prior probability of \( \Pi_i \) such that \( p_1 + p_2 = 1 \).

\[ P(\text{the object is correctly classified into } \Pi_i) = P(\text{the object comes from } \Pi_i \text{ and is correctly classified into } \Pi_i) \]
\[ = P(\mathbf{X} \in R_i/\Pi_i)P(\Pi_i) \]
\[ = P(i/i)p_i, \quad i = 1, 2. \]

\[ P(\text{the object is misclassified to } \Pi_i) = P(\text{the object comes from } \Pi_j \text{ and is misclassified to } \Pi_i) \]
\[ = P(\mathbf{X} \in R_i/\Pi_j)P(\Pi_j) \]
\[ = P(i/j)p_j, \quad i, j = 1, 2 \text{ and } i \neq j. \]

The costs of misclassification can be defined by a cost matrix

<table>
<thead>
<tr>
<th></th>
<th>( \Pi_1 )</th>
<th>( \Pi_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pi_1 )</td>
<td>0</td>
<td>( c(2/1) )</td>
</tr>
<tr>
<td>( \Pi_2 )</td>
<td>( c(1/2) )</td>
<td>0</td>
</tr>
</tbody>
</table>

We shall consider the expected cost of misclassification (ECM) given by

\[ \text{ECM} = c(1/2)P(1/2)p_2 + c(2/1)P(2/1)p_1. \]

**Result 4.1** The regions \( R_1 \) and \( R_2 \) that minimize the ECM are defined by the values \( x \) for which the following inequalities hold, that is,

\[ R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{c(1/2)}{c(2/1)} p_2 \quad \text{and } R_2 : \text{otherwise.} \]

Under equal prior probabilities and equal misclassification costs, these regions reduce to

\[ R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \quad \text{and } R_2 : \text{otherwise.} \quad (4.1) \]
4.2.1 Classification Rules Under Kotz Type Distributions and General Covariance Structure

We now assume that $f_1(x)$ and $f_2(x)$ are Kotz type densities with parameters $\mu_1$, $\Sigma_1$ and $\mu_2$, $\Sigma_2$ respectively. We consider two cases based on whether the variance covariance matrices are equal or not. First we shall consider the case of equal variance covariance matrices. That is,

**Case 1.** $\Sigma_1 = \Sigma_2 = \Sigma$

Suppose the Kotz type densities of $X$ for populations $\Pi_1$ and $\Pi_2$ are given by

$$f_i(x) = c |\Sigma|^{-\frac{1}{2}} \exp \{-[(x - \mu_i)^\top \Sigma^{-1}(x - \mu_i)]^{\frac{1}{2}}\}, \quad \mu_i \in \mathbb{R}^p, \Sigma p.d., i = 1, 2,$$

where $c = \frac{\Gamma(p)}{2\pi^{\frac{p}{2}}\Gamma(p)}$.

Suppose $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$ are two random samples of sizes $n_1$ and $n_2$ from Kotz type populations, $\Pi_1$ and $\Pi_2$ with unknown parameters $\mu_1$ and $\mu_2$ and a positive definite common variance covariance matrix $(p + 1)\Sigma$, respectively, with the conditions $n_1 + n_2 - 2 \geq p$. Then the minimum ECM regions determined by samples of sizes $n_1$ and $n_2$ respectively from (4.1) are simplified to

$$\hat{R}_1 : \exp \{-\sqrt{(x - \hat{\mu}_1)^\top \hat{\Sigma}^{-1}(x - \hat{\mu}_1)} + \sqrt{(x - \hat{\mu}_2)^\top \hat{\Sigma}^{-1}(x - \hat{\mu}_2)}\} \geq 1$$

and $\hat{R}_2 :$ otherwise.

Given these regions $\hat{R}_1$ and $\hat{R}_2$, the sample classification rule proceeds as in the following result.

**Result 4.2** Let the populations $\Pi_1$ and $\Pi_2$ be Kotz type populations with equal general covariance matrices. Then the linear classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$, a new observation vector to be classified, to $\Pi_1$ if

$$(\hat{\mu}_1 - \hat{\mu}_2)^\top \hat{\Sigma}^{-1} x_0 - \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)^\top \hat{\Sigma}^{-1}(\hat{\mu}_1 + \hat{\mu}_2) \geq 0, \quad (4.2)$$

where $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\Sigma}$ are MLEs of $\mu_1$, $\mu_2$ and $\Sigma$ respectively.

Allocate $x_0$ to $\Pi_2$ otherwise.
We note that this rule is similar to the rule based on the normal densities except that here we estimate $\mu_1$, $\mu_2$ and $\Sigma$ differently; that is, by the optimization method mentioned in Chapter 2.

**Case 2. $\Sigma_1 \neq \Sigma_2$**

Suppose the Kotz type densities of $X$ for populations $\Pi_1$ and $\Pi_2$ are given by

$$f_i(x) = c / |\Sigma_i|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)'\Sigma_i^{-1}(x - \mu_i) \right\}, \quad \mu_i \in \mathbb{R}^p, \Sigma_i \text{ p. d., } i = 1, 2,$$

where $c = \frac{\Gamma(p)}{2\pi^{\frac{p}{2}}\Gamma(p)}$.

Let $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$ be two random samples of sizes $n_1$ and $n_2$ from Kotz type populations $\Pi_1$ and $\Pi_2$ with unknown parameters $\mu_1, \Sigma_1$ and $\mu_2, \Sigma_2$, respectively, with $n_1 - 1 \geq p$ and $n_2 - 1 \geq p$. Then the minimum ECM regions determined by samples of sizes $n_1$ and $n_2$ respectively from (4.1) are simplified to

$$\hat{R}_1 : \left( \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_2|} \right)^{\frac{1}{2}} \exp \left\{ -\sqrt{(x - \hat{\mu}_1)'\hat{\Sigma}_1^{-1}(x - \hat{\mu}_1)} + \sqrt{(x - \hat{\mu}_2)'\hat{\Sigma}_2^{-1}(x - \hat{\mu}_2)} \right\} \geq 1$$

$$\hat{R}_2 : \text{otherwise.}$$

Given the regions $\hat{R}_1$ and $\hat{R}_2$, we can construct the sample classification rule as given in the following result.

**Result 4.3** Let the populations $\Pi_1$ and $\Pi_2$ be Kotz type populations with unequal general covariance matrices. Then the non-linear classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$ to $\Pi_1$ if

$$\sqrt{(x_0 - \hat{\mu}_2)'\hat{\Sigma}_2^{-1}(x_0 - \hat{\mu}_2)} - \sqrt{(x_0 - \hat{\mu}_1)'\hat{\Sigma}_1^{-1}(x_0 - \hat{\mu}_1)} - \frac{1}{2} \ln \left( \frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_2|} \right) \geq 0,$$

(4.3)

that is, if

$$\sqrt{(x_0 - \hat{\mu}_1)'\hat{\Sigma}_1^{-1}(x_0 - \hat{\mu}_1)} + \frac{1}{2} \ln |\hat{\Sigma}_1| \leq \sqrt{(x_0 - \hat{\mu}_2)'\hat{\Sigma}_2^{-1}(x_0 - \hat{\mu}_2)} + \frac{1}{2} \ln |\hat{\Sigma}_2|,$$

where $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are MLEs of $\mu_1, \mu_2, \Sigma_1$ and $\Sigma_2$ respectively.

Allocate $x_0$ to $\Pi_2$ otherwise.
**Example 4.1.** For illustration, here we consider Fisher's famous Iris data. The two species of Irises we have selected are Iris Versicolor and Iris Virginica and the data are displayed in Table 4.1. Data on four variables, $x_1 = \text{sepal length}$, $x_2 = \text{sepal width}$, $x_3 = \text{petal length}$ and $x_4 = \text{petal width}$ are available. There are 50 observations in each sample. These data have appeared in many places, but we have taken the data from Table 11.5 of Johnson and Wichern (1998, pp. 715 - 716).

These data were found to have unequal covariance matrices and hence we used Case 2 to illustrate our procedure. Next, we computed the MLEs of $\mu_1, \mu_2, \Sigma_1$ and $\Sigma_2$ under Kotz type distributions. The results are as follows:

$\Pi_1$ (Iris Versicolor) : $n_1 = 50$,

\[
\hat{\mu}_1 = \begin{pmatrix} 5.9301907, & 2.7846629, & 4.2389524, & 1.3164927 \end{pmatrix}',
\]

\[
\hat{\Sigma}_1 = \begin{pmatrix}
0.0651039 & 0.0220169 & 0.0426881 & 0.0135329 \\
0.0220169 & 0.022449 & 0.0206865 & 0.0101087 \\
0.0426881 & 0.0206865 & 0.0489653 & 0.0167524 \\
0.0135329 & 0.0101087 & 0.0167524 & 0.0088665
\end{pmatrix},
\]

and $\hat{\Sigma}_1^{-1} = \begin{pmatrix}
-18.48802 & 101.38476 & 8.9867497 & -104.3503 \\
-37.56574 & 8.9867497 & 94.079695 & -130.6641 \\
31.451537 & -104.3503 & -130.6641 & 430.62776
\end{pmatrix}^{-1}$.

$\Pi_2$ (Iris Virginica) : $n_2 = 50$,

\[
\hat{\mu}_2 = \begin{pmatrix} 6.5506952, & 2.9737207, & 5.5073221, & 2.0322709 \end{pmatrix}',
\]

\[
\hat{\Sigma}_2 = \begin{pmatrix}
0.0836885 & 0.0193137 & 0.0613979 & 0.0109602 \\
0.0193137 & 0.021707 & 0.0147842 & 0.0108312 \\
0.0613979 & 0.0147842 & 0.0625200 & 0.0110904 \\
0.0109602 & 0.0108312 & 0.0110904 & 0.0175368
\end{pmatrix},
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 4.1: Iris data: Iris Versicolor and Iris Virginica.

<table>
<thead>
<tr>
<th></th>
<th>Sepal length</th>
<th>Sepal width</th>
<th>Petal Length</th>
<th>Petal width</th>
<th>Sepal length</th>
<th>Sepal width</th>
<th>Petal Length</th>
<th>Petal width</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>3.2</td>
<td>4.7</td>
<td>1.4</td>
<td>6.3</td>
<td>3.3</td>
<td>6.9</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>3.2</td>
<td>4.5</td>
<td>1.5</td>
<td>5.8</td>
<td>2.7</td>
<td>5.1</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>3.1</td>
<td>4.9</td>
<td>1.5</td>
<td>7.1</td>
<td>3.0</td>
<td>5.9</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>2.3</td>
<td>4.0</td>
<td>1.3</td>
<td>6.3</td>
<td>2.9</td>
<td>5.6</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>2.8</td>
<td>4.6</td>
<td>1.5</td>
<td>6.5</td>
<td>3.0</td>
<td>5.8</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>2.8</td>
<td>4.5</td>
<td>1.3</td>
<td>7.6</td>
<td>3.0</td>
<td>6.6</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>3.3</td>
<td>4.7</td>
<td>1.6</td>
<td>4.9</td>
<td>2.5</td>
<td>4.5</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>4.9</td>
<td>2.4</td>
<td>3.3</td>
<td>1.0</td>
<td>7.3</td>
<td>2.9</td>
<td>6.3</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>2.9</td>
<td>4.6</td>
<td>1.3</td>
<td>6.7</td>
<td>2.5</td>
<td>5.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>2.7</td>
<td>3.9</td>
<td>1.4</td>
<td>7.2</td>
<td>3.6</td>
<td>6.1</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>2.0</td>
<td>3.5</td>
<td>1.0</td>
<td>6.5</td>
<td>3.2</td>
<td>5.1</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>3.0</td>
<td>4.2</td>
<td>1.5</td>
<td>6.4</td>
<td>2.7</td>
<td>5.3</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>2.2</td>
<td>4.0</td>
<td>1.0</td>
<td>6.8</td>
<td>3.0</td>
<td>5.5</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>2.9</td>
<td>4.7</td>
<td>1.4</td>
<td>5.7</td>
<td>2.5</td>
<td>5.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>2.9</td>
<td>3.6</td>
<td>1.3</td>
<td>5.8</td>
<td>2.8</td>
<td>5.1</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>3.1</td>
<td>4.4</td>
<td>1.4</td>
<td>6.4</td>
<td>3.2</td>
<td>5.3</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>3.0</td>
<td>4.5</td>
<td>1.5</td>
<td>6.5</td>
<td>3.0</td>
<td>5.5</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>2.7</td>
<td>4.1</td>
<td>1.0</td>
<td>7.7</td>
<td>3.8</td>
<td>6.7</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>2.2</td>
<td>4.5</td>
<td>1.5</td>
<td>7.7</td>
<td>2.6</td>
<td>6.9</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>2.5</td>
<td>3.9</td>
<td>1.1</td>
<td>6.0</td>
<td>2.2</td>
<td>5.0</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>3.2</td>
<td>4.8</td>
<td>1.8</td>
<td>6.9</td>
<td>3.2</td>
<td>5.7</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>2.8</td>
<td>4.0</td>
<td>1.3</td>
<td>5.6</td>
<td>2.8</td>
<td>4.9</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>2.5</td>
<td>4.9</td>
<td>1.5</td>
<td>7.7</td>
<td>2.8</td>
<td>6.7</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>2.8</td>
<td>4.7</td>
<td>1.2</td>
<td>6.3</td>
<td>2.7</td>
<td>4.9</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>2.9</td>
<td>4.3</td>
<td>1.3</td>
<td>6.7</td>
<td>3.3</td>
<td>5.7</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>3.0</td>
<td>4.4</td>
<td>1.4</td>
<td>7.2</td>
<td>3.2</td>
<td>6.0</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>2.8</td>
<td>4.8</td>
<td>1.4</td>
<td>6.2</td>
<td>2.8</td>
<td>4.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>3.0</td>
<td>5.0</td>
<td>1.7</td>
<td>6.1</td>
<td>3.0</td>
<td>4.9</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>2.9</td>
<td>4.5</td>
<td>1.5</td>
<td>6.4</td>
<td>2.8</td>
<td>5.6</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>2.6</td>
<td>3.5</td>
<td>1.0</td>
<td>7.2</td>
<td>3.0</td>
<td>5.8</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>2.4</td>
<td>3.8</td>
<td>1.1</td>
<td>7.4</td>
<td>2.8</td>
<td>6.1</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>2.4</td>
<td>3.7</td>
<td>1.0</td>
<td>7.0</td>
<td>3.8</td>
<td>6.4</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>2.7</td>
<td>3.9</td>
<td>1.2</td>
<td>6.4</td>
<td>2.8</td>
<td>5.6</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>2.7</td>
<td>5.1</td>
<td>1.6</td>
<td>6.3</td>
<td>2.8</td>
<td>5.1</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>3.0</td>
<td>4.5</td>
<td>1.5</td>
<td>6.1</td>
<td>2.6</td>
<td>5.6</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>3.4</td>
<td>4.5</td>
<td>1.6</td>
<td>7.7</td>
<td>3.0</td>
<td>6.1</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>6.7</td>
<td>3.1</td>
<td>4.7</td>
<td>1.5</td>
<td>6.3</td>
<td>3.4</td>
<td>5.6</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>6.3</td>
<td>2.3</td>
<td>4.4</td>
<td>1.3</td>
<td>6.4</td>
<td>3.1</td>
<td>5.5</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>3.0</td>
<td>4.1</td>
<td>1.3</td>
<td>6.0</td>
<td>3.0</td>
<td>4.8</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>2.5</td>
<td>4.0</td>
<td>1.3</td>
<td>6.0</td>
<td>3.1</td>
<td>5.4</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>2.6</td>
<td>4.4</td>
<td>1.2</td>
<td>6.7</td>
<td>3.1</td>
<td>5.6</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>3.0</td>
<td>4.6</td>
<td>1.4</td>
<td>6.0</td>
<td>3.1</td>
<td>5.1</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>2.6</td>
<td>4.0</td>
<td>1.2</td>
<td>5.8</td>
<td>2.7</td>
<td>5.1</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>2.3</td>
<td>3.3</td>
<td>1.0</td>
<td>6.8</td>
<td>3.2</td>
<td>5.9</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>2.7</td>
<td>4.2</td>
<td>1.3</td>
<td>6.7</td>
<td>3.3</td>
<td>5.7</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>3.0</td>
<td>4.2</td>
<td>1.2</td>
<td>6.7</td>
<td>3.0</td>
<td>5.2</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>2.9</td>
<td>4.2</td>
<td>1.3</td>
<td>6.3</td>
<td>2.5</td>
<td>5.0</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>2.9</td>
<td>4.3</td>
<td>1.3</td>
<td>6.5</td>
<td>3.0</td>
<td>5.2</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>2.5</td>
<td>3.0</td>
<td>1.1</td>
<td>6.2</td>
<td>3.4</td>
<td>5.4</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>2.8</td>
<td>4.1</td>
<td>1.3</td>
<td>5.9</td>
<td>3.0</td>
<td>5.1</td>
<td>1.8</td>
<td></td>
</tr>
</tbody>
</table>
and \( \hat{\Sigma}_2^{-1} = \begin{pmatrix} 45.953273 & -15.75067 & -42.84096 & 8.1010053 \\ -15.75067 & 77.428683 & 4.3874471 & -40.75300 \\ -42.84096 & 4.3874471 & 59.427269 & -13.51723 \\ 8.1010053 & -40.75300 & -13.51723 & 85.678729 \end{pmatrix} \).

If \( x'_0 = (x_{01}, ..., x_{04}) \) is data on a new specie, then we have the distance from \( x_0 \) to \( \hat{\mu}_1 \) as

\[
\sqrt{(x_0 - \hat{\mu}_1)' \hat{\Sigma}_1^{-1} (x_0 - \hat{\mu}_1)}
\]

\[
= \sqrt{\left[(x_{01} - 5.9301907, ..., x_{04} - 1.3164927)' \begin{pmatrix} 39.70618 & ... & 31.451537 \\ ... & ... & ... \\ 31.451537 & ... & 430.62776 \end{pmatrix} \begin{pmatrix} x_{01} - 5.9301907 \\ ... \\ x_{04} - 1.3164927 \end{pmatrix} \right]^\frac{1}{2}}
\]

and distance from \( x_0 \) to \( \hat{\mu}_2 \) as

\[
\sqrt{(x_0 - \hat{\mu}_2)' \hat{\Sigma}_2^{-1} (x_0 - \hat{\mu}_2)}
\]

\[
= \sqrt{\left[(x_{01} - 6.5506952, ..., x_{04} - 2.0322709)' \begin{pmatrix} 45.953273 & ... & 8.1010053 \\ ... & ... & ... \\ 8.1010053 & ... & 85.678729 \end{pmatrix} \begin{pmatrix} x_{01} - 6.5506952 \\ ... \\ x_{04} - 2.0322709 \end{pmatrix} \right]^\frac{1}{2}}
\]

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by "allocate \( x_0 \) to Iris Versicolor specie (\( \Pi_1 \)) if the distance from \( x_0 \) to \( \hat{\mu}_1 \) plus \( \frac{1}{2} \ln | \hat{\Sigma}_1 | = -8.496731 \) is smaller than or equal to the distance from \( x_0 \) to \( \hat{\mu}_2 \) plus \( \frac{1}{2} \ln | \hat{\Sigma}_2 | = -7.519708. " Otherwise, allocate \( x_0 \) to Iris Virginica specie (\( \Pi_2 \)).

### 4.2.2 Under AR(1) Structure

Sometimes when the data are repeated measures of a single variable or when they are of time series type then instead of using a general variance covariance matrix, a special structure like AR(1) may be more appropriate. In this section, we provide discriminant functions under AR(1) structure.

Suppose the Kotz type densities of \( X \) for populations \( \Pi_1 \) and \( \Pi_2 \) are given by

\[
f_i(x) = c \left( \frac{\sigma^2}{1 - \rho^2} \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{\sigma_i} \sqrt{(x - \mu_i)' \left[I_p + \rho^2 C_1 - \rho C_2 \right] (x - \mu_i)} \right\}, \quad i = 1, 2,
\]
where \( c = \frac{\Gamma(\frac{p}{2})}{2^{\frac{p}{2}}\Gamma(p)} \), \( C_1 \) and \( C_2 \) are defined in (2.5), (2.6).

**Case 1.** \( \Sigma_1 = \Sigma_2 = \Sigma \), that is, \( \sigma_1^2 = \sigma_2^2 = \sigma^2 \) and \( \rho_1 = \rho_2 = \rho \).

Let \( x_{11}, x_{12}, ..., x_{1n_1} \) and \( x_{21}, x_{22}, ..., x_{2n_2} \) be two random samples of sizes \( n_1 \) and \( n_2 \) from Kotz type populations \( \Pi_1 \) and \( \Pi_2 \) with unknown parameters \( \mu_1 \) and \( \mu_2 \) and a positive definite common covariance matrix \( (p + 1)\Sigma \), respectively, with \( n_1 + n_2 - 2 \geq p \). Then the minimum ECM regions determined by samples of sizes \( n_1 \) and \( n_2 \) respectively from (4.1) are simplified to

\[
\hat{R}_1 : \exp \left\{ -\frac{1}{\sigma} \sqrt{D_{11} + \hat{\rho}^2 D_{21} - \hat{\rho} D_{31}} + \frac{1}{\sigma} \sqrt{D_{12} + \hat{\rho}^2 D_{22} - \hat{\rho} D_{32}} \right\} \geq 1
\]

and \( \hat{R}_2 : \) otherwise.

Here

\[
\begin{align*}
D_{11} &= (x - \hat{\mu}_1)' (x - \hat{\mu}_1), \quad (4.4) \\
D_{21} &= (x - \hat{\mu}_1)' C_1 (x - \hat{\mu}_1), \quad (4.5) \\
D_{31} &= (x - \hat{\mu}_1)' C_2 (x - \hat{\mu}_1), \quad (4.6) \\
D_{12} &= (x - \hat{\mu}_2)' (x - \hat{\mu}_2), \quad (4.7) \\
D_{22} &= (x - \hat{\mu}_2)' C_1 (x - \hat{\mu}_2), \quad (4.8) \\
D_{32} &= (x - \hat{\mu}_2)' C_2 (x - \hat{\mu}_2). \quad (4.9)
\end{align*}
\]

Given these regions \( \hat{R}_1 \) and \( \hat{R}_2 \), we can construct the sample classification rule as given in the following result.

**Result 4.4** Let the populations \( \Pi_1 \) and \( \Pi_2 \) be Kotz type populations with equal AR(1) structured covariance matrices. Then the linear classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate \( x_0 \) to \( \Pi_1 \) if

\[
-\frac{1}{\sigma} \sqrt{D_{11} + \hat{\rho}^2 D_{21} - \hat{\rho} D_{31}} + \frac{1}{\sigma} \sqrt{D_{12} + \hat{\rho}^2 D_{22} - \hat{\rho} D_{32}} \geq 0
\]

This classification function is equivalent to allocating \( x_0 \) to \( \Pi_1 \) if

\[
(\hat{\mu}_1 - \hat{\mu}_2)' x_0 + \hat{\rho}^2 (\hat{\mu}_1 - \hat{\mu}_2)' C_1 x_0 - \hat{\rho} (\hat{\mu}_1 - \hat{\mu}_2)' C_2 x_0 - k \geq 0, \quad (4.10)
\]
where \( \hat{\mu}_1, \hat{\mu}_2 \) and \( \hat{\rho} \) are MLEs of \( \mu_1, \mu_2 \) and \( \rho \) respectively and \( C_1 \) and \( C_2 \) are as in (2.5) and (2.6) and

\[
k = \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)'(\hat{\mu}_1 + \hat{\mu}_2) + \frac{1}{2}\hat{\rho}^2(\hat{\mu}_1 - \hat{\mu}_2)'C_1(\hat{\mu}_1 + \hat{\mu}_2) + \frac{1}{2}\hat{\rho}(\hat{\mu}_1 - \hat{\mu}_2)'C_2(\hat{\mu}_1 + \hat{\mu}_2).
\]

Allocate \( \mathbf{x}_0 \) to \( \Pi_2 \) otherwise.

**Case 2.** \( \Sigma_1 \neq \Sigma_2 \), that is, \( \sigma_i^2 \)'s and \( \rho_i \)'s are different.

Let \( \mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1} \) and \( \mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2} \) be two random samples of sizes \( n_1 \) and \( n_2 \) from Kotz type populations, \( \Pi_1 \) and \( \Pi_2 \) with unknown parameters \( \mu_1, \Sigma_1 \) and \( \mu_2, \Sigma_2 \), respectively, with \( n_1 - 1 \geq p \) and \( n_2 - 1 \geq p \). Then the minimum ECM regions determined by samples of sizes \( n_1 \) and \( n_2 \) respectively for (4.1) are simplified to

\[
\hat{R}_1 : \exp \left\{ -\frac{1}{\hat{\sigma}_1} \sqrt{D_{11} + \hat{\rho}_1^2D_{21} - \hat{\rho}_1D_{31}} + \frac{1}{\hat{\sigma}_2} \sqrt{D_{12} + \hat{\rho}_2^2D_{22} - \hat{\rho}_2D_{32}} \right\} \geq 1
\]

and \( \hat{R}_2 : \) otherwise.

Given these regions \( \hat{R}_1 \) and \( \hat{R}_2 \), we can construct the sample classification rule as given in the following result.

**Result 4.5** Let the populations \( \Pi_1 \) and \( \Pi_2 \) be Kotz type populations with unequal AR(1) structured covariance matrices. Then the non-linear classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate \( \mathbf{x}_0 \) to \( \Pi_1 \) if

\[
\frac{1}{\hat{\sigma}_2} \sqrt{D_{12} + \hat{\rho}_2^2D_{22} - \hat{\rho}_2D_{32}} - \frac{1}{\hat{\sigma}_1} \sqrt{D_{11} + \hat{\rho}_1^2D_{21} - \hat{\rho}_1D_{31}} - k \geq 0,
\]

where \( D_{11}, D_{21}, D_{31}, D_{12}, D_{22} \) and \( D_{32} \) are as in (4.4) - (4.9) and

\[
k = \frac{p}{2} \ln \left( \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right) + \frac{1}{2} \ln \left( \frac{1 - \hat{\rho}_2^2}{1 - \hat{\rho}_1^2} \right).
\]

Alternatively, allocate \( \mathbf{x}_0 \) to \( \Pi_1 \) if
Example 4.2. Although, AR(1) structure may not be appropriate for this example, we illustrate this procedure using Iris data of Example 4.1. These data were found to have unequal AR(1) covariance matrices and hence we used Case 2 to illustrate our procedure. Next, we computed the MLEs of \( \mu_1, \mu_2, \sigma_1, \sigma_2, \rho_1 \) and \( \rho_2 \) under Kotz type distributions. The results are as follows:

\[
\Pi_1(\text{Iris Versicolor}) : n_1 = 50,
\begin{align*}
\hat{\mu}_1 &= (5.9048405, 2.7879834, 4.2520208, 1.3172935)', \\
\hat{\sigma}_1 &= 0.1523249 \quad \text{and} \quad \hat{\rho}_1 = 0.5391383.
\end{align*}
\]

\[
\Pi_2(\text{Iris Virginica}) : n_2 = 50,
\begin{align*}
\hat{\mu}_2 &= (6.5320442, 2.9815112, 5.4915763, 2.04259659)', \\
\hat{\sigma}_2 &= 0.1909728 \quad \text{and} \quad \hat{\rho}_2 = 0.389897.
\end{align*}
\]

Substituting these values in (4.11), we obtain the non-linear classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate \( x_0 \) to Iris Versicolor specie (\( \Pi_1 \)) if

\[
\frac{1}{\hat{\sigma}_1} \sqrt{D_{11} + \hat{\rho}_1^2 D_{21} - \hat{\rho}_1 D_{31} + \frac{1}{2} \ln \left( \frac{(\hat{\sigma}_1^2)^p}{1 - \hat{\rho}_1^2} \right)} \leq \frac{1}{\hat{\sigma}_2} \sqrt{D_{12} + \hat{\rho}_2^2 D_{22} - \hat{\rho}_2 D_{32} + \frac{1}{2} \ln \left( \frac{(\hat{\sigma}_2^2)^p}{1 - \hat{\rho}_2^2} \right)},
\]

otherwise, allocate \( x_0 \) to \( \Pi_2 \).

Allocate \( x_0 \) to Iris Virginica specie (\( \Pi_2 \)) otherwise.
4.2.3 Evaluating Classification Functions for Two Populations

Because parent populations are rarely known, we consider the error rates as a measure of judging the performance of any classification procedure. The performance of sample classification functions can be evaluated by calculating the actual error rate (AER) given by

\[ AER = p_1 \int_{\hat{R}_2} f_1(x)dx + p_2 \int_{\hat{R}_1} f_2(x)dx, \]

where \( \hat{R}_1 \) and \( \hat{R}_2 \) are the classification regions determined by samples of sizes \( n_1 \) and \( n_2 \) respectively.

The AER indicates how the sample classification function will perform in future samples. In general, it cannot be calculated because it depends on the unknown density functions \( f_1(x) \) and \( f_2(x) \). However, a nearly unbiased estimate of the expected actual error rate (\( \hat{E}(AER) \)) can be calculated by using Lachenbruch's holdout procedure under the equal prior probabilities and equal misclassification costs (see Lachenbruch and Mickey (1963)). Constructing the error-rate estimates do not require distributional assumptions and can be calculated from what is called as confusion matrix. This procedure is sometimes referred to as jackknifing or cross-validation. For \( n_1 \) and \( n_2 \) observations from Kotz type populations \( \Pi_1 \) and \( \Pi_2 \) respectively, the confusion matrix has the form

<table>
<thead>
<tr>
<th>Predicted membership</th>
<th>( \Pi_1 )</th>
<th>( \Pi_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pi_1 )</td>
<td>( n_{1C} )</td>
<td>( n_{1M} = n_1 - n_{1C} )</td>
</tr>
<tr>
<td>( \Pi_2 )</td>
<td>( n_{2M} = n_2 - n_{2C} )</td>
<td>( n_{2C} )</td>
</tr>
</tbody>
</table>

where

\( n_{iC} = \text{number of } \Pi_i \text{ items correctly classified into } \Pi_i \text{ items, } i = 1, 2 \)
\( n_{iM} = \text{number of } \Pi_i \text{ items misclassified into } \Pi_j \text{ items, } i, j = 1, 2 \) and \( i \neq j \).

A nearly unbiased estimate of the expected actual error rate, \( \hat{E}(AER) \) is given by

\[ \hat{E}(AER) = \frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2}, \]
where \( n_{iM}^{(H)} \) are the number of holdout (H) observations misclassified in the \( i^{th} \) group, \( i = 1, 2 \).

When we adopted the Lachenbruch's holdout procedure using Kotz type densities, we obtained the confusion matrix as follows: (\( \Pi_1 \): Iris Versicolor, \( \Pi_2 \): Iris Virginica).

\[
\begin{array}{ccc}
\text{Predicted membership} & \Pi_1 & \Pi_2 \\
\text{Actual membership} & \Pi_1 & 46 & 4 & 50 \\
& \Pi_2 & 1 & 49 & 50 \\
\end{array}
\]

and consequently, \( \hat{E}(AER) = \frac{5}{100} = 0.05 \).

When we adopted the quadratic discriminant rule under multivariate normal distributions with unequal covariance matrices, we obtained the confusion matrix as follows:

\[
\begin{array}{ccc}
\text{Predicted membership} & \Pi_1 & \Pi_2 \\
\text{Actual membership} & \Pi_1 & 47 & 3 & 50 \\
& \Pi_2 & 1 & 49 & 50 \\
\end{array}
\]

and consequently, \( \hat{E}(AER) = \frac{4}{100} = 0.04 \).

However, Iris data were found to have come from multivariate normal distributions with unequal covariance matrices. In general, for the data that satisfy multivariate normal distribution assumption, the error rate seems to be lower for the rules based on multivariate normal densities than those based on Kotz type densities.

It would be interesting to compare these two error rates, that is, the two error rates based on multivariate normal and Kotz type distribution assumptions when the data are from Kotz type distributions. For this purpose, we adopt the following samples
simulated from Kotz type densities.

**Example 4.3.** We generate data with four variables and 30 observations each from two Kotz type populations $\Pi_1$ and $\Pi_2$ with the following means and variance covariance matrices; where

$$\mu_1 = \left( 5.9301910, 2.7846634, 4.2389532, 1.3164930 \right)^\prime,$$

$$\Sigma_1 = \begin{pmatrix}
0.0651017 & 0.0220158 & 0.042686 & 0.0135321 \\
0.0220158 & 0.024482 & 0.0206853 & 0.0101082 \\
0.042686 & 0.0206853 & 0.0489631 & 0.0167515 \\
0.0135321 & 0.0101082 & 0.0167515 & 0.0088661
\end{pmatrix},$$

$$\mu_2 = \left( 6.5506954, 2.9737208, 5.5073221, 2.0322710 \right)^\prime,$$

$$\Sigma_2 = \begin{pmatrix}
0.0836884 & 0.0193137 & 0.0613978 & 0.0109602 \\
0.0193137 & 0.021707 & 0.0147842 & 0.0108312 \\
0.0613978 & 0.0147842 & 0.0625199 & 0.0110904 \\
0.0109602 & 0.0108312 & 0.0110904 & 0.0175368
\end{pmatrix}. $$

The SAS program to compute the estimates and discriminant rule is provided in Program 3 in APPENDIX.

The Lachenbruch's holdout procedure using Kotz type densities provides the following confusion matrix:

<table>
<thead>
<tr>
<th>Predicted membership</th>
<th>$\Pi_1$</th>
<th>$\Pi_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pi_1$</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>$\Pi_2$</td>
<td>1</td>
<td>29</td>
</tr>
</tbody>
</table>

and consequently, $\hat{E}(AER) = \frac{1}{60} = 0.0167$.

When we adopted the quadratic discriminant rule under multivariate normal distributions with unequal covariance matrices, we obtained the confusion matrix as follows:
and consequently, $\hat{E}(AER) = \frac{2}{60} = 0.0333$. This rate is considerably higher than 0.0167. Thus, the discriminant function using Kotz type densities is a better measure of performance and can achieve effective classification when the samples are from Kotz type populations.

4.3 Separation and Classification for Several Populations

Let $f_i(x)$ be the probability density function of a $p \times 1$ random vector $X$ under population $\Pi_i$, $i = 1, \ldots, g$.

Let $p_i = \text{the prior probability of population } \Pi_i$,

$c(j/i) = \text{the cost of allocating an object into } \Pi_j \text{ when it comes from } \Pi_i$,

$R_j = \text{the set of } x \text{'s classified into } \Pi_j$ and

$P(j/i) = P(\text{classifying an object into } \Pi_j \text{ when it is from } \Pi_i)$

$$= \int_{R_j} f_i(x) dx \quad \text{with } P(i/i) = 1 - \sum_{j=1, j\neq i}^{g} P(j/i).$$

We shall consider the expected cost of misclassification,

$$ECM = \sum_{i=1}^{g} p_i \sum_{j=1, j\neq i}^{g} P(j/i) c(j/i),$$

as our criteria for assessing various classification procedures.
Result 4.6  The classification regions that minimize the ECM are defined by allocating \( x \) to population \( \Pi_j, j = 1, \ldots, g \) for which

\[
\sum_{i=1, i \neq j}^{g} p_i f_i(x) \, c(j/i),
\]

is smallest.

Under the equal prior probabilities and equal misclassification costs allocate \( x_0 \), a new observation, into population \( \Pi_j \) if

\[
f_j(x_0) > f_i(x_0), \text{ for } i = 1, \ldots, g, \ i \neq j.
\]

Alternatively, allocate \( x_0 \) into population \( \Pi_j \) if

\[
\ln f_j(x_0) > \ln f_i(x_0), \text{ for } i = 1, \ldots, g, \ i \neq j.
\]

See Johnson and Wichern (1998) for proofs and other details.

4.3.1 Classification Rules Under Kotz Type Distributions and General Covariance Structure

We now assume that \( f_i(x) \) is a Kotz type density with parameters \( \mu_i \) and \( \Sigma_i, \ i = 1, \ldots, g \). We consider the two cases, of equal and unequal covariance assumptions.

Case 1. \( \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g = \Sigma \).

The Kotz type densities of \( X \) for population \( \Pi_i \) with parameters \( \mu_i \) and a common variance covariance matrix \( (p + 1)\Sigma, \ i = 1, \ldots, g \) are given by

\[
f_i(x) = c \, |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\left[ (x - \mu_i)'\Sigma^{-1}(x - \mu_i) \right]^{\frac{1}{2}} \right\}, \quad \mu_i \in \mathbb{R}^p, \ \Sigma \ p. \ d., \ i = 1, \ldots, g,
\]

where \( c = \frac{\Gamma\left(\frac{p+1}{2}\right)}{2\pi^{p/2}\Gamma(p/2)} \).

Let \( x_{i1}, x_{i2}, \ldots, x_{in_i} \) be random samples of sizes \( n_i \) from \( f_i(.) \) with \( \sum_{i=1}^{g} (n_i - 1) \geq p \).

We provide the classification rule below.

Result 4.7  Let the populations \( \Pi_i, \ i = 1, \ldots, g \) be Kotz type populations with equal
general covariance matrices. Then the classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$ to $\Pi_j$ if

$$\hat{d}_j(x_0) = \max_{1 \leq i \leq g} \hat{d}_i(x_0),$$

where

$$\hat{d}_i(x_0) = \ln f_i(x_0) = -\frac{1}{2} \ln |\Sigma| - \sqrt{(x_0 - \mu_i)'\Sigma^{-1}(x_0 - \mu_i)}.$$

This classification function is equivalent to allocating $x_0$ to $\Pi_j$ if

$$\hat{d}_{ji}(x_0) = d_j(x_0) - d_i(x_0) = (\hat{\mu}_j - \hat{\mu}_i)'\Sigma^{-1}x_0 - \frac{1}{2}(\hat{\mu}_j - \hat{\mu}_i)'\Sigma^{-1}(\hat{\mu}_j + \hat{\mu}_i) \geq 0, \text{ for } i = 1, \ldots, g, \ i \neq j,$$

(4.12)

where $\hat{\mu}_i, \hat{\mu}_j$ and $\hat{\Sigma}$ are MLEs of $\mu_i, \mu_j$ and $\Sigma$ respectively.

The result is obtained by following similar steps as in Johnson and Wichen (1998). We note that this rule is similar to the rule based on the normal densities except that here we estimate $\mu_i, \mu_j$ and $\Sigma$ by the optimization method mentioned in Chapter 2.

Case 2. $\Sigma_i$’s are different

The Kotz type densities of $X$ for population $\Pi_i$ with parameters $\mu_i$ and $\Sigma_i$, $i = 1, \ldots, g$ are given by

$$f_i(x) = c |\Sigma_i|^{-\frac{1}{2}} \exp \{-[(x - \mu_i)'\Sigma_i^{-1}(x - \mu_i)]^{\frac{1}{2}}\}, \ \mu_i \in \mathbb{R}^p, \ \Sigma_i \text{ p. d.}, \ i = 1, \ldots, g,$$

where $c = \frac{\Gamma(\frac{p+1}{2})}{2^{p/2} \Gamma(p)}. $

Suppose $x_{i1}, x_{i2}, \ldots, x_{in_i}$ are random samples of sizes $n_i$ from $f_i(.)$ with $n_i - 1 \geq p, \ i = 1, \ldots, g$. Then we have the following result.

Result 4.8 Let the populations $\Pi_i, \ i = 1, \ldots, g$ be Kotz type populations with unequal general covariance matrices. Then the classification rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$ to $\Pi_j$ if

$$\hat{d}_j(x_0) = \max_{1 \leq i \leq g} \hat{d}_i(x_0),$$
where \( d_i(x_0) = \ln f_i(x_0) = -\frac{1}{2} \ln |\hat{\Sigma}_i| - \sqrt{(x_0 - \hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x_0 - \hat{\mu}_i)}, \ i = 1, \ldots, g. \)

This classification function is equivalent to allocating \( x_0 \) to \( \Pi_j \) if

\[
\hat{d}_j(x_0) = d_j(x_0) - d_i(x_0) \\
= \sqrt{(x_0 - \hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x_0 - \hat{\mu}_i)} - \sqrt{(x_0 - \hat{\mu}_j)'\hat{\Sigma}_j^{-1}(x_0 - \hat{\mu}_j)} - \frac{1}{2} \ln \left( \frac{|\hat{\Sigma}_j|}{|\hat{\Sigma}_i|} \right) \geq 0,
\]

that is, if

\[
\sqrt{(x_0 - \hat{\mu}_j)'\hat{\Sigma}_j^{-1}(x_0 - \hat{\mu}_j)} + \frac{1}{2} \ln |\hat{\Sigma}_j| \quad \text{is smallest, (that is, if } \ln f_j(x_0) \text{ is largest)}
\]

where \( \hat{\mu}_i, \hat{\mu}_j, \hat{\Sigma}_i \) and \( \hat{\Sigma}_j \) are MLEs of \( \mu_i, \mu_j, \Sigma_i \) and \( \Sigma_j \) respectively.

**Example 4.4.** For illustration of the results we once again consider Fisher’s Iris data, but now with all the three species. The three species of Irises data considered are Iris Versicolor (\( \Pi_2 \)), Iris Virginica (\( \Pi_3 \)) displayed in Table 4.1 and Iris Sesota (\( \Pi_1 \)) displayed in Table 4.2 respectively. Data on four variables, \( x_1 = \) sepal length, \( x_2 = \) sepal width, \( x_3 = \) petal length and \( x_4 = \) petal width for 50 observations each from the three species are available. Again, we have taken the data from Table 11.5 of Johnson and Wichern (1998, pp. 715 - 716). These data were found to have unequal covariance matrices and hence we use Case 2 to illustrate our procedure.

Here we provide ML estimates of \( \mu \) and \( \Sigma \) and the discrimination rule for the new group that we included in this example (as compared to Example 4.1).

\( \Pi_1 \) (Iris Setosa) : \( n_1 = 50 \),

\[
\hat{\mu}_1 = \begin{pmatrix} 5.0003585, 3.4047069, 1.4585232, 0.2283304 \end{pmatrix},
\]

\[
\hat{\Sigma}_1 = \begin{pmatrix} 0.0276767 & 0.0213725 & 0.0033927 & 0.0020632 \\ 0.0213725 & 0.0301198 & 0.0023413 & 0.0022393 \\ 0.0033927 & 0.0023413 & 0.0057416 & 0.0009577 \\ 0.0020632 & 0.0022393 & 0.0009577 & 0.0021534 \end{pmatrix},
\]
Table 4.2: Data on Iris Setosa.

<table>
<thead>
<tr>
<th>Sepal length</th>
<th>Sepal width</th>
<th>Petal Length</th>
<th>Petal width</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>3.5</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>4.9</td>
<td>3.0</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>4.7</td>
<td>3.2</td>
<td>1.3</td>
<td>0.2</td>
</tr>
<tr>
<td>4.6</td>
<td>3.1</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.6</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>5.4</td>
<td>3.9</td>
<td>1.7</td>
<td>0.4</td>
</tr>
<tr>
<td>4.6</td>
<td>3.4</td>
<td>1.4</td>
<td>0.3</td>
</tr>
<tr>
<td>5.0</td>
<td>3.4</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>4.4</td>
<td>2.9</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>4.9</td>
<td>3.1</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>5.4</td>
<td>3.7</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>4.8</td>
<td>3.4</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>4.8</td>
<td>3.0</td>
<td>1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>4.3</td>
<td>3.0</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td>5.8</td>
<td>4.0</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>5.7</td>
<td>4.4</td>
<td>1.5</td>
<td>0.4</td>
</tr>
<tr>
<td>5.4</td>
<td>3.9</td>
<td>1.3</td>
<td>0.4</td>
</tr>
<tr>
<td>5.1</td>
<td>3.5</td>
<td>1.4</td>
<td>0.3</td>
</tr>
<tr>
<td>5.7</td>
<td>3.8</td>
<td>1.7</td>
<td>0.3</td>
</tr>
<tr>
<td>5.1</td>
<td>3.8</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td>5.4</td>
<td>3.4</td>
<td>1.7</td>
<td>0.2</td>
</tr>
<tr>
<td>5.1</td>
<td>3.7</td>
<td>1.5</td>
<td>0.4</td>
</tr>
<tr>
<td>4.6</td>
<td>3.6</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>5.1</td>
<td>3.3</td>
<td>1.7</td>
<td>0.5</td>
</tr>
<tr>
<td>4.8</td>
<td>3.4</td>
<td>1.9</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.0</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.4</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>5.2</td>
<td>3.5</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5.2</td>
<td>3.4</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>4.7</td>
<td>3.2</td>
<td>1.8</td>
<td>0.2</td>
</tr>
<tr>
<td>4.8</td>
<td>3.1</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>5.4</td>
<td>3.4</td>
<td>1.5</td>
<td>0.4</td>
</tr>
<tr>
<td>5.2</td>
<td>4.1</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>5.5</td>
<td>4.2</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>4.0</td>
<td>3.1</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.2</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>5.5</td>
<td>3.5</td>
<td>1.3</td>
<td>0.2</td>
</tr>
<tr>
<td>4.9</td>
<td>3.6</td>
<td>1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>4.4</td>
<td>3.0</td>
<td>1.3</td>
<td>0.2</td>
</tr>
<tr>
<td>5.1</td>
<td>3.4</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.5</td>
<td>1.3</td>
<td>0.3</td>
</tr>
<tr>
<td>4.5</td>
<td>2.3</td>
<td>1.3</td>
<td>0.3</td>
</tr>
<tr>
<td>4.4</td>
<td>3.2</td>
<td>1.3</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.5</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>5.1</td>
<td>3.8</td>
<td>1.9</td>
<td>0.4</td>
</tr>
<tr>
<td>4.8</td>
<td>3.0</td>
<td>1.4</td>
<td>0.3</td>
</tr>
<tr>
<td>5.1</td>
<td>3.8</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>4.6</td>
<td>3.2</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td>5.3</td>
<td>3.7</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>3.3</td>
<td>1.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

If $x'_0 = (x_{01}, \ldots, x_{04})$, then we have

$$\sqrt{(x_0 - \mu_1)'\Sigma_1^{-1}(x_0 - \mu_1) + \frac{1}{2} \ln | \Sigma_1 |}$$

$$= \left[ (x_{01} - 5.0003585, \ldots, x_{04} - 0.2283304)' \begin{pmatrix} 83.726660 & \ldots & -10.27257 \\ \vdots & \ddots & \vdots \\ -10.27257 & \ldots & 534.18177 \end{pmatrix} \begin{pmatrix} x_{01} - 5.0003585 \\ \vdots \\ x_{04} - 0.2283304 \end{pmatrix} \right]^{\frac{1}{2}}$$

$$- 9.700359,$$

$$\sqrt{(x_0 - \mu_2)'\Sigma_2^{-1}(x_0 - \mu_2) + \frac{1}{2} \ln | \Sigma_2 |}$$

$$= \left[ (x_{01} - 5.9301907, \ldots, x_{04} - 1.3164927)' \begin{pmatrix} 39.70618 & \ldots & 31.451537 \\ \vdots & \ddots & \vdots \\ 31.451537 & \ldots & 430.62776 \end{pmatrix} \begin{pmatrix} x_{01} - 5.9301907 \\ \vdots \\ x_{04} - 1.3164927 \end{pmatrix} \right]^{\frac{1}{2}}$$

$$- 8.496731,$$

and $\sqrt{(x_0 - \mu_3)'\Sigma_3^{-1}(x_0 - \mu_3) + \frac{1}{2} \ln | \Sigma_3 |}$

$$= \left[ (x_{01} - 6.5506952, \ldots, x_{04} - 2.0322709)' \begin{pmatrix} 45.953273 & \ldots & 8.1010053 \\ \vdots & \ddots & \vdots \\ 8.1010053 & \ldots & 85.678729 \end{pmatrix} \begin{pmatrix} x_{01} - 6.5506952 \\ \vdots \\ x_{04} - 2.0322709 \end{pmatrix} \right]^{\frac{1}{2}}$$

$$- 7.5197089.$$

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by "allocate $x_0$ to $\Pi_j$ if the distance from $x_0$ to $\mu_j$ plus $\frac{1}{2} \ln | \Sigma_j |$, $j = 1, 2, 3$ is smallest."
4.3.2 Under AR(1) Structure

The Kotz type densities of $X$ for populations $\Pi_i$, with parameters $\mu_i$ and $\Sigma_i$, $i = 1, \ldots, g$, are given by

$$f_i(x) = c \left( \frac{\sigma_i^2 \rho_i^2}{1 - \rho_i^2} \right)^{-\frac{1}{2}} \exp \left\{ - \frac{1}{\sigma_i} \sqrt{(x - \mu_i)' [I_2 + \rho_i^2 \Sigma_i - \rho_i \Sigma_2] (x - \mu_i)} \right\}, \quad i = 1, \ldots, g,$$

where $c = \frac{\Gamma\left(\frac{p}{2}\right)}{2\sigma_i^{2p}\Gamma(p)}$, $C_1$ and $C_2$ are as in (2.5), (2.6).

**Case 1.** $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_g = \Sigma$, that is, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_g^2 = \sigma^2$ and $\rho_1 = \rho_2 = \ldots = \rho_g = \rho$.

Suppose $x_{i1}, x_{i2}, \ldots, x_{in_i}$ are random samples of sizes $n_i$ from $f_i(.)$ with $\sum_{i=1}^{g}(n_i - 1) \geq p$.

**Result 4.9** Let the populations $\Pi_i$, $i = 1, \ldots, g$ be Kotz type densities with equal AR(1) covariance matrices. Then the estimated minimum ECM rule under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$ to $\Pi_j$ if

$$\hat{d}_j(x_0) = \max_{1 \leq i \leq g} \hat{d}_i(x_0),$$

where $\hat{d}_i(x_0) = -\frac{1}{\hat{\sigma}} \sqrt{D_{1i} + \hat{\rho}^2D_{2i} - \hat{\rho}D_{3i}}$ and

$$D_{1i} = (x_0 - \hat{\mu}_i)'(x_0 - \hat{\mu}_i),$$
$$D_{2i} = (x_0 - \hat{\mu}_i)'C_1(x_0 - \hat{\mu}_i),$$
$$D_{3i} = (x_0 - \hat{\mu}_i)'C_2(x_0 - \hat{\mu}_i).$$

This classification function is equivalent to allocating $x_0$ to $\Pi_j$ if

$$\hat{d}_{ji}(x_0) = (\hat{\mu}_j - \hat{\mu}_i)'x_0 + \hat{\rho}^2(\hat{\mu}_j - \hat{\mu}_i)'C_1x_0 - \hat{\rho}(\hat{\mu}_j - \hat{\mu}_i)'C_2x_0 - k_1 \geq 0, \quad (4.14)$$

where $\hat{\mu}_j$, $\hat{\mu}_i$, $\hat{\sigma}$ and $\hat{\rho}$ are MLEs of $\mu_j$, $\mu_i$, $\sigma$ and $\rho$ respectively and

$$k_1 = \frac{1}{2}(\hat{\mu}_j - \hat{\mu}_i)'(\hat{\mu}_j + \hat{\mu}_i) + \frac{1}{2}\hat{\rho}^2(\hat{\mu}_j - \hat{\mu}_i)'C_1(\hat{\mu}_j + \hat{\mu}_i) + \frac{1}{2}\hat{\rho}(\hat{\mu}_j - \hat{\mu}_i)'C_2(\hat{\mu}_j + \hat{\mu}_i).$$

**Case 2.** $\Sigma_i$'s are different, that is, $\sigma_i^2$'s and $\rho_i$'s are different.
Result 4.10 Let the populations $\Pi_i$, $i = 1, ..., g$ be Kotz type densities with unequal AR(1) covariance matrices. Then the estimated minimum ECM rule under equal prior probabilities and equal misclassification costs is as follows:

Allocate $x_0$ to $\Pi_j$ if

$$\hat{d}_j(x_0) = \max_{1 \leq i \leq g} \hat{d}_i(x_0),$$

where $\hat{d}_i(x_0) = -\frac{1}{2} \ln \left( \frac{\sigma^2_i}{1 - \rho_i^2} \right) - \frac{1}{\sigma_i} \sqrt{D_{11} + \rho_i^2 D_{21} - \hat{\rho}_i D_{31}}.$$

This classification function is equivalent to allocating $x_0$ to $\Pi_j$ if

$$\hat{d}_{ij}(x_0) = \frac{1}{\sigma_i} \sqrt{D_{11} + \rho_i^2 D_{21} - \hat{\rho}_i D_{31}} - \frac{1}{\sigma_j} \sqrt{D_{1j} + \rho_j^2 D_{2j} - \hat{\rho}_j D_{3j}} - k_1 \geq 0$$

where $k_1 = \frac{\xi}{2} \ln \left( \frac{\sigma^2_i}{\sigma^2_j} \right) + \frac{1}{2} \ln \left( \frac{1 - \rho_i^2}{1 - \rho_j^2} \right)$, that is, if

$$\frac{1}{\sigma_j} \sqrt{D_{1j} + \rho_j^2 D_{2j} - \hat{\rho}_j D_{3j}} + \frac{1}{2} \ln \left( \frac{\sigma^2_j}{\sigma^2_i} \right)$$

is smallest. (4.15)

Example 4.5. We illustrate this procedure by using the Iris data considered in Example 4.4. The MLEs for Iris Sesota group and the classification rules are provided below.

$\Pi_1$ (Iris Sesota) : $n_1 = 50,$

$$\hat{\mu}_1 = \left( 5.0105823, 3.4134009, 1.4686691, 0.2409602 \right)',$n_1 = 50,$

$$\hat{\sigma}_1 = 0.1077129 \quad \text{and} \quad \hat{\rho}_1 = 0.456636.$$

If $x_0' = (x_{01}, ..., x_{04})$, then we have

$$\frac{1}{\hat{\sigma}_1} \left( D_{11} + \hat{\rho}_1^2 D_{21} - \hat{\rho}_1 D_{31} \right)^{\frac{1}{2}} + \frac{1}{2} \ln \left( \frac{\hat{\sigma}_1^2}{1 - \hat{\rho}_1^2} \right)$$

$$= \frac{1}{0.1077129} \left( D_{11} + 0.208516436 D_{21} - 0.456636 D_{31} \right)^{\frac{1}{2}} - 8.796221$$

$$\frac{1}{\hat{\sigma}_2} \left( D_{12} + \hat{\rho}_2^2 D_{22} - \hat{\rho}_2 D_{32} \right)^{\frac{1}{2}} + \frac{1}{2} \ln \left( \frac{\hat{\sigma}_2^2}{1 - \hat{\rho}_2^2} \right)$$

$$= \frac{1}{0.1523249} \left( D_{12} + 0.290670106 D_{22} - 0.5391383 D_{32} \right)^{\frac{1}{2}} - 7.355241$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[
\frac{1}{\sigma_3}
\left( D_{13} + \hat{\rho}_3^2 D_{23} - \hat{\rho}_3 D_{33} \right) \right)^{\frac{3}{2}} + \frac{1}{2} \ln \left( \frac{\sigma_j^2}{\hat{\rho}_j^2} \right) \\
= \frac{1}{0.190928} \left( D_{13} + 0.152019436 D_{23} - 0.3898967 D_{33} \right)^{\frac{3}{2}} - 6.540048
\]

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by

allocate \( x_0 \) to \( \Pi_j \) if
\[
\frac{1}{\sigma_j} \left( D_{1j} + \hat{\rho}_j^2 D_{2j} - \hat{\rho}_j D_{3j} \right) \right)^{\frac{3}{2}} + \frac{1}{2} \ln \left( \frac{\sigma_j^2}{\hat{\rho}_j^2} \right), \; j = 1, 2, 3 \text{ is smallest.}
\]

4.3.3 Evaluating Classification Functions for Several Populations

In a similar way as in the case of two populations, we can judge the performance of the discriminant function by evaluating the error rate using Lachenbruch's holdout procedure.

If \( n_{iM}^{(H)} \) is the number of misclassified holdout observations in the \( i^{th} \) group, \( i = 1, \ldots, g \) then an estimate of the expected actual error rate, \( \hat{E}(AER) \), is given by

\[
\hat{E}(AER) = \frac{\sum_{i=1}^{g} n_{iM}^{(H)}}{\sum_{i=1}^{g} n_i}.
\]

For Example 4.4, that is, for the Iris data, we adopted the Lachenbruch's holdout procedure using Kotz type densities, and obtained the following confusion matrix (\( \Pi_1 \): Iris Setosa, \( \Pi_2 \): Iris Versicolor, \( \Pi_3 \): Iris Virginica).

<table>
<thead>
<tr>
<th>Predicted membership</th>
<th>( \Pi_1 )</th>
<th>( \Pi_2 )</th>
<th>( \Pi_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Pi_1 )</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \Pi_2 )</td>
<td>0</td>
<td>46</td>
<td>4</td>
</tr>
<tr>
<td>( \Pi_3 )</td>
<td>0</td>
<td>1</td>
<td>49</td>
</tr>
</tbody>
</table>

and consequently, \( \hat{E}(AER) = \frac{5}{150} = 0.0333333 \).

When we used the multivariate normal densities for the same example, the sample quadratic classification rule provided the the confusion matrix as follows:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
and consequently, the estimate of the expected actual error rate = 0.0267. It shows that the discriminant function using Kotz type densities provides slightly higher error rate than that using quadratic discriminant rule based on multivariate normal distributions with unequal covariance matrices. Once again, this could be due to the fact that the Iris data are found to follow multivariate normal distribution.

To evaluate the performance of the discriminant rules when data are from Kotz type distributions we consider the simulated data in the following example.

**Example 4.6.** We generate data with three variables for 30, 35 and 40 observations from three Kotz type populations $\Pi_1$, $\Pi_2$ and $\Pi_3$ with the following means and variance covariance matrices respectively where

$$
\mu_1 = \begin{pmatrix} 5.006, 3.428, 1.462 \end{pmatrix},
$$

$$
\Sigma_1 = \begin{pmatrix} 0.121764 & 0.097232 & 0.016028 \\ 0.097232 & 0.140816 & 0.011464 \\ 0.016028 & 0.011464 & 0.029556 \end{pmatrix},
$$

$$
\mu_2 = \begin{pmatrix} 5.936, 2.77, 4.26 \end{pmatrix},
$$

$$
\Sigma_2 = \begin{pmatrix} 0.261104 & 0.08348 & 0.17924 \\ 0.08348 & 0.0965 & 0.081 \\ 0.17924 & 0.081 & 0.2164 \end{pmatrix},
$$

$$
\mu_3 = \begin{pmatrix} 6.588, 2.974, 5.552 \end{pmatrix},
$$
Now for these data the Lachenbruch’s holdout procedure using Kotz type densities provides the following confusion matrix.

\[
\Sigma_3 = \begin{pmatrix}
0.396256 & 0.091888 & 0.297224 \\
0.091888 & 0.101924 & 0.069952 \\
0.297224 & 0.069952 & 0.298496
\end{pmatrix}
\]

and consequently, \( \hat{E}(AER) = \frac{13}{105} = 0.1238 \).

When we used the sample quadratic classification rule under multivariate normal densities with unequal \( \Sigma_i \) for the same example, we obtained the confusion matrix as follows:

\[
\begin{array}{c|c|c|c}
& \Pi_1 & \Pi_2 & \Pi_3 \\
\hline
\Pi_1 & 29 & 1 & 0 & 30 \\
\Pi_2 & 1 & 28 & 6 & 35 \\
\Pi_3 & 0 & 9 & 31 & 40 \\
\end{array}
\]

and consequently, the estimate of the expected actual error rate = 0.1619. This rate is considerably higher than 0.1238. Thus, the discriminant function using Kotz type densities performs better and can achieve effective classification when the samples are from Kotz type populations.
4.4 Concluding Remarks

Before implementing a sample classification rule, multivariate normality of the data should be checked. If multivariate normality holds, then already well established procedures based on normality can be utilized for classification. If normality fails and Kotz type distribution assumption holds, then we can use the discrimination function rules based on Kotz type densities that we have developed in this chapter as an alternative procedure. If we are not sure about the appropriateness of using normal or Kotz type densities, rules based on both the densities can be constructed and the one that provides smaller error rate can be used.
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


APPENDIX

PROGRAM 1: This program is to compute the power comparison between Mardia’s skewness and kurtosis measures and Henze-Zirkler test.

```
options ls=70 nodate;
proc iml;
count1=0; count2=0; count3=0;
count4=0; count5=0; count6=0;
seed=0;
sample=5000;
n=100;
q1_95= 18.234;
q2_95= 1.366;
qtnb_5_95= 0.120;
qtnb1_95= 0.681;
qtnb3_95= 1.098;
qtnb_n_95= 0.985;
p=3;
kappal=j(sample,1,0);
kappa2=j(sample,1,0);
tnb_5=j(sample,1,0);
tnb1=j(sample,1,0);
tnb3=j(sample,1,0);
tnb_n=j(sample,1,0);
* compute beta for HZ statistic;
tt1= ( (2*p + 1)/4 )**(1/(p+4));
tt2=n**(1/(p+4));
beta_n=(1/sqrt(2))*tt1*tt2; print beta_n;
beta_5=0.5; print beta_5;
beta1=1; print beta1;
beta3=3; print beta3;
```
* generate new normal samples;
   do k=1 to sample;
      z=j(n,p,0);
      x=j(n,p,0);
      g_matrix=j(n,n,0); newg_matrix=j(n,n,0);
      x=normal(repeat(seed,n,p));
      dfchi=p*(p+1)*(p+2)/6;
      q=i(n) - (1/n)*j(n,n,1);
      s=(1/n)*t(x)*q*x;
      s_inv= inv(s);
      g_matrix=q*x*s_inv*t(x)*q;
      do i=1 to n;
         do j=1 to n;
            newg_matrix[i,j]=g_matrix[i,i] -2*g_matrix[i,j]+g_matrix[j,j];
         end;
      end;
      betahat=( sum(g_matrix#g_matrix#g_matrix) )/(n*n);
      beta2hat=trace( g_matrix#g_matrix )/n;
      kappa1[k]=n*betahat/6;
      kappa2[k]=(beta2hat - p*(p+2))/sqrt(8*p*(p+2)/n);
      if (kappa1[k] > q1_95) then
         count1=count1+1;
      if (kappa2[k] > q2_95) then
         count2=count2+1;
   * compute tnb,HZ statistic for beta0.5;
      term1=sum( exp( -0.5*(beta_5**2)*newg_matrix ) )/(n**2);
      t1=-(2/n)*(1+beta_5**2)**(-p/2);
      t2=trace( exp( -0.5*(beta_5**2/(1+beta_5**2))*g_matrix ) );
      term2=t1*t2;
      term3=(1+2*beta_5**2)**(-p/2);
      tnb_5[k]=n*(term1 + term2 + term3);
* compute tnb,HZ statistic for beta1;
  term1=sum( exp( -0.5*(beta1**2)*newg_matrix ) )/(n**2);
  t1=-(2/n)*(1+beta1**2)**(-p/2);
  t2=trace( exp( -0.5*(beta1**2/(1+beta1**2))*g_matrix ) );
  term2=t1*t2;
  term3=(1+2*beta1**2)**(-p/2);
  tnb1[k]=n*(term1 + term2 + term3);

* compute tnb,HZ statistic for beta3;
  term1=sum( exp( -0.5*(beta3**2)*newg_matrix ) )/(n**2);
  t1=-(2/n)*(1+beta3**2)**(-p/2);
  t2=trace( exp( -0.5*(beta3**2/(1+beta3**2))*g_matrix ) );
  term2=t1*t2;
  term3=(1+2*beta3**2)**(-p/2);
  tnb3[k]=n*(term1 + term2 + term3);

* compute tnb,HZ statistic for beta_n,;
  term1=sum( exp( -0.5*(beta_n**2)*newg_matrix ) )/(n**2);
  t1=-(2/n)*(1+beta_n**2)**(-p/2);
  t2=trace( exp( -0.5*(beta_n**2/(1+beta_n**2))*g_matrix ) );
  term2=t1*t2;
  term3=(1+2*beta_n**2)**(-p/2);
  tnb_n[k]=n*(term1 + term2 + term3);

if (tnb_5[k] > qtnb_5_95) then
  count3=count3+1;
if (tnb1[k] > qtnb1_95) then
  count4=count4+1;
if (tnb3[k] > qtnb3_95) then
  count5=count5+1;
if (tnb_n[k] > qtnb_n_95) then
  count6=count6+1;
end;

power_blp=countl/sample;
power_b2p=count2/sample;
power_tnb_5=count3/sample;
power_tnb1=count4/sample;
power_tnb3=count5/sample;
power_tnb_n=count6/sample;
title2 'Power for normal data';
print seed; print p;
print 'The number of' sample 'samples of size' n;
print 'by (1) using kappa1, kappa2, tnb';
print 'Based on skewness: ' count1 power_b1p;
print 'Based on kurtosis: ' count2 power_b2p;
print 'Based on Henze-zirkler: ' count3 power_tnb_5;
print 'Based on Henze-zirkler: ' count4 power_tnb1;
print 'Based on Henze-zirkler: ' count5 power_tnb3;
print 'Based on Henze-zirkler: ' count6 power_tnb_n;
quit;
PROGRAM 2: This program is to compute MLEs for $\mu$ and general covariance $\Sigma$.

options ls=70;
proc iml;
/* Log-likelihood function:f_ha */
start f_ha(x) global(data,optn,con,xbar,p,n);
t=j(n,1,0); v=j(p,p,0); d=j(p,p,0);
sum=0.;
pi=3.1416;
c=1./(2*pi);
k=1;
do i=1 to p;
do j=1 to p;
  if (i<=j) then do;
    v[i,j]=x[p+k];
    v[j,i]=x[p+k];
    k=k+1;
  end;
end;
end;
varcov=v;
* check if the varcov is pd;
call eigen(lambda,u,varcov);
d=diag(lambda);
do i=1 to p;
  if (d[i,i]<=0) then d[i,i]=0.0001;
end;
varcov=u*d*t(u);
inv_var=inv(varcov);
detsig=det(varcov);
do i=1 to p;
   xbar[i]=x[i];
end;
* use original data of X;
do i=1 to n;
   diff=data[,i] - xbar;
   tt= t(diff)*inv_var*diff;
   t[i]=sqrt( tt );
   sum=sum + t[i];
end;
f_ha=n*log(c) - (n/2)*log(detsig) - sum;
return(f_ha);
finish f_ha;
/* Main program */
n=30;
p=4;
cc=p*(p+1)/2;
par_ha=p+cc; par_ho1=cc; par_ho2=1+cc;
par_ho3=2+cc;
data=j(p,n,0); xbar=j(p,1,0);
x_ha0=j(par_ha,1,0); muopt_ha=j(p,1,0);
v1=j(cc,1,0); v=j(p,0,0);
a_hat=j(p,0,0); b_hat=j(p,0,0); asym_mu=j(p,0,0);
sigma2=0; rho=0;
sum1=0; sum2=0; sum3=0;
alpha=0.05;
x_ho10=j(par_ho1,1,0);
x_ho20=j(par_ho2,1,0); muop_ho2=j(1,1,0);
x_ho30=j(par_ho3,1,0); muop_ho3=j(2,1,0);
sigma=0;
asy_lmu1=j(p,0,0); h1=j(p-1,1,0);
asy_lmnu3=j(3,3,0); h3=j(3,1,0);
asyl1=j(p,1,0); lowl=j(p,1,0); up1=j(p,1,0);
asylm2=j(p,1,0); low2=j(p-1,1,0); up2=j(p-1,1,0);
asyl3=j(3,1,0); low3=j(3,1,0); up3=j(3,1,0);
s1=j(p,1,0);
* board data;
tdata={ 1889 1651 1561 1778,
   2403 2048 2087 2197,
   2119 1700 1815 2222,
   1645 1627 1110 1533,
   1976 1916 1614 1883,
   1712 1712 1439 1546,
   1943 1685 1271 1671,
   2104 1820 1717 1874,
   2983 2794 2412 2581,
   1745 1600 1384 1508,
   1710 1591 1518 1667,
   2046 1907 1627 1898,
   1840 1841 1595 1741,
   1867 1685 1493 1678,
   1859 1649 1389 1714,
   1954 2149 1180 1281,
   1325 1170 1002 1176,
   1419 1371 1252 1308,
   1828 1634 1602 1755,
   1725 1594 1313 1646,
   2276 2189 1547 2111,
   1899 1614 1422 1477,
   1633 1513 1290 1516,
   2061 1867 1646 2037,
   1856 1493 1356 1533,}
1727 1412 1238 1469,
2168 1896 1701 1834,
1655 1675 1414 1597,
2326 2301 2065 2234,
1490 1382 1214 1284};
data=t(tdata);
xbar=(1/n)*data*j(n,1,1);
q=i(n) - (1/n)*j(n,n,1);
ss=(1/(n-1))*data*q*tdata;
do i=1 to p;
   sigma2=sigma2 + ss[i,i];
end;
sigma2=sigma2/p;
* initial value of (mu,sigma) under Ha;
k=1;
do i=1 to p;
do j=1 to p;
   if (i<=j) then do;
      v1[k]=ss[i,j];
k=k+1;
   end;
end;
end;
x_ha0=xbar//v1;
tx_ha0=t(x_ha0);
optn={1 2}; * for maximizing ;
con_ha={. . . .
   1.e-6 . .
   1.e-6 .
   1.e-6 .
   1.e-6,
call nlprnra(rc,x_hares,"f_ha",x_ha0,optn,con_ha);

* Save result in xopt_ha, fopt_ha;

xopt_ha=x_hares;
fopt_ha=f_ha(xopt_ha);
print rc;
print xopt_ha;
print fopt_ha;

* Solve for muopt,sigmaopt under Ha;
do i=1 to p;
    muopt_ha[i]=xopt_ha[i];
end; sigop_ha=xopt_ha[p+1:par_ha]; print muopt_ha sigop_ha;

* Compute the optimal covariance using the optimal values of MU,SIGMA;
k=1; do i=1 to p;
    do j=1 to p;
        if (i<=j) then do;
            v[i,j]=xopt_ha[p+k];
            v[j,i]=xopt_ha[p+k];
            k=k+1;
        end;
    end;
end;
bigsopt=v;
print bigsopt;
quit;
PROGRAM 3: In this program, we are computing $\text{APER}$ and $\hat{E}(\text{AER})$ using Kotz type densities for unequal general covariances for $p=4$ and $g=2$.

```plaintext
options ls=70;
proc iml;
/* Log-likelihood function: f_ha_1 */
start f_ha_1(x)
   global(data1,data2,optn,
         con,p1,p2,n,n1,n2,g);
   v_1=j(p,p,0);  v_2=j(p,p,0);
   t1=j(n1,1,0);  t2=j(n2,1,0);
   sum1=0.; sum2=0.;
   xbar=j(p2,1,0);  d=j(p,p,0);
   pi=3.1416;
   c=1./(2*pi);
   * find Sigma1;
   k1=1;
   do i=1 to p;
      do j=1 to p;

         if (i<=j) then do;
            v_1[i,j]=x[p2+k1];
            v_1[j,i]=x[p2+k1];
            k1=k1+1;
         end;
      end;
   varcov1=v_1;
   * check if the varcov1 is pd;
   call eigen(lambda,u,varcov1);
   d=diag(lambda);
```
do i=1 to p;
if (d[i,i]<=0) then d[i,i]=0.0001;
end;
varcov1=u*d*t(u);
inv_var1= inv(varcov1);
detsig1=det(varcov1);
* find Sigma2;
cc=p*(p+1)/2;
k2=cc + 1;
do i=1 to p;
   do j=1 to p;
      if (i<=j) then do;
         v_2[i,j]=x[p2+k2];
         v_2[j,i]=x[p2+k2];
         k2=k2+1;
      end;
   end;
end;
varcov2=v_2;
* check if the varcov2 is pd;
call eigen(lambda,u,varcov2);
d=diag(lambda);
do i=1 to p;
   if (d[i,i]<=0) then d[i,i]=0.0001;
end;
varcov2=u*d*t(u);
inv_var2= inv(varcov2);
detsig2=det(varcov2);
do i=1 to p2;
   xbar[i]=x[i];
end;
* use original data of X;
do i=1 to n1;
    diff1=data1[,i] - xbar[1:p];
    tt1= t(diff1)*inv_var1*diff1;
    t1[i]=sqrt( tt1 );
    sum1=sum1 + t1[i];
end;
do i=1 to n2;
    diff2=data2[,i] - xbar[p+1:p2];
    tt2= t(diff2)*inv_var2*diff2;
    t2[i]=sqrt( tt2 );
    sum2=sum2 + t2[i];
end;
sum=sum1+sum2;
sum_sig=n1*log(detsig1) + n2*log(detsig2);
f_ha_1=n*log(c) - (1/2)*sum_sig - sum; * maximized fn ;
return(f_ha_1);
finish f_ha_1;
start check_dist(data,muopt,inv_sig1,inv_sig2,
   detsig1,detsig2,i,p,g1,g2,n12_m);
diff=j(p,1,0);
n12_m=0;
diff = data[,i] - muopt[,g1];
distsq1 = sqrt( t(diff)*inv_sig1*diff ) + 0.5*log(detsig1);
diff = data[,i] - muopt[,g2];
distsq2 = sqrt( t(diff)*inv_sig2*diff ) + 0.5*log(detsig2);
if (distsq2 < distsq1) then n12_m = 1;
finish check_dist;
start get_xbar_sigma(muopt_ha,inv_sig1,inv_sig2,
   detsig1,detsig2)
global(tdata1,tdata2,data1,data2,
cc,p,p2,n1,n2,g);
v1 = j(cc,1,0); v2 = j(cc,1,0);
muopt_ha=j(p,g,0);
v1v1=j(p,p,0); vv2=j(p,p,0);
xbar1=j(p,1,0);
xbar2=j(p,1,0);
* find sample means for 2 groups;
xbar1=data1[,+] / ncol(data1);
xbar2=data2[,+] / ncol(data2);
* find sample var-cov matrix;
q1=i(n1) - (1/n1)*j(n1,n1,1);
q2=i(n2) - (1/n2)*j(n2,n2,1);
ss1=(1/(n1-1))*data1*q1*tdata1;
ss2=(1/(n2-1))*data2*q2*tdata2;
ss= ( (n1-1)*ss1+(n2-1)*ss2)/(n1+n2-2);
* find the initial values for sigma2;
k1=1;
do i=1 to p;
   do j=1 to p;
      if (i<=j) then do;
         v1[k1]=ss1[i,j];
         k1=k1+1;
      end;
   end;
end;
k2=1;
do i=1 to p;
   do j=1 to p;
      if (i<=j) then do;
         v2[k2]=ss2[i,j];
         k2=k2+1;
      end;
end;
end;

end;

end;

* initial value of (mu11,..,mu15, mu41,...,mu45,
   sigma1,...,sigma3);

xbar=xbar1//xbar2;
sigma_po=v1//v2;
x_ha_10=xbar//sigma_po;
tx_ha_10=t(x_ha_10);
optn={1 0};         * for maximizing ;
con_ha={

   1.e-6 . .   
   1.e-6 . .   
   1.e-6 . 
   1.e-6 . 
   1.e-6 .  
   1.e-6 . 

   1.e-6 . .   
   1.e-6 . .   
   1.e-6 . 
   1.e-6 . 
   1.e-6, 


   . . .
   .    
   .    
   .    

}
call nlprna(rc,x_hares,"f_ha_1",x_ha_10,optn,con_ha);
* Save result in xopt, fopt;
xopt_ha=x_hares;
fopt_ha=f_ha_1(xopt_ha);
* Solve for mu1opt,...,mu2opt, sigma1opt,...,sigma3opt;
do i=1 to g;
  muopt_ha[,i]=xopt_ha[((i-1)*p)+1:p*i];
end;
* find Sigma1;
k1=1;
do i=1 to p;
do j=1 to p;
  if (i<j) then do;
    vv1[i,j]=xopt_ha[p2+k1];
    vv1[j,i]=xopt_ha[p2+k1];
k1=k1+1;
  end;
end;
end;
* find Sigma2;
cc=p*(p+1)/2;
k2=cc + 1;
do i=1 to p;
do j=1 to p;
  if (i<=j) then do;
    vv2[i,j]=xopt_ha[p2+k2];
  end;
vv2[j,i]=xopt_ha[p2+k2];
k2=k2+1;
end;
end;
sig1=vv1;
invsig1= inv(sig1);
detsig1=det(sig1);
sig2=vv2;
invsig2= inv(sig2);
detsig2=det(sig2);
finish get_xbar_sigma;
/* Main program */
g=2;
p=4; p2=2*p;
cc=p*(p+1)/2;
par_ha=p2+2*cc;
tdata1={
5.7806923 2.9711524 4.7236083 1.5028512,
5.4164124 2.8093463 4.1048688 1.3894683,
5.8224671 2.6413571 4.1156555 1.2004034,
5.915706 2.7520598 4.4367974 1.3211563,
7.0615442 3.1057031 5.0010325 1.4469908,
5.5535799 2.922983 4.5625164 1.4017037,
6.0405752 2.5955045 4.2735683 1.2460578,
5.9646149 2.6004935 3.6010724 1.1288722,
6.100764 2.7878842 4.1367692 1.2718204,
5.9264868 2.8248937 4.2337807 1.3092521,
5.6443527 2.3811418 4.2256821 1.2690787,
5.9006483 3.1013458 4.1848988 1.4086161,
6.6374569 2.7429361 4.1684296 1.1353632,
5.4346199 2.4490424 4.0686659 1.1920723,
<table>
<thead>
<tr>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.068537</td>
<td>2.7037714</td>
<td>4.3215518</td>
<td>1.3376282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5398378</td>
<td>2.9877456</td>
<td>4.1977785</td>
<td>1.4275273</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5792779</td>
<td>2.8897342</td>
<td>4.3448149</td>
<td>1.2068549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.7580576</td>
<td>2.6495915</td>
<td>4.3011515</td>
<td>1.2496377</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.7195076</td>
<td>2.7513127</td>
<td>3.8838301</td>
<td>1.2760272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2880995</td>
<td>2.7209261</td>
<td>4.5542073</td>
<td>1.4095898</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.6827247</td>
<td>2.9219009</td>
<td>4.2515982</td>
<td>1.4155149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9629738</td>
<td>2.8588019</td>
<td>4.3858777</td>
<td>1.3956366</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2534427</td>
<td>2.8480558</td>
<td>4.1071094</td>
<td>1.2321512</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2880453</td>
<td>2.7655088</td>
<td>4.1234011</td>
<td>1.2536415</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.4859478</td>
<td>3.2043381</td>
<td>4.2138811</td>
<td>1.366392</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.9656021</td>
<td>3.0616776</td>
<td>4.1505009</td>
<td>1.312617</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.8308631</td>
<td>2.8428823</td>
<td>4.0287503</td>
<td>1.2345323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3202378</td>
<td>2.9035702</td>
<td>4.2867455</td>
<td>0.9928669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.246117</td>
<td>3.2109154</td>
<td>4.0325683</td>
<td>1.3407778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.284534</td>
<td>2.5278402</td>
<td>3.9808005</td>
<td>1.1471642</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
5.0316135 1.8806569 4.4691248 2.1438683,
6.8310662 2.9633427 5.7297421 2.0299421,
5.6569992 2.8223333 4.8557879 1.9384321,
5.5815036 1.9986782 4.5067316 1.9977182,
7.2810534 3.317477 5.83851 1.8618643,
5.2051024 2.4717544 4.6737335 1.1315235,
6.8058637 3.1057605 5.9300564 2.0530363,
6.6884489 2.758086 5.5829555 1.9810277,
6.418012 3.0816724 5.6730993 1.9943875,
6.7143569 3.0428895 5.7003529 2.0546334,
6.3792509 2.8108183 5.6470089 1.7982288,
6.1085266 2.6748243 5.1651656 2.0292148,
4.4734502 3.2315595 3.769565 1.9606387,
5.9880721 2.8316502 5.1363329 2.1609173,
7.4433548 2.6297364 6.5193664 1.7712821};

sig1=j(p,p,0);
sig2=j(p,p,0);
n1=nrow(tdata1); n2=nrow(tdata2);
n=n1+n2;
data1=j(p,n1,0); data2=j(p,n2,0);
data1=t(tdata1);
data2=t(tdata2);
n12_mm = 0; n21_mm = 0;
run get_xbar_sigma(muopt_ha,invsig1, invsig2,
detsig1,detsig2);
sig1 = inv(invsig1);
sig2 = inv(invsig2);
print muopt_ha sig1 invsig1 sig2 invsig2 ;
do i = 1 to n1;
    run check_dist(data1,muopt_ha,invsig1, invsig2,
        detsig1,detsig2,i,p,1,2,n12_m);
n12_mm = n12_mm + n12_m;
end;
n11_t = n1 - n12_mm;
do i = 1 to n2;
   run check_dist(data2, muopt_ha, invsig2, invsig1,
                  detsig2, detsig1, i, p, 2, 1, n12_m);
   n21_mm = n21_mm + n12_m;
end;
n22_t = n2 - n21_mm;
count_mm = n12_mm + n21_mm;
aper = count_mm / n;
print 'The confusion matrix';
print 'Classify as';
print '   Pop1   Pop2';
print 'True Pop' n11_t n12_mm ;
print '   ' n21_mm n22_t ;
print 'APER' aper;
store_data1 = tdata1;
store_data2 = tdata2;
no1 = nrow(store_data1);
tdata1 = j(no1 - 1, p, 0);
n11_true = 0; n12_miss = 0;
do i = 1 to no1;
   k = 0;
do j = 1 to no1;
      if i ^= j then do;
         k = k + 1;
         tdata1[k,] = store_data1[j,];
      end;
end;
data1 = t(tdata1);
n1=ncol(data1); n=n1+n2;
run get_xbar_sigma(muopt_ha,invsig1,invsig2,
detsig1,detsig2);
run check_dist(t(store_data1),muopt_ha,invsig1,invsig2,
detsig1,detsig2,i,p,1,2,n12_m);
n12_miss = n12_miss + n12_m;
end;
n11_true = no1 - n12_miss ;
tdata1=store_data1;
data1=t(tdata1);
n1=ncol(data1);
no2=nrow(store_data2);
tdata2=j(no2-1,p,0);
n22_true = 0; n21_miss = 0;
do i = 1 to no2;
k = 0;
do j = 1 to no2;
   if i ^= j then do;
      k = k + 1;
      tdata2[k,] = store_data2[j,];
   end;
end;
data2=t(tdata2);
n2=ncol(data2); n=n1+n2;
run get_xbar_sigma(muopt_ha,invsig1,invsig2,
detsig1,detsig2);
run check_dist(t(store_data2),muopt_ha,invsig1,invsig2,
detsig1,detsig2,i,p,2,1,n12_m);
n21_miss = n21_miss + n12_m;
end;
n22_true = no2 - n21_miss ;
count_miss=n12_miss + n21_miss;
tot=n01 + n02;
ex_aer=count_miss/tot;
print 'The confusion matrix';
print 'Classify as';
print 'Pop1   Pop2';
print 'True Pop' n11_true n12_miss;
print '   ' n21_miss n22_true;
print 'The estimate of E(AER)' ex_aer;
quit;
VITA

Place of Birth: Bangkok, Thailand.

Education: 1997 M.S. in Statistics, North Carolina State University, Raleigh, NC.
           1986 M.S. in Applied Statistics, National Institute of Development Administration (NIDA), Bangkok, Thailand.
           1983 B.S. in Statistics, Kasetsart University, Bangkok, Thailand.

Experience: 2001 - 2003 Graduate Research Assistant, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA.
           1991 - present Lecturer, Department of Mathematics, Faculty of Sciences, Silpakorn University, Bangkok, Thailand.
           1994 - 1995 Invited Lecturer, Krungthep University, Bangkok, Thailand.

This document was prepared by the author using \LaTeX.