Summer 2003

Analysis of Multivariate Data Using Kotz Type Distribution

Kusaya Plungpongpun

Old Dominion University

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 N. Naik (Director)

N. Rao Chaganty (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

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp \; \{-[(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})]^{\frac{1}{2}}\},$$

where $\boldsymbol{\mu} \in \Re^p$, $\boldsymbol{\Sigma}$ is a positive definite matrix and $c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$. 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 generalized spatial median (GSM) estimators of C. R. Rao (1988).

(iii) Using the asymptotic distribution of the estimates, multivariate analysis of variance (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 expected 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

Page

1 Introduction **1**

1.1 Tests for Multivariate Normality 1

1.1.1 Power Studies 6

1.1.2 General Comments and Concluding Remarks 10

1.2 Overview of Thesis 17

2 A Kotz Type Distribution and Estimation of Parameters **19**

2.1 Introduction 19

2.1.1 Definition 19

2.1.2 Moments and Other Properties 21

2.1.3 Marginal and Conditional Distributions 22

2.2 Simulation Algorithm 23

2.3 Goodness-of-fit Tests for Kotz Type Distribution 26

2.4 Estimation of Parameters 27

2.4.1 Generalized Spatial Median (GSM) 27

2.4.2 Computation of GSM and $\widehat{\Sigma}$ 28

2.4.3 Computation of GSM and $\widehat{\Sigma}$ under Different Covariance Structures 30

2.5 Example 34

2.6 Comparisons of ML Estimates Between Kotz Type and Normal Populations 39

2.7 Concluding Remarks 45

3 Statistical Inference Using Kotz Type Distribution **46**

3.1 Introduction 46

3.2 Testing of Hypothesis for One Population Under the General Covariance Structure 46

3.3 Testing of Hypothesis for One Population Under AR(1) Structure . . . 48

Page

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

LIST OF TABLES

Table Page

1.1 The approximate and empirical percentage points of K_1 and K_2 of size $1-\alpha$ and sample sizes n for $p=3$. 7

1.2 The approximate and empirical percentage points of $T_{n,\beta}$ of size $1-\alpha$, sample sizes n and parameters β for $p=3$. 7

1.3 The approximate and empirical percentage points of K_1 and K_2 of size $1-\alpha$ and sample sizes n for $p=5$. 8

1.4 The approximate and empirical percentage points of $T_{n,\beta}$ of size $1-\alpha$, sample sizes n and parameters β for $p=5$. 8

1.5 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, χ^2 and gamma distributions. 11

1.6 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. 12

1.7 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, χ^2 and gamma distributions. 13

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. 14

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, χ^2 and gamma distributions. 15

Table Page

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 $Exp(1)$, 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 $\hat{\boldsymbol{\mu}}$ and $\bar{\mathbf{x}}$ and Pitman nearness probability of $\hat{\boldsymbol{\mu}}$ Relative to $\bar{\mathbf{x}}$ for Kotz type samples. 43

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

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

2.5 The estimates of bias and root MSE for $(p+1)\widehat{\boldsymbol{\Sigma}}$ and $\mathbf{S}_n$ and Pitman nearness probability of $\mathbf{S}_n$ Relative to $(p+1)\widehat{\boldsymbol{\Sigma}}$ 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,

The model journal used for this dissertation is *Statistica Sinica*.

developed by researchers over the years, into the following 5 categories.

(I) Procedures based on graphical plots or correlation coefficients (due to Healy (1968), Gnanadesikan and Kittenring (1972), Gnanadesikan (1977), Cox and Small (1978), Koziol (1982), Tsai and Koziol (1988), Ahn (1992), Singh (1993), Fang et al. (1998), Liang and Bentler (1999) and Beirlant et al. (1999)).

(II) Tests based on skewness and kurtosis measures (due to Mardia (1970), Malkovich and Afifi (1973), Small (1980), Isogai (1982), Bera and John (1983), Mardia and Foster (1983), Srivastava (1984), Koziol (1983, 1986, 1989), Isogai (1989), Mardia and Kent (1991), Móri et al. (1993), Jarque and Mckenzie (1995) and Kariya and George (1995)).

(III) Goodness-of-fit tests (due to Weiss (1958), Malkovich and Afifi (1973), Hawkins (1981), Moore and Stubblebine (1981), Koziol (1982), Royston (1983), Machado (1983), Fattorini (1986), Paulson et al. (1987), Ward (1988), Quiroz and Dudley (1991), Mudholkar et al. (1992), Romeu and Ozturk (1993, 1996), and Kariya and George (1995)).

(IV) Consistent procedures based on the empirical characteristic function (due to Csörgö (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' exension 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 $\mathbf{x}_j,\ j=1,...,n$ be a random sample from a multivariate normal distribution $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Then

$$\begin{aligned} \mathbf{z}_j &= \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}_j-\boldsymbol{\mu}),\ j=1,...,n \text{ are } \textit{i.i.d. } N_p(\mathbf{0},\mathbf{I}) \\ \text{and hence each} \quad \delta_j^2 &= \mathbf{z}_j^{'}\mathbf{z}_j \\ &= (\mathbf{x}_j-\boldsymbol{\mu})^{'}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_j-\boldsymbol{\mu}),\ j=1,...,n \end{aligned}$$

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

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

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 45^0 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

$$\begin{aligned} \beta_{1p} &= E[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})]^3, \\ \text{and } \beta_{2p} &= E[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^2, \end{aligned}$$

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

For a sample $\mathbf{x}_1,...,\mathbf{x}_n$ of size n, the estimates of β_{1p} and β_{2p} can be obtained as

$$\begin{aligned} \widehat{\beta}_{1p} &= b_{1p} = \frac{1}{n^2}\sum_{j,k=1}^{n}(\mathbf{r}_j'\mathbf{r}_k)^3 = \frac{1}{n^2}\sum_{j,k=1}^{n} g_{jk}^3, \\ \widehat{\beta}_{2p} &= b_{2p} = \frac{1}{n}\sum_{j=1}^{n} g_{jj}^2 = \frac{1}{n}\sum_{j=1}^{n} d_j^4, \end{aligned}$$

where $\mathbf{r}_j=\mathbf{S}_n^{-\frac{1}{2}}(\mathbf{x}_j-\bar{\mathbf{x}})$, $g_{jk}=(\mathbf{x}_j-\bar{\mathbf{x}})'\mathbf{S}_n^{-1}(\mathbf{x}_k-\bar{\mathbf{x}})$, $d_j=\sqrt{g_{jj}}$, $j,\ k=1,...,n$, $\bar{\mathbf{x}}=\frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$ and $\mathbf{S}_n=\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'$.

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,

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

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

$$D_{n,\beta} = \int_{\Re^p} \Big| \Psi_n(\mathbf{t}) - \exp\Big(-\frac{1}{2}\|\mathbf{t}\|^2\Big)\Big|^2 \varphi_\beta(\mathbf{t})\, d\mathbf{t},$$

where β denotes a positive number,

$$\Psi_n(\mathbf{t}) = \frac{1}{n}\sum_{j=1}^{n} \exp(i\mathbf{t}'\mathbf{r}_j) \qquad (\mathbf{t} \in \Re^p)$$

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

$$\varphi_\beta(\mathbf{t}) = \Big(2\pi\beta^2\Big)^{-\frac{p}{2}} \exp\Big(-\frac{\|\mathbf{t}\|^2}{2\beta^2}\Big), \ \mathbf{t} \in \Re^p,$$

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

Henze and Zirkler (1990) suggested the smoothing parameter β to be

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

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

$$\begin{aligned} D_{n,\beta} \ &= \ \frac{1}{n^2}\sum_{j,k=1}^{n} \exp\Big(-\frac{\beta^2}{2}\|\mathbf{r}_j - \mathbf{r}_k\|^2\Big) \\ &- \ 2(1+\beta^2)^{-\frac{p}{2}}\ \frac{1}{n}\sum_{j=1}^{n} \exp\Big(-\frac{\beta^2}{2(1+\beta^2)}\|\mathbf{r}_j\|^2\Big) + (1+2\beta^2)^{-\frac{p}{2}}, \end{aligned}$$

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

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

$$\begin{aligned} E\Big[T_\beta(p)\Big] \ &= \ 1 - (1+2\beta^2)^{-\frac{p}{2}}\Big[1 + \frac{p\beta^2}{1+2\beta^2} + \frac{p\,(p+2)\beta^4}{2(1+2\beta^2)^2}\Big] \quad \text{and} \\ Var\Big[T_\beta(p)\Big] \ &= \ 2(1+4\beta^2)^{-\frac{p}{2}} \ + \ 2(1+2\beta^2)^{-p}\Big[1 + \frac{2p\beta^4}{(1+2\beta^2)^2} + \frac{3p\,(p+2)\beta^8}{4(1+2\beta^2)^4}\Big] \end{aligned}$$

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

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 $\mu_{\beta,p} = E\Big[T_\beta(p)\Big]$ and $\sigma^2_{\beta,p} = Var\Big[T_\beta(p)\Big]$. That is, the $(1-\alpha)$ quantile of the asymptotic distribution of $T_{n,\beta}$, that is, of $T_\beta(p)$, is given by

$$q_{\beta,p}(1-\alpha) = \mu_{\beta,p}\Big(1 + \frac{\sigma^2_{\beta,p}}{\mu^2_{\beta,p}}\Big)^{-\frac{1}{2}} \exp\Big[\Phi^{-1}(1-\alpha)\sqrt{\log\Big(1+\frac{\sigma^2_{\beta,p}}{\mu^2_{\beta,p}}\Big)}\,\Big],$$

where $\Phi^{-1}(.)$ 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 β 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 β depends on n) is the χ^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 $p = 3$.

	K_1			K_2		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99
$n = 20$	13.196	15.428	20.654	0.234	0.559	1.245
$n = 30$	14.222	16.659	22.595	0.465	0.852	1.682
$n = 50$	14.988	17.518	23.439	0.702	1.132	2.067
$n = 100$	15.660	18.234	23.978	0.918	1.366	2.352
$q_{\beta,p}(1-\alpha)$	15.987	18.307	23.209	1.282	1.645	2.326

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

	$\beta = 0.5$			$\beta = 1$			$\beta = 3$			$\beta = \beta_3(n)$		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$n = 20$	0.097	0.112	0.145	0.613	0.670	0.790	1.054	1.092	1.172	0.756	0.819	0.953
$n = 30$	0.100	0.114	0.150	0.614	0.671	0.802	1.055	1.093	1.175	0.801	0.867	1.010
$n = 50$	0.102	0.118	0.151	0.616	0.676	0.805	1.057	1.094	1.169	0.852	0.921	1.067
$n = 100$	0.104	0.120	0.155	0.619	0.681	0.821	1.061	1.098	1.171	0.915	0.985	1.137
$q_{\beta,3}(1-\alpha)$	0.106	0.121	0.156	0.620	0.684	0.823	1.062	1.096	1.165			

n, p and α. We use the above obtained approximate cutoff points as an approximate critical value for a level α 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 α. We notice that four Henze-Kirkler tests based on different parameters β have a rejection rate below or close to the α level but the Mardia's skewness and kurtosis measures based on K_1 and K_2 tests have a rejection rate slightly above the α 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 $p = 5$.

	K_1			K_2		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99
$n = 20$	38.467	42.029	49.048	-0.176	0.102	0.643
$n = 30$	41.450	45.569	54.156	0.151	0.484	1.145
$n = 50$	43.576	47.733	56.797	0.468	0.833	1.595
$n = 100$	45.139	49.230	58.408	0.769	1.168	2.022
$q_{\beta,p}(1-\alpha)$	46.059	49.802	57.342	1.282	1.645	2.326

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

	$\beta = 0.5$			$\beta = 1$			$\beta = 3$			$\beta = \beta_5(n)$		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$n = 20$	0.186	0.200	0.229	0.813	0.843	0.905	1.008	1.018	1.044	0.880	0.909	0.971
$n = 30$	0.192	0.207	0.241	0.813	0.847	0.920	1.013	1.020	1.042	0.905	0.937	1.005
$n = 50$	0.195	0.211	0.246	0.814	0.847	0.919	1.014	1.021	1.037	0.932	0.963	1.026
$n = 100$	0.197	0.214	0.249	0.814	0.848	0.924	1.015	1.021	1.035	0.961	0.991	1.053
$q_{\beta,3}(1-\alpha)$	0.201	0.217	0.249	0.814	0.849	0.920	1.015	1.021	1.031			

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 ≥ 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 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 α 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 β. 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 β 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 $\mathbf{X}$, that is, K_1 test is consistent against all distribution satisfying $\beta_{1p} > 0$ (Henze and Zirkler, 1990).

The test based on K_2 is consistent against all alternative distributions satisfying $\beta_{2p} \neq p(p+2)$. 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 $T_{n,\beta}$. Tests of sizes $\alpha = 0.01$ against multivariate normal, Cauchy, t, χ^2 and gamma distributions.

Alternative	n	p=3						p=5					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
$N_p(\mathbf{0}, \mathbf{I})$	20	0.009	0.010	0.009	0.008	0.009	0.008	0.011	0.008	0.010	0.009	0.010	0.009
	30	0.011	0.011	0.010	0.009	0.009	0.011	0.012	0.011	0.009	0.009	0.009	0.008
	50	0.013	0.012	0.011	0.010	0.010	0.010	0.011	0.008	0.009	0.009	0.008	0.010
	100	0.008	0.010	0.008	0.009	0.011	0.009	0.009	0.008	0.010	0.011	0.009	0.010
p-variate Cauchy	20	0.961	0.979	0.969	0.975	0.935	0.973	0.995	0.998	0.995	0.992	0.892	0.990
	30	0.992	0.999	0.996	0.998	0.994	0.998	0.999	1.000	0.999	1.000	0.991	1.000
	50	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(2)	20	0.715	0.782	0.718	0.695	0.462	0.669	0.875	0.904	0.851	0.772	0.328	0.744
	30	0.872	0.937	0.893	0.886	0.710	0.868	0.970	0.988	0.969	0.948	0.673	0.938
	50	0.966	0.995	0.986	0.988	0.936	0.983	0.999	1.000	0.999	1.000	0.956	0.998
	100	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(5)	20	0.231	0.271	0.220	0.158	0.054	0.135	0.366	0.393	0.315	0.182	0.034	0.156
	30	0.360	0.442	0.346	0.259	0.093	0.214	0.552	0.637	0.502	0.323	0.062	0.283
	50	0.550	0.706	0.557	0.466	0.180	0.387	0.798	0.908	0.777	0.656	0.163	0.582
	100	0.731	0.950	0.829	0.793	0.439	0.707	0.940	0.996	0.971	0.952	0.489	0.913
p-variate $\chi^2(5)$	20	0.223	0.143	0.281	0.246	0.070	0.210	0.218	0.140	0.300	0.204	0.028	0.172
	30	0.481	0.263	0.602	0.518	0.140	0.436	0.487	0.291	0.604	0.428	0.044	0.349
	50	0.839	0.453	0.913	0.853	0.336	0.755	0.868	0.511	0.943	0.834	0.116	0.724
	100	1.000	0.743	1.000	0.998	0.820	0.987	1.000	0.840	1.000	1.000	0.441	0.992
p-variate $\chi^2(10)$	20	0.107	0.072	0.126	0.100	0.029	0.084	0.087	0.051	0.101	0.064	0.018	0.053
	30	0.202	0.119	0.259	0.187	0.036	0.142	0.188	0.117	0.233	0.127	0.023	0.099
	50	0.490	0.209	0.575	0.428	0.084	0.310	0.495	0.244	0.589	0.362	0.035	0.260
	100	0.922	0.394	0.953	0.861	0.269	0.686	0.954	0.461	0.975	0.837	0.085	0.656
p-variate G(2, 1)	20	0.311	0.193	0.381	0.344	0.109	0.302	0.303	0.195	0.393	0.285	0.040	0.251
	30	0.593	0.328	0.708	0.644	0.227	0.566	0.616	0.374	0.738	0.573	0.076	0.491
	50	0.930	0.546	0.968	0.935	0.503	0.879	0.953	0.647	0.981	0.933	0.204	0.871
	100	1.000	0.841	1.000	1.000	0.948	0.999	1.000	0.926	1.000	1.000	0.698	0.999
p-variate G(5, 1)	20	0.095	0.064	0.118	0.097	0.025	0.080	0.089	0.056	0.115	0.071	0.023	0.061
	30	0.207	0.110	0.258	0.187	0.041	0.147	0.190	0.111	0.243	0.129	0.021	0.105
	50	0.482	0.213	0.575	0.433	0.085	0.309	0.498	0.228	0.591	0.364	0.035	0.263
	100	0.920	0.395	0.951	0.862	0.268	0.685	0.957	0.469	0.976	0.862	0.092	0.675

Table 1.6: 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.

Alternative	n	p=3						p=5					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
p-variate	20	0.631	0.421	0.746	0.751	0.410	0.713	0.647	0.456	0.767	0.691	0.154	0.647
$Exp(1)$	30	0.894	0.609	0.959	0.956	0.733	0.935	0.931	0.713	0.974	0.946	0.380	0.923
	50	0.998	0.842	1.000	1.000	0.978	0.999	0.999	0.923	1.000	1.000	0.832	1.000
	100	1.000	0.990	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000
p-variate	20	0.169	0.200	0.161	0.125	0.063	0.110	0.188	0.197	0.162	0.098	0.029	0.089
$\frac{1}{2}N(0,1)$	30	0.283	0.332	0.278	0.220	0.113	0.194	0.327	0.387	0.280	0.170	0.049	0.147
$+\frac{1}{2}Exp(1)$	50	0.433	0.542	0.437	0.395	0.226	0.353	0.559	0.655	0.520	0.396	0.114	0.346
	100	0.689	0.846	0.742	0.755	0.612	0.733	0.828	0.930	0.847	0.800	0.362	0.745
p-variate	20	0.105	0.104	0.087	0.051	0.019	0.039	0.105	0.097	0.086	0.037	0.015	0.030
$\frac{1}{2}N(0,1)$	30	0.165	0.185	0.141	0.073	0.023	0.055	0.197	0.207	0.150	0.055	0.017	0.041
$+\frac{1}{2}t(5)$	50	0.277	0.346	0.237	0.128	0.032	0.090	0.337	0.403	0.259	0.106	0.028	0.076
	100	0.423	0.602	0.396	0.245	0.066	0.161	0.544	0.714	0.462	0.252	0.039	0.169
p-variate	20	0.299	0.168	0.427	0.476	0.201	0.441	0.280	0.161	0.427	0.379	0.065	0.339
$\frac{1}{2}N(0,1)$	30	0.565	0.254	0.752	0.775	0.432	0.738	0.572	0.289	0.771	0.712	0.146	0.648
$+\frac{1}{2}\chi^2(5)$	50	0.929	0.441	0.984	0.983	0.812	0.970	0.944	0.516	0.989	0.979	0.415	0.958
	100	1.000	0.728	1.000	1.000	0.999	1.000	1.000	0.820	1.000	1.000	0.945	1.000
p-variate	20	0.164	0.219	0.146	0.122	0.051	0.107	0.158	0.187	0.135	0.075	0.021	0.063
Kotz Type	30	0.229	0.337	0.224	0.181	0.089	0.164	0.266	0.375	0.219	0.140	0.037	0.128
	50	0.359	0.594	0.393	0.394	0.191	0.360	0.417	0.639	0.391	0.317	0.081	0.276
	100	0.461	0.879	0.662	0.736	0.482	0.699	0.617	0.924	0.740	0.727	0.256	0.657

Table 1.7: 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, χ^2 and gamma distributions.

Alternative	n	p=3						p=5					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
$N_p(\mathbf{0}, \mathbf{I})$	20	0.046	0.053	0.048	0.047	0.045	0.048	0.055	0.056	0.050	0.050	0.051	0.050
	30	0.052	0.054	0.050	0.046	0.044	0.046	0.051	0.050	0.048	0.043	0.046	0.043
	50	0.050	0.052	0.046	0.043	0.048	0.045	0.050	0.048	0.044	0.048	0.045	0.049
	100	0.050	0.054	0.050	0.050	0.045	0.044	0.058	0.046	0.050	0.046	0.045	0.048
p-variate Cauchy	20	0.980	0.992	0.982	0.983	0.957	0.982	0.997	0.998	0.998	0.995	0.942	0.994
	30	0.996	0.999	0.998	0.998	0.995	0.998	1.000	1.000	1.000	1.000	0.997	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(2)	20	0.833	0.890	0.836	0.799	0.609	0.777	0.936	0.958	0.926	0.859	0.524	0.840
	30	0.930	0.974	0.945	0.939	0.812	0.927	0.989	0.996	0.986	0.976	0.791	0.968
	50	0.985	0.999	0.995	0.994	0.965	0.993	1.000	1.000	1.000	1.000	0.974	0.999
	100	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(5)	20	0.413	0.426	0.377	0.288	0.127	0.251	0.543	0.592	0.499	0.311	0.120	0.278
	30	0.543	0.638	0.523	0.425	0.196	0.363	0.722	0.808	0.683	0.508	0.172	0.456
	50	0.675	0.845	0.691	0.614	0.318	0.548	0.890	0.964	0.881	0.775	0.303	0.707
	100	0.832	0.983	0.910	0.891	0.602	0.827	0.980	0.999	0.991	0.982	0.647	0.962
p-variate $\chi^2(5)$	20	0.436	0.298	0.534	0.479	0.198	0.434	0.407	0.305	0.509	0.393	0.112	0.355
	30	0.707	0.440	0.801	0.738	0.309	0.661	0.688	0.470	0.796	0.650	0.149	0.573
	50	0.954	0.641	0.981	0.955	0.571	0.911	0.996	0.702	0.987	0.939	0.295	0.876
	100	1.000	0.873	1.000	1.000	0.932	0.998	1.000	0.933	1.000	1.000	0.685	0.998
p-variate $\chi^2(10)$	20	0.253	0.181	0.297	0.245	0.100	0.217	0.216	0.167	0.271	0.184	0.072	0.170
	30	0.437	0.267	0.506	0.414	0.145	0.343	0.380	0.256	0.463	0.309	0.078	0.257
	50	0.721	0.374	0.780	0.659	0.230	0.539	0.717	0.414	0.802	0.600	0.115	0.476
	100	0.978	0.592	0.991	0.958	0.486	0.864	0.991	0.663	0.998	0.953	0.250	0.852
p-variate G(2, 1)	20	0.536	0.368	0.625	0.584	0.255	0.535	0.522	0.387	0.626	0.500	0.149	0.462
	30	0.802	0.535	0.881	0.827	0.425	0.771	0.809	0.578	0.900	0.775	0.202	0.711
	50	0.981	0.722	0.994	0.984	0.726	0.958	0.985	0.792	0.996	0.980	0.417	0.953
	100	1.000	0.935	1.000	1.000	0.987	1.000	1.000	0.974	1.000	1.000	0.856	1.000
p-variate G(5, 1)	20	0.253	0.183	0.292	0.242	0.096	0.216	0.219	0.170	0.266	0.195	0.075	0.178
	30	0.410	0.252	0.489	0.397	0.137	0.325	0.393	0.257	0.472	0.307	0.079	0.255
	50	0.703	0.368	0.767	0.653	0.229	0.540	0.721	0.414	0.798	0.599	0.123	0.480
	100	0.981	0.579	0.990	0.952	0.489	0.859	0.989	0.666	0.996	0.946	0.246	0.844

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.

Alternative	n	$p=3$						$p=5$					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
$Exp(1)$	20	0.816	0.595	0.898	0.891	0.623	0.868	0.819	0.654	0.904	0.849	0.367	0.820
	30	0.974	0.783	0.992	0.991	0.879	0.984	0.987	0.862	0.997	0.986	0.602	0.975
	50	1.000	0.936	1.000	1.000	0.995	1.000	1.000	0.974	1.000	1.000	0.929	1.000
	100	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate	20	0.332	0.383	0.318	0.247	0.158	0.224	0.378	0.420	0.339	0.211	0.112	0.194
$\frac{1}{2}N(0,1)$	30	0.444	0.533	0.434	0.362	0.229	0.333	0.522	0.602	0.480	0.321	0.136	0.288
$+\frac{1}{2}Exp(1)$	50	0.617	0.754	0.627	0.594	0.424	0.562	0.720	0.816	0.687	0.570	0.264	0.526
	100	0.813	0.942	0.857	0.883	0.775	0.874	0.910	0.980	0.927	0.903	0.538	0.865
p-variate	20	0.215	0.234	0.195	0.137	0.070	0.120	0.231	0.241	0.204	0.109	0.070	0.099
$\frac{1}{2}N(0,1)$	30	0.305	0.347	0.273	0.184	0.086	0.146	0.347	0.382	0.299	0.151	0.071	0.129
$+\frac{1}{2}t(5)$	50	0.408	0.508	0.375	0.252	0.111	0.191	0.487	0.576	0.418	0.224	0.080	0.174
	100	0.561	0.752	0.549	0.402	0.160	0.299	0.691	0.840	0.640	0.417	0.111	0.307
p-variate	20	0.564	0.334	0.688	0.716	0.408	0.687	0.500	0.328	0.666	0.600	0.200	0.566
$\frac{1}{2}N(0,1)$	30	0.795	0.447	0.920	0.919	0.649	0.891	0.786	0.493	0.915	0.875	0.324	0.843
$+\frac{1}{2}\chi^2(5)$	50	0.983	0.625	0.998	0.997	0.931	0.993	0.987	0.710	0.999	0.995	0.655	0.989
	100	1.000	0.860	1.000	1.000	0.999	1.000	1.000	0.920	1.000	1.000	0.983	1.000
p-variate	20	0.334	0.417	0.300	0.248	0.137	0.230	0.351	0.424	0.308	0.193	0.104	0.179
Kotz Type	30	0.425	0.580	0.416	0.359	0.204	0.326	0.464	0.600	0.408	0.279	0.117	0.243
	50	0.542	0.795	0.582	0.570	0.350	0.529	0.628	0.833	0.607	0.502	0.195	0.452
	100	0.662	0.965	0.819	0.865	0.663	0.845	0.779	0.986	0.863	0.852	0.425	0.802

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, χ^2 and gamma distributions.

Alternative	n	p=3						p=5					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
$N_p(\mathbf{0}, \mathbf{I})$	20	0.100	0.102	0.096	0.100	0.096	0.098	0.097	0.100	0.094	0.096	0.100	0.094
	30	0.102	0.099	0.097	0.098	0.098	0.097	0.097	0.096	0.095	0.092	0.094	0.094
	50	0.095	0.094	0.096	0.096	0.096	0.098	0.101	0.092	0.098	0.100	0.106	0.099
	100	0.099	0.103	0.100	0.098	0.092	0.098	0.091	0.087	0.100	0.095	0.096	0.096
p-variate Cauchy	20	0.986	0.995	0.990	0.989	0.965	0.988	0.998	0.999	0.998	0.998	0.966	0.996
	30	0.998	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.998	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(2)	20	0.869	0.925	0.865	0.836	0.663	0.823	0.957	0.975	0.948	0.893	0.638	0.877
	30	0.956	0.984	0.961	0.956	0.846	0.943	0.993	0.999	0.992	0.984	0.842	0.977
	50	0.990	1.000	0.997	0.997	0.974	0.996	1.000	1.000	1.000	0.999	0.984	0.999
	100	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate t(5)	20	0.506	0.580	0.483	0.375	0.202	0.340	0.644	0.701	0.588	0.403	0.206	0.367
	30	0.630	0.743	0.615	0.509	0.267	0.453	0.806	0.881	0.760	0.587	0.256	0.539
	50	0.770	0.899	0.773	0.698	0.409	0.632	0.922	0.977	0.913	0.827	0.397	0.776
	100	0.879	0.990	0.942	0.925	0.690	0.880	0.986	0.999	0.996	0.988	0.730	0.978
p-variate $\chi^2(5)$	20	0.566	0.408	0.647	0.596	0.278	0.552	0.537	0.412	0.638	0.512	0.200	0.465
	30	0.795	0.532	0.867	0.808	0.423	0.751	0.797	0.584	0.878	0.761	0.247	0.692
	50	0.973	0.716	0.988	0.974	0.680	0.940	0.980	0.792	0.994	0.964	0.412	0.925
	100	1.000	0.922	1.000	1.000	0.959	0.999	1.000	0.958	1.000	1.000	0.772	0.999
p-variate $\chi^2(10)$	20	0.372	0.281	0.426	0.360	0.169	0.323	0.334	0.264	0.382	0.291	0.149	0.267
	30	0.542	0.357	0.614	0.520	0.228	0.451	0.515	0.368	0.579	0.449	0.151	0.385
	50	0.812	0.472	0.860	0.768	0.340	0.664	0.821	0.527	0.884	0.708	0.204	0.599
	100	0.992	0.679	0.995	0.978	0.630	0.921	0.997	0.755	0.999	0.970	0.351	0.898
p-variate G(2, 1)	20	0.668	0.476	0.750	0.701	0.360	0.647	0.621	0.485	0.721	0.605	0.236	0.565
	30	0.864	0.613	0.926	0.893	0.544	0.849	0.874	0.675	0.940	0.861	0.304	0.811
	50	0.992	0.801	0.997	0.992	0.814	0.977	0.995	0.867	1.000	0.992	0.537	0.972
	100	1.000	0.959	1.000	1.000	0.993	1.000	1.000	0.988	1.000	1.000	0.917	1.000
p-variate G(5, 1)	20	0.362	0.272	0.407	0.352	0.178	0.322	0.328	0.269	0.395	0.303	0.144	0.276
	30	0.546	0.351	0.614	0.513	0.224	0.441	0.513	0.364	0.593	0.450	0.157	0.393
	50	0.811	0.480	0.864	0.762	0.339	0.657	0.809	0.523	0.877	0.708	0.199	0.602
	100	0.990	0.674	0.996	0.978	0.620	0.915	0.996	0.758	0.999	0.975	0.359	0.910

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 $Exp(1)$, contaminated normal and $t(5)$, contaminated normal and $\chi^2(5)$, and Kotz type distributions.

Alternative	n	p=3						p=5					
		K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$	K_1	K_2	$T_{0.5}$	T_1	T_3	$T_{n,\beta}$
$Exp(1)$	20	0.892	0.696	0.945	0.936	0.725	0.923	0.895	0.757	0.950	0.905	0.480	0.884
	30	0.989	0.852	0.997	0.994	0.923	0.990	0.993	0.901	0.999	0.994	0.717	0.989
	50	1.000	0.962	1.000	1.000	0.996	1.000	1.000	0.985	1.000	1.000	0.956	1.000
	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
p-variate	20	0.438	0.517	0.417	0.353	0.236	0.329	0.479	0.539	0.429	0.293	0.185	0.270
$\frac{1}{2}N(0,1)$	30	0.558	0.668	0.540	0.476	0.340	0.442	0.636	0.712	0.585	0.432	0.231	0.395
$+\frac{1}{2}Exp(1)$	50	0.705	0.838	0.708	0.676	0.515	0.648	0.798	0.882	0.771	0.656	0.345	0.609
	100	0.868	0.972	0.912	0.926	0.842	0.920	0.949	0.990	0.978	0.932	0.651	0.908
p-variate	20	0.299	0.322	0.280	0.203	0.114	0.180	0.320	0.337	0.298	0.180	0.122	0.162
$\frac{1}{1}N(0,1)$	30	0.382	0.439	0.359	0.265	0.148	0.224	0.437	0.480	0.382	0.222	0.127	0.190
$+\frac{1}{2}t(5)$	50	0.505	0.605	0.469	0.347	0.184	0.285	0.587	0.663	0.529	0.307	0.149	0.259
	100	0.641	0.826	0.636	0.505	0.242	0.393	0.762	0.894	0.722	0.503	0.192	0.394
p-variate	20	0.671	0.434	0.803	0.819	0.523	0.793	0.618	0.449	0.763	0.709	0.334	0.677
$\frac{1}{1}N(0,1)$	30	0.885	0.566	0.956	0.953	0.761	0.937	0.879	0.596	0.959	0.932	0.452	0.904
$+\frac{1}{2}\chi^2(5)$	50	0.997	0.727	0.999	0.999	0.958	0.998	0.996	0.786	1.000	0.997	0.749	0.995
	100	1.000	0.921	1.000	1.000	1.000	1.000	1.000	0.955	1.000	1.000	0.993	1.000
p-variate	20	0.448	0.552	0.415	0.336	0.204	0.303	0.460	0.534	0.406	0.265	0.174	0.248
Kotz Type	30	0.543	0.713	0.539	0.468	0.285	0.432	0.580	0.713	0.519	0.368	0.192	0.336
	50	0.631	0.868	0.666	0.657	0.436	0.616	0.725	0.903	0.697	0.589	0.276	0.548
	100	0.751	0.987	0.891	0.919	0.750	0.895	0.856	0.992	0.913	0.891	0.507	0.848

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

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})]^{\frac{1}{2}}\},$$

where $\boldsymbol{\mu} \in \Re^p$, $\boldsymbol{\Sigma}$ is a positive definite matrix and $c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$.

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

$$f(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = c \mid \Sigma \mid^{-\frac{1}{2}} \exp \ \{-[(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{\frac{1}{2}}\}, \ \boldsymbol{\mu} \in \Re^p, \ \Sigma \ p.\ d. \qquad (2.1)$$

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

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

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

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

$$\phi_{\mathbf{x}-\boldsymbol{\mu}}(t) = \exp(i\mathbf{t}'\boldsymbol{\mu})\psi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$$

for some function ψ.

Therefore, the elliptically symmetric distributions denoted by $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, have the *pdf* in the form

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

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

$$k_p = \frac{p\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}}.$$

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 $PE_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$, with $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and β, where $\boldsymbol{\mu} \in \Re^p$, $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite symmetric matrix and $\beta \in (0, \infty)$. Its density function is

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

where $k = \frac{p\Gamma(\frac{p}{2})}{\pi^{\frac{p}{2}}\Gamma(1+\frac{p}{2\beta})2^{1+\frac{p}{2\beta}}}$.

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

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

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

$g(u) = c_p u^{N-1} \exp(-ru^s)$, $r, s > 0$, $2N + p > 2$ then we say that $\mathbf{x}$ possesses a symmetric Kotz distribution. The *pdf* of $\mathbf{x}$ is given by

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

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

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 refered 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, $\mathbf{x}_1, ..., \mathbf{x}_n$ 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 formulae in Baringhaus and Henze (1992).

$$\begin{aligned} E(\mathbf{x}) &= \boldsymbol{\mu}, \\ Var(\mathbf{x}) &= (p+1)\boldsymbol{\Sigma}, \\ \beta_{1p} &= 0, \text{ Mardia's multivariate skewness measure,} \\ \beta_{2p} &= \frac{p(p+2)(p+3)}{(p+1)}, \text{ Mardia's multivariate kurtosis measure.} \end{aligned}$$

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

$$\psi_{\mathbf{x}}(\mathbf{t}) = \frac{1}{\Gamma(p)} e^{i\mathbf{t}'\boldsymbol{\mu}} \int_0^\infty \Psi_p(r\sqrt{\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}) r^{p-1} e^{-r} dr,$$

where $\Psi_1(x) = \cos x$, $\Psi_p(x) = \frac{\Gamma(\frac{p}{2})}{\pi^{\frac{1}{2}}\Gamma(\frac{p-1}{2})} \int_0^\pi \exp\{ix\cos\theta\} \sin^{p-2}\theta d\theta$, 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 $\mathbf{x}$ is partitioned as $\mathbf{x} = (\mathbf{x}'_{(1)}, \mathbf{x}'_{(2)})'$, where $\mathbf{x}_{(1)} = (\mathbf{x}_1, ..., \mathbf{x}_k)'$, $\mathbf{x}_{(2)} = (\mathbf{x}_{k+1}, ..., \mathbf{x}_p)'$ with $k < p$ and similarly, $\boldsymbol{\mu} = (\boldsymbol{\mu}'_{(1)}, \boldsymbol{\mu}'_{(2)})'$ with $\boldsymbol{\mu}_{(1)} = (\mu_1, ..., \mu_k)'$ and $\boldsymbol{\mu}_{(2)} = (\mu_{k+1}, ..., \mu_p)'$. Further suppose $\boldsymbol{\Sigma}$ is accordingly partitioned as $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, where $\boldsymbol{\Sigma}_{11}$ is a $k \times k$ p. d. matrix and $\boldsymbol{\Sigma}_{22}$ is a $p-k \times p-k$ p. d. matrix and $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$. Then

1. $\mathbf{x}_{(1)}$ has an elliptically symmetric $EC_k(\boldsymbol{\mu}_{(1)}, \boldsymbol{\Sigma}_{11}, g_1)$ distribution with

   $$g_1(\mathbf{t}) = \mathbf{t}^{\frac{p-k}{2}} \int_0^1 \omega^{\frac{k-p}{2}-1} (1-\omega)^{\frac{p-k}{2}-1}\, e^{-\sqrt{\frac{\mathbf{t}}{\omega}}}\, d\omega.$$

   The marginal characteristics of $\mathbf{x}_{(1)}$ are

   $$\begin{aligned} E(\mathbf{x}_{(1)}) &= \boldsymbol{\mu}_{(1)}, \\ Var(\mathbf{x}_{(1)}) &= (p+1)\boldsymbol{\Sigma}_{11}, \\ \beta_{1p}(\mathbf{x}_{(1)}) &= 0, \text{ and} \\ \beta_{2p}(\mathbf{x}_{(1)}) &= \frac{k(k+2)(p+3)}{p+1}. \end{aligned}$$

2. The conditional distribution of $\mathbf{x}_{(2)}$ given $\mathbf{x}_{(1)}$ is elliptically contoured $EC_{p-k}(\boldsymbol{\mu}_{2.1},\ \boldsymbol{\Sigma}_{22.1},\ g_{2.1})$, where

$$\begin{aligned}
\boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}_{(2)} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)}), \\
\boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}, \quad \text{and} \\
g_{2.1}(t) &= \exp\ \{-[t + (\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)})'\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_{(1)} - \boldsymbol{\mu}_{(1)})]^{\frac{1}{2}}\}.
\end{aligned}$$

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 $\mathbf{y}' = (y_1, ..., y_p)$ having the density

$$f(\mathbf{y}) = c\ \exp\Big\{-\sqrt{\mathbf{y}'\mathbf{y}}\Big\},$$

where $-\infty < y_i < \infty$, and $c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$. Note that $f(\mathbf{y})$ is the standardized version of Kotz type distribution given in (2.1).

The simulation of $\mathbf{y}$ is achieved by using the polar coordinate transformation,

$$\begin{aligned}
y_1 &= R\cos\theta_1 \\
y_2 &= R\sin\theta_1\cos\theta_2 \\
&\vdots \\
y_{p-1} &= R\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\cos\theta_{p-1} \\
y_p &= R\sin\theta_1\sin\theta_2\cdots\sin\theta_{p-2}\sin\theta_{p-1},
\end{aligned}$$

where $\theta_j \in [\ 0, \pi)$ for $1 \le j \le p-2$ 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 $\theta_j,\ j = 1, ..., p-1,$ are independently distributed with the probability density function given by

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

$$
\begin{aligned}
g(\theta_1) &= \frac{p-2}{2}\left[\frac{(p-4)\cdots 3\cdot 1}{(p-3)\cdots 4\cdot 2}\right]\sin^{p-2}(\theta_1)\\
g(\theta_2) &= \frac{2^{p-3}}{\pi}\frac{\left[(\frac{p-3}{2})!\right]^2}{(p-3)!}\sin^{p-3}(\theta_2)\\
g(\theta_3) &= \frac{p-4}{2}\left[\frac{(p-6)\cdots 3\cdot 1}{(p-5)\cdots 4\cdot 2}\right]\sin^{p-4}(\theta_3)\\
g(\theta_4) &= \frac{2^{p-5}}{\pi}\frac{\left[(\frac{p-5}{2})!\right]^2}{(p-5)!}\sin^{p-5}(\theta_4)\\
&\vdots\\
g(\theta_{p-2}) &= \frac{1}{2}\sin(\theta_{p-2})\\
g(\theta_{p-1}) &= \frac{1}{2\pi}.
\end{aligned}
$$

For an even number p, R and θ_j, $j = 1, ..., p-1$, are independently distributed with the probability density function given by

$$
\begin{aligned}
g(r) &= \frac{1}{\Gamma(p)}\, r^{p-1}e^{-r}, \text{ that is, } R \sim G(p,1) \text{ and}\\
g(\theta_1) &= \frac{2^{p-2}}{\pi}\frac{\left[(\frac{p-2}{2})!\right]^2}{(p-2)!}\sin^{p-2}(\theta_1)\\
g(\theta_2) &= \frac{p-3}{2}\left[\frac{(p-5)\cdots 3\cdot 1}{(p-4)\cdots 4\cdot 2}\right]\sin^{p-3}(\theta_2)\\
g(\theta_3) &= \frac{2^{p-4}}{\pi}\frac{\left[(\frac{p-4}{2})!\right]^2}{(p-4)!}\sin^{p-4}(\theta_3)\\
g(\theta_4) &= \frac{p-5}{2}\left[\frac{(p-7)\cdots 3\cdot 1}{(p-6)\cdots 4\cdot 2}\right]\sin^{p-5}(\theta_4)\\
&\vdots\\
g(\theta_{p-2}) &= \frac{1}{2}\sin(\theta_{p-2})\\
g(\theta_{p-1}) &= \frac{1}{2\pi}.
\end{aligned}
$$

We note that $E(\mathbf{y}) = \mathbf{0}$ and $Var(\mathbf{y}) = (p+1)\mathbf{I}$.

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 θ

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

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

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

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

where

$$\begin{aligned} y_1 &= R\cos\theta_1 \\ y_2 &= R\sin\theta_1\cos\theta_2 \\ y_3 &= R\sin\theta_1\sin\theta_2\cos\theta_3 \\ y_4 &= R\sin\theta_1\sin\theta_2\sin\theta_3\cos\theta_4 \\ y_5 &= R\sin\theta_1\sin\theta_2\sin\theta_3\sin\theta_4 \end{aligned}$$

and R and θ_j are independently distributed with

$$\begin{aligned} g(r) &= \frac{1}{24}r^4e^{-r}, \text{ that is, } R \sim G(5,1) \text{ and} \\ g(\theta_1) &= \frac{3}{4}\sin^3(\theta_1) \\ g(\theta_2) &= \frac{2}{\pi}\sin^2(\theta_2) \\ g(\theta_3) &= \frac{1}{2}\sin(\theta_3) \\ g(\theta_4) &= \frac{1}{2\pi} \end{aligned}$$

where $\theta_j \in [\,0,\pi)$ for $j = 1,2,3$ and $\theta_4 \in [\,0,2\pi)$. Then we obtain $\mathbf{x}$ by making the transformation $\mathbf{x} = \mathbf{\Gamma y} + \boldsymbol{\mu}$ for fixed $\boldsymbol{\mu}$ and $\mathbf{\Gamma}$.

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 χ^2 variates. That is,

$$nb_{1p} \xrightarrow{\mathcal{D}} \alpha_1\chi^2_p \;+\; \alpha_2\chi^2_{\frac{p(p-1)(p+4)}{6}},$$

where

$$\begin{aligned}\alpha_1 &= \frac{3}{p}\Big[\frac{m_6}{p+2} - 2m_4 + p(p+2)\Big] \text{ and} \\ \alpha_2 &= \frac{6m_6}{p(p+2)(p+4)}.\end{aligned}$$

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

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

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

For the distribution given in (2.1),

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

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)

$$f(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = c \mid \Sigma \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})]^{\frac{1}{2}}\},\ \boldsymbol{\mu} \in \Re^p,\ \Sigma\ p.d.$$

First, we note that this distribution has fatter tail regions than that of multivariate normal distribution. Secondly, if we observe $\mathbf{x}_1, ..., \mathbf{x}_n$ *iid* sample from this distribution then the joint *pdf* of $\mathbf{x}_1, ..., \mathbf{x}_n$ 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 $\boldsymbol{\mu}$ and Σ are not the same as those in the multivariate normal case. But for those elliptically symmetric distributions for which the joint *pdf* of $\mathbf{x}_1, ..., \mathbf{x}_n$ is also elliptically symmetric, the MLEs of $\boldsymbol{\mu}$ and Σ (except for a constant) are the same as those for multivariate normal distribution. See Fang and Anderson (1990).

Let $\mathbf{x}_1, ..., \mathbf{x}_n$ be a random sample from Kotz type distribution. Then the log-likelihood function is given by

$$\ln\, L(\boldsymbol{\mu}, \Sigma) \;=\; n\, \ln c - \frac{n}{2}\, \ln |\Sigma| - \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}$$

The MLEs of $\boldsymbol{\mu}$ and Σ are obtained by minimizing

$$\frac{n}{2} \ln \mid \Sigma \mid + \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i-\boldsymbol{\mu})}$$

w.r.t. $\boldsymbol{\mu}$ and Σ simultaneously.

When $\Sigma = \mathbf{I}$, the solution to the above problem or the MLE of $\boldsymbol{\mu}$ is the spatial median introduced by Haldane (1948) and for general Σ 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 $\boldsymbol{\mu}$ of the Kotz type distribution given in (2.1). Haldane (1948) defined the spatial median of the points $\mathbf{x}_1, ..., \mathbf{x}_n$, as a point $\hat{\boldsymbol{\mu}} \in \Re^p$ which minimizes

$$\sum_{i=1}^{n} \|\mathbf{x}_i - \boldsymbol{\mu}\| = \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i-\boldsymbol{\mu})'(\mathbf{x}_i-\boldsymbol{\mu})}$$

with respect to $\boldsymbol{\mu}$. For $p > 1$, the vector $\hat{\boldsymbol{\mu}}$ 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 $\hat{\boldsymbol{\mu}}$ which minimizes

$$\sum_{i=1}^{n} \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})'\mathbf{S}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})} \tag{2.2}$$

with respect to $\boldsymbol{\mu}$, where $\mathbf{S}$ is the usual sample variance covariance matrix.

2. A vector $\hat{\boldsymbol{\mu}}$ which minimizes

$$\frac{n}{2} \ln|\Sigma| + \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})} \tag{2.3}$$

simultaneously with respect to $\boldsymbol{\mu}$ and Σ.

Thus, we note that the MLE of $\boldsymbol{\mu}$ under the assumption of Kotz type distribution of $\mathbf{x}_1, ..., \mathbf{x}_n$ is same as the generalized spatial median defined by Rao (1988).

2.4.2 Computation of GSM and $\widehat{\Sigma}$

Let $\mathbf{x}_1, ..., \mathbf{x}_n$ be a random sample from (2.1). Then the log-likelihood function can be written as

$$\ln\, L(\boldsymbol{\mu}, \Sigma) = n\, \ln c - \frac{n}{2} \ln|\Sigma| - \sum_{i=1}^{n} \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})}. \tag{2.4}$$

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

$$\mathbf{y}_i = G^{-1}\mathbf{x}_i \quad \text{and} \quad \boldsymbol{\nu} = G^{-1}\boldsymbol{\mu},$$

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

$$\hat{\boldsymbol{\mu}} = G\hat{\boldsymbol{\nu}}.$$

Next the MLE of $\boldsymbol{\Sigma}$ is obtained as the matrix $\widehat{\boldsymbol{\Sigma}}$ which minimizes (2.3) with respect to $\boldsymbol{\Sigma}$ as a solution to the non-linear equations given by

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

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

Step 1. Take the initial values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, say $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\widehat{\boldsymbol{\Sigma}} = \mathbf{S}$ (for general covariance structure) or when $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a structured covariance matrix depending on the parameter vector $\boldsymbol{\theta}$, take $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the vector of initial estimators obtained using the sample. For example, if $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\sigma, \delta)$, (equicorrelation structure with σ as the common standard deviation and δ 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 $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$.

Step 3. Taking $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ as the initial values repeat Step 2. Continue until termination rules for all the estimators are met.

For example, for equicorrelation structure, continue until

$$\begin{aligned} \| \hat{\boldsymbol{\mu}}_{i+1} - \hat{\boldsymbol{\mu}}_i \| &\leq \epsilon_1, \\ | \hat{\sigma}_{i+1}\hat{\sigma}_i^{-1} - 1| &\leq \epsilon_2, \\ | \hat{\delta}_{i+1} - \hat{\delta}_i | &\leq \epsilon_3, \end{aligned}$$

where $\epsilon_1 = \epsilon_2 = \epsilon_3 = 10^{-3}$, (say) and $\hat{\boldsymbol{\mu}}_i, \hat{\sigma}_i$, and $\hat{\delta}_i$ are the estimators of $\boldsymbol{\mu}_i, \sigma_i$, and δ_i respectively at the i^{th} iteration.

2.4.3 Computation of GSM and $\widehat{\Sigma}$ under Different Covariance Structures

The computation of GSM can become easy under simpler structures of Σ. In the following we present computation formulae for different structures.

The General Covariance Structure:

$$\text{Suppose } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{pp} \end{pmatrix} = (\sigma_{ij}),$$

where for $i, j = 1, ..., p$, $\sigma_{ij} > 0$ if $i = j$ and $\sigma_{ij} \in \Re$ 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.

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

where

$$Q_i = (\frac{x_{1i}-\hat{\mu}_1}{\sigma_1})^2 - 2\rho(\frac{x_{1i}-\hat{\mu}_1}{\sigma_1})(\frac{x_{2i}-\hat{\mu}_2}{\sigma_2}) + (\frac{x_{2i}-\hat{\mu}_2}{\sigma_2})^2,\ i = 1, ..., n.$$

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:

$$\text{The structure} \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{pmatrix},$$

where $\sigma_{ii} > 0$ for $i = 1, ..., p$ 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 δ. The covariance matrix has the form,

$$\begin{aligned} \mathbf{\Sigma} &= \sigma^2 \left[(1-\delta)\mathbf{I}_p + \delta \mathbf{J}_p\right] \\ &= \sigma^2 \begin{pmatrix} 1 & \delta & \dots & \delta \\ \delta & 1 & \dots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \dots & 1 \end{pmatrix} = \sigma^2 V(\delta). \end{aligned}$$

The determinant and inverse of $\mathbf{\Sigma}$ respectively are

$$|\mathbf{\Sigma}| = (\sigma^2)^p \left[(1-\delta)^{p-1}(1+(p-1)\delta)\right] \qquad \text{and}$$

$$\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2(1-\delta)}[\mathbf{I}_p - \frac{\delta}{(1-(p-1)\delta)}\mathbf{J}_p],$$

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

Naik (1993) derived the estimators of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$; that is, of $\boldsymbol{\mu}$, σ and δ by maximizing the log-likelihood function of the distribution with respect to $\boldsymbol{\mu}$, σ and δ. The MLE of σ and δ were obtained by solving the following non-linear equations:

$$[1-(p-1)(1-\hat{\delta})]\sum_{i=1}^{n}\sqrt{Q_i(\hat{\delta})}+$$

$$[1+(p-1)\hat{\delta}](1-\hat{\delta})\sum_{i=1}^{n}\frac{(p-1)(\mathbf{x}_i-\hat{\boldsymbol{\mu}})'(\mathbf{x}_i-\hat{\boldsymbol{\mu}})-[\mathbf{1}'(\mathbf{x}_i-\hat{\boldsymbol{\mu}})]^2}{\sqrt{Q_i(\hat{\delta})}}=0$$

for $\hat{\delta}$, and

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

for $\hat{\sigma}$.

Here $Q_i(\hat{\delta})=[1+(p-1)\hat{\delta}](\mathbf{x}_i-\hat{\boldsymbol{\mu}})'(\mathbf{x}_i-\hat{\boldsymbol{\mu}})-\hat{\delta}[(\mathbf{x}_i-\hat{\boldsymbol{\mu}})'\mathbf{1}]^2$.

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, σ and ρ and the structure is given by

$$\begin{aligned}\boldsymbol{\Sigma} &= \frac{\sigma^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{p-1} \\ \rho & 1 & \dots & \rho^{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \dots & 1 \end{pmatrix} \\ &= \frac{\sigma^2}{1-\rho^2}V(\rho),\end{aligned}$$

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

The determinant and inverse respectively are

$$|\boldsymbol{\Sigma}| = \frac{(\sigma^2)^p}{1-\rho^2} \quad \text{and}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\,[\mathbf{I}_p+\rho^2\mathbf{C}_1-\rho\,\mathbf{C}_2],$$

where

$$\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}, \tag{2.5}$$

$$\mathbf{C}_2 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}. \tag{2.6}$$

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:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_1 & 1 \end{pmatrix} \text{ and } \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_2 & \rho_1 & 1 \end{pmatrix},$$

where $\sigma > 0$ and $-1 < \rho_i < 1, i = 1,\ 2.$

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 $p = 5$:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{pmatrix} \text{ and } \Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 & \rho^2 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho^2 & \rho & 1 \end{pmatrix}.$$

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, $\mathbf{\Sigma} = \mathbf{R}$, where

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & . & . & 1 \end{pmatrix}$$

and $-1 < \rho_{ij} < 1, \ i, j = 1, ..., p$.

2.5 Example

To illustrate the estimation of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ 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 $\mathbf{x}_1$ involves sending a shock wave down the board, the second measurement $\mathbf{x}_2$ is determined while vibrating the board and the last two measurements $\mathbf{x}_3$ and $\mathbf{x}_4$ 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, $\bar{\mathbf{x}}$, covariance matrix, $\mathbf{S}$, and correlation matrix, $\mathbf{R}$, and then provide ML estimates of the parameters involved.

Table 2.1: Board data: measurements on stiffness.

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

The sample mean, covariance matrix, and correlation matrix, respectively are

$$\bar{\mathbf{x}} = \Big(1906.1,\ 1749.5333,\ 1509.1333,\ 1724.9667\Big)',$$

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

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

In the following we provide ML estimates of $\boldsymbol{\mu}$ and parameters of $\boldsymbol{\Sigma}$ under various structures for $\boldsymbol{\Sigma}$ for the example.

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

$$\hat{\boldsymbol{\mu}} = \Big(1850.4159,\ 1691.1202,\ 1481.2977,\ 1685.424\Big)',$$

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 91485.308 & 81017.783 & 77644.475 & 86380.137 \\ 81017.783 & 75470.307 & 69361.802 & 77060.054 \\ 77644.475 & 69361.802 & 70111.928 & 75515.426 \\ 86380.137 & 77060.054 & 75515.426 & 85505.102 \end{pmatrix}.$$

Under Diagonal Covariance Structure for $\boldsymbol{\Sigma}$, we have $2p$ parameters to estimate. The estimation results are

$$\hat{\boldsymbol{\mu}} = \Big(1852.7454,\ 1685.772,\ 1474.3039,\ 1681.4894\Big)',$$

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 14678.514 & 0 & 0 & 0 \\ 0 & 13766.555 & 0 & 0 \\ 0 & 0 & 14022.531 & 0 \\ 0 & 0 & 0 & 16301.33 \end{pmatrix}.$$

Under Equicorrelation Structure for Σ, we have $p+2$ parameters to estimate. The estimation results are

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \left(1859.9366,\ 1697.3872,\ 1484.3861,\ 1691.7187\right)', \\ \hat{\sigma} &= 139.29078, \\ \hat{\delta} &= 0.8898684, \end{aligned}$$

$$\text{so } \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 19401.922 & 17265.158 & 17265.158 & 17265.158 \\ 17265.158 & 19401.922 & 17265.158 & 17265.158 \\ 17265.158 & 17265.158 & 19401.922 & 17265.158 \\ 17265.158 & 17265.158 & 17265.158 & 19401.922 \end{pmatrix}.$$

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

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \left(1859.3134,\ 1694.2681,\ 1485.1081,\ 1688.5484\right)', \\ \hat{\sigma} &= 62.804267, \\ \hat{\rho} &= 0.8946798, \end{aligned}$$

$$\text{so } \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 19766.542 & 17684.726 & 15822.166 & 14155.772 \\ 17684.726 & 19766.542 & 17684.726 & 15822.166 \\ 15822.166 & 17684.726 & 19766.542 & 17684.726 \\ 14155.772 & 15822.166 & 17684.726 & 19766.542 \end{pmatrix}.$$

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

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \left(1858.0636,\ 1695.4335,\ 1482.6518,\ 1690.9761\right)', \\ \hat{\sigma} &= 112.65213, \\ \hat{\rho} &= \left(0.7999,\ 0.7693483\right)', \end{aligned}$$

$$\text{so } \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 12690.502 & 10151.132 & 9763.4154 & 10151.132 \\ 10151.132 & 12690.502 & 10151.132 & 9763.4154 \\ 9763.4154 & 10151.132 & 12690.502 & 10151.132 \\ 10151.132 & 9763.4154 & 10151.132 & 12690.502 \end{pmatrix}.$$

Under the Circular Covariance (2) Structure for $\boldsymbol{\Sigma}$, we have $p + 2$ parameters to estimate. The estimation results are

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \left(1850.1956,\ 1692.4255,\ 1480.7527,\ 1689.693\right)', \\ \hat{\sigma} &= 115.34328, \\ \hat{\rho} &= 0.7292725, \end{aligned}$$

$$\text{so } \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 13304.072 & 9702.2945 & 7075.6168 & 9702.2945 \\ 9702.2945 & 13304.072 & 9702.2945 & 7075.6168 \\ 7075.6168 & 9702.2945 & 13304.072 & 9702.2945 \\ 9702.2945 & 7075.6168 & 9702.2945 & 13304.072 \end{pmatrix}.$$

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

$$\hat{\boldsymbol{\mu}} = \left(1856.2708,\ 1691.7989,\ 1481.0653,\ 1689.4673\right)',$$

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

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 $\boldsymbol{\mu}$ and the general covariance $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We shall focus on the comparisons of ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the general covariance structure are the GSM $\hat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ for Kotz type population and they are $\bar{\mathbf{x}}$ and $\mathbf{S}_n$ for normal population. When the simulation of data is from Kotz type population, we compute $\hat{\boldsymbol{\mu}}_j = (\hat{\mu}_{1j}, ..., \hat{\mu}_{pj})'$ and $\widehat{\boldsymbol{\Sigma}}_j$ as Kotz type ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and $\bar{\mathbf{x}}_j = (\bar{x}_{1j}, ..., \bar{x}_{pj})'$ and $\frac{1}{p+1}\mathbf{S}_{nj}$ as normal ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for the j^{th} simulation, $j = 1, ..., 1000$. However, when the simulation of data is from normal population, we compute $\bar{\mathbf{x}}_j$, $\mathbf{S}_{nj}$ as normal ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\mu}}_j$, $(p+1)\widehat{\boldsymbol{\Sigma}}_j$ as Kotz type ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for the j^{th} simulation, $j = 1, ..., 1000$. A reason for doing these adjustments is that $Var(\mathbf{x}) = \boldsymbol{\Sigma}$ if $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $Var(\mathbf{x}) = (p+1)\boldsymbol{\Sigma}$ if $\mathbf{x}$ is distributed as Kotz type.

Also for convenience of presentation, suppose all the $\nu = \frac{p(p+1)}{2}$ unknown parameters in $\boldsymbol{\Sigma}$ 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 σ_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 $\boldsymbol{\mu}$:

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

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

The estimated bias for $\boldsymbol{\Sigma}$:

For Kotz type distribution,

$$bias(\widehat{\boldsymbol{\Sigma}}) = \sqrt{\sum_{i=1}^{\nu} (\bar{\hat{\sigma}}_i - \sigma_i)^2},$$

$$bias(\frac{1}{p+1}\mathbf{S}_n) = \sqrt{\sum_{i=1}^{\nu} (\frac{1}{p+1}\bar{s}_i - \sigma_i)^2}.$$

For Normal distribution,

$$bias\Big((p+1)\widehat{\boldsymbol{\Sigma}}\Big) = \sqrt{\sum_{i=1}^{\nu} \Big((p+1)\bar{\hat{\sigma}}_i - \sigma_i\Big)^2},$$

$$bias(\mathbf{S}_n) = \sqrt{\sum_{i=1}^{\nu} (\bar{s}_i - \sigma_i)^2},$$

where for any estimate, $\bar{\theta} = \frac{1}{1000}\sum_{j=1}^{1000}\hat{\theta}_j$, 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 $\boldsymbol{\mu}$:

$$root\ MSE(\hat{\boldsymbol{\mu}}) = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{1000} (\hat{\mu}_{ij} - \mu_i)^2}{1000}},$$

$$root\ MSE(\bar{\mathbf{x}}) = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{1000}(\bar{x}_{ij}-\mu_i)^2}{1000}}.$$

The estimated root MSE for $\boldsymbol{\Sigma}$:

For Kotz type distribution,

$$root\ MSE(\widehat{\boldsymbol{\Sigma}}) = \sqrt{\frac{\sum_{i=1}^{\nu}\sum_{j=1}^{1000}(\hat{\sigma}_{ij}-\sigma_i)^2}{1000}},$$

$$root\ MSE(\frac{1}{p+1}\mathbf{S}_n) = \sqrt{\frac{\sum_{i=1}^{\nu}\sum_{j=1}^{1000}(\frac{1}{p+1}s_{ij}-\sigma_i)^2}{1000}}.$$

For Normal distribution,

$$root\ MSE\Big((p+1)\widehat{\boldsymbol{\Sigma}}\Big) = \sqrt{\frac{\sum_{i=1}^{\nu}\sum_{j=1}^{1000}\Big((p+1)\hat{\sigma}_{ij}-\sigma_i\Big)^2}{1000}},$$

$$root\ MSE(\mathbf{S}_n) = \sqrt{\frac{\sum_{i=1}^{\nu}\sum_{j=1}^{1000}(s_{ij}-\sigma_i)^2}{1000}}.$$

(iii) *Pitman Nearness (PN) Probability :*

For Kotz type distribution, the estimate of Pitman Nearness probability of $\hat{\boldsymbol{\mu}}$ relative to $\bar{\mathbf{x}}$ is computed as

$$PN(\hat{\boldsymbol{\mu}},\bar{\mathbf{x}}) = \frac{1}{1000}\Big(\#\text{of times }\Big[\|\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}\| < \|\bar{\mathbf{x}}_j-\boldsymbol{\mu}\|\Big]\text{ in 1000 samples}\Big),$$

where

$$\|\hat{\boldsymbol{\mu}}-\boldsymbol{\mu}\| = \sqrt{\sum_{i=1}^{p}(\hat{\mu}_i-\mu_i)^2} \quad \text{and}$$

$$\|\bar{\mathbf{x}}-\boldsymbol{\mu}\| = \sqrt{\sum_{i=1}^{p}(\bar{x}_i-\mu_i)^2}$$

and the estimate of Pitman Nearness probability of $\widehat{\boldsymbol{\Sigma}}$ relative to $\frac{1}{p+1}\mathbf{S}_n$ is computed as

$$PN(\widehat{\boldsymbol{\Sigma}},\frac{1}{p+1}\mathbf{S}_n) = \frac{1}{1000}\Big(\#\text{of times }\Big[\|\widehat{\boldsymbol{\Sigma}}_j-\boldsymbol{\Sigma}\| < \|\frac{1}{p+1}\mathbf{S}_{nj}-\boldsymbol{\Sigma}\|\Big]\text{ in 1000 samples}\Big),$$

where

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| = \sqrt{\sum_{i=1}^{\nu} (\hat{\sigma}_i - \sigma_i)^2} \quad \text{and}$$

$$\|\frac{1}{p+1}\mathbf{S}_n - \boldsymbol{\Sigma}\| = \sqrt{\sum_{i=1}^{\nu} (\frac{1}{p+1}s_i - \sigma_i)^2}.$$

For normal distribution, the estimate of Pitman Nearness probability of $\bar{\mathbf{x}}$ relative to $\hat{\boldsymbol{\mu}}$ is computed as

$$PN(\bar{\mathbf{x}}, \hat{\boldsymbol{\mu}}) = \frac{1}{1000}\Big(\#\text{of times}\Big[\|\bar{\mathbf{x}}_j - \boldsymbol{\mu}\| < \|\hat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}\|\Big] \text{ in 1000 samples}\Big),$$

where

$$\|\bar{\mathbf{x}} - \boldsymbol{\mu}\| = \sqrt{\sum_{i=1}^{p} (\bar{x}_i - \mu_i)^2} \quad \text{and}$$

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

and the estimate of Pitman Nearness probability of $\mathbf{S}_n$ relative to $(p+1)\widehat{\boldsymbol{\Sigma}}$ is computed as

$$PN\Big(\mathbf{S}_n, (p+1)\widehat{\boldsymbol{\Sigma}}\Big) = \frac{1}{1000}\Big(\#\text{of times}\Big[\|\mathbf{S}_{nj} - \boldsymbol{\Sigma}\| < \|(p+1)\widehat{\boldsymbol{\Sigma}}_j - \boldsymbol{\Sigma}\|\Big] \text{ in 1000 samples}\Big),$$

where

$$\|\mathbf{S}_n - \boldsymbol{\Sigma}\| = \sqrt{\sum_{i=1}^{\nu} (s_i - \sigma_i)^2} \quad \text{and}$$

$$\|(p+1)\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| = \sqrt{\sum_{i=1}^{\nu} \Big((p+1)\hat{\sigma}_i - \sigma_i\Big)^2}.$$

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):

$$bias(\hat{\boldsymbol{\mu}}) < bias(\bar{\mathbf{x}}),$$

$$bias(\widehat{\boldsymbol{\Sigma}}) < bias(\frac{1}{p+1}\mathbf{S}_n).$$

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:

$$\begin{aligned} root\ MSE(\hat{\boldsymbol{\mu}}) &< root\ MSE(\bar{\mathbf{x}}), \\ root\ MSE(\widehat{\boldsymbol{\Sigma}}) &< root\ MSE(\frac{1}{p+1}\mathbf{S}_n). \end{aligned}$$

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 $\hat{\boldsymbol{\mu}}$ and $\bar{\mathbf{x}}$ and Pitman nearness probability of $\hat{\boldsymbol{\mu}}$ relative to $\bar{\mathbf{x}}$ for Kotz type samples.

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

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):

$$\begin{aligned} bias(\bar{\mathbf{x}}) &< bias(\hat{\boldsymbol{\mu}}), \\ bias(\mathbf{S}_n) &< bias\Big((p+1)\widehat{\boldsymbol{\Sigma}}\Big). \end{aligned}$$

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:

$$root\ MSE(\bar{\mathbf{x}}) < root\ MSE(\hat{\boldsymbol{\mu}}),$$

Table 2.3: The estimates of bias and root MSE for $\widehat{\mathbf{\Sigma}}$ and $\frac{1}{p+1}\mathbf{S}_n$ and Pitman nearness probability of $\widehat{\mathbf{\Sigma}}$ relative to $\frac{1}{p+1}\mathbf{S}_n$ for Kotz type samples.

p	n	$Bias(\widehat{\mathbf{\Sigma}})$	$Bias(\frac{1}{p+1}\mathbf{S}_n)$	$root\ MSE(\widehat{\mathbf{\Sigma}})$	$root\ MSE(\frac{1}{p+1}\mathbf{S}_n)$	$PN(\widehat{\mathbf{\Sigma}}, \frac{1}{p+1}\mathbf{S}_n)$
2	20	0.2173286	0.6136360	4.5729571	4.6180806	0.580
	30	0.0594558	0.2972934	3.6564364	3.8758817	0.603
	50	0.0397885	0.2024529	2.8092322	3.0098940	0.595
	100	0.0423176	0.0496837	1.9725625	2.1130978	0.571
5	20	2.5227599	7.5841423	59.32383	59.885291	0.558
	30	1.3631267	4.8771157	47.940337	48.826692	0.578
	50	0.8710543	1.4684184	37.831514	39.230765	0.603
	100	0.6455298	1.4271476	25.851964	27.179129	0.616

$$root\ MSE(\mathbf{S}_n) \quad < \quad root\ MSE\Big((p+1)\widehat{\mathbf{\Sigma}}\Big).$$

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 $\hat{\boldsymbol{\mu}}$ and $\bar{\mathbf{x}}$ and Pitman nearness probability of $\bar{\mathbf{x}}$ relative to $\hat{\boldsymbol{\mu}}$ for normal samples.

p	n	$Bias(\hat{\boldsymbol{\mu}})$	$Bias(\bar{\mathbf{x}})$	$root\ MSE(\hat{\boldsymbol{\mu}})$	$root\ MSE(\bar{\mathbf{x}})$	$PN(\bar{\mathbf{x}}, \hat{\boldsymbol{\mu}})$
2	20	0.0419632	0.0250159	0.9201970	0.8175815	0.598
	30	0.0235707	0.0146184	0.7436633	0.6521578	0.613
	50	0.0103899	0.0067315	0.5750868	0.5067936	0.620
	100	0.0036084	0.0033503	0.4029865	0.3635994	0.600
5	20	0.0685258	0.0337121	3.5255275	3.3563840	0.589
	30	0.0653346	0.0331537	2.8187549	2.6684190	0.614
	50	0.0612400	0.0431314	2.1983321	2.0978606	0.580
	100	0.0313821	0.0227773	1.5524146	1.4771821	0.600

Table 2.5: The estimates of bias and root MSE for $(p+1)\widehat{\Sigma}$ and $\mathbf{S}_n$ and Pitman nearness probability of $\mathbf{S}_n$ relative to $(p+1)\widehat{\Sigma}$ for normal samples.

p	n	$Bias\left((p+1)\widehat{\Sigma}\right)$	$Bias(\mathbf{S}_n)$	$root\ MSE\left((p+1)\widehat{\Sigma}\right)$	$root\ MSE(\mathbf{S}_n)$	$PN\left(\mathbf{S}_n,(p+1)\widehat{\Sigma}\right)$
2	20	1.5617102	0.4167722	4.3717427	3.2530058	0.599
	30	1.4783830	0.3349251	3.7999676	2.8374782	0.602
	50	1.7676229	0.1653396	3.2272657	2.1500709	0.654
	100	1.7092323	0.1373975	2.5760868	1.5487331	0.711
5	20	7.6239726	6.9183541	56.9054100	49.532877	0.594
	30	9.5955013	4.6867674	48.4888900	41.584469	0.584
	50	8.5435641	4.2156948	35.7407750	31.135813	0.589
	100	10.706864	1.7581084	28.4005960	23.347784	0.632

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 $\boldsymbol{\mu}$ obtained by maximizing the log-likelihood function of Kotz type distribution is same as the generalized spatial median of $\boldsymbol{\mu}$ introduced by Rao (1988). The estimators of $\boldsymbol{\mu}$ and Σ 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 $\boldsymbol{\mu}$ and Σ 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 $\mathbf{x}_1, ..., \mathbf{x}_n$ be a random sample from Kotz type distribution with the parameters $\boldsymbol{\mu}$ and Σ. Let $\boldsymbol{\theta} = (\mu_1, ..., \mu_p\ , \sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1,p}, \sigma_{pp})'$ be the vector of all unknown parameters and let $L(\boldsymbol{\theta})$ be the likelihood function given the samples $\mathbf{x}_1, ..., \mathbf{x}_n$. Suppose under the null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_0$, where $\Theta_0 \subset \Theta$ and Θ is the parameter space.

Under H_0, let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\mu}_1, ..., \widetilde{\mu}_p\ , \widetilde{\sigma}_{11}, ..., \widetilde{\sigma}_{1p}, ..., \widetilde{\sigma}_{p-1,p}, \widetilde{\sigma}_{pp})'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum likelihood function under H_0 is given by

$$L(\widetilde{\boldsymbol{\theta}}) \;=\; c^n|\widetilde{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^{n}\sqrt{(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})}\right\}.$$

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

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

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

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

that is, if

$$\Lambda = \left[\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widetilde{\boldsymbol{\Sigma}}|}\right]^{\frac{n}{2}} \exp\left\{-\textstyle\sum_{i=1}^{n}\left[\sqrt{(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})} - \sqrt{(\mathbf{x}_i - \hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})}\,\right]\right\} < c,$$

where c is a suitably chosen constant.

When the sample size n is large,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\right) \text{ is approximately distributed as } \chi^2_r$$

random variable, where the degrees of freedom, r = (dimension of Θ) - (dimension of Θ_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 $H_0 : \boldsymbol{\mu} = \mathbf{0}$ *vs.* $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$.

Under H_0, the maximum likelihood estimate of $\boldsymbol{\mu}$ is $\widetilde{\boldsymbol{\mu}} = \mathbf{0}$. 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 $\boldsymbol{\mu} \neq \mathbf{0}$.

(ii) Suppose we want to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_4 = \gamma$ (say) vs. $H_1 : \boldsymbol{\mu}$ is arbitrary. That is, test $H_0 : \boldsymbol{\mu} = \gamma\mathbf{1}$ vs. $H_1 : \boldsymbol{\mu} \neq \gamma\mathbf{1}$, where $\mathbf{1} = (1, ..., 1)'$. Under H_0, $\widetilde{\boldsymbol{\mu}} = \widetilde{\gamma}\mathbf{1}$, where $\widetilde{\gamma}$ is the MLE of γ.

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 $\boldsymbol{\mu}$ are zero. That is, $H_0 : \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix}$ vs. $H_1 : \boldsymbol{\mu}$ is arbitrary.

Under H_0, $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \widetilde{\mu}_2, 0, 0)'$ 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 $\mathbf{x}_1, ..., \mathbf{x}_n$ be a random sample from Kotz type distribution with the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta} = (\mu_1, ..., \mu_p, \sigma, \rho)'$ be the vector of all unknown parameters and let $L(\boldsymbol{\theta})$ be the likelihood function. Suppose under the null hypothesis $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, where $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$.

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

$$
\begin{aligned}
L(\widetilde{\boldsymbol{\theta}}) &= c^n \left[\frac{(\widetilde{\sigma}^2)^p}{1-\widetilde{\rho}^2}\right]^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^{n} \sqrt{(\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})' \frac{1}{\widetilde{\sigma}^2}[\mathbf{I}_p + \widetilde{\rho}^2\mathbf{C}_1 - \widetilde{\rho}\mathbf{C}_2](\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})}\right\} \\
&= c^n \left[\frac{(\widetilde{\sigma}^2)^p}{1-\widetilde{\rho}^2}\right]^{-\frac{n}{2}} \exp\left\{-\frac{1}{\widetilde{\sigma}} \sum_{i=1}^{n} \sqrt{A_{1i} + \widetilde{\rho}^2 A_{2i} - \widetilde{\rho} A_{3i}}\right\},
\end{aligned}
$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are as defined in (2.5), (2.6) and

$$\begin{aligned} A_{1i} &= (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})' \, (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}), \\ A_{2i} &= (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})' \, \mathbf{C}_1 \, (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}), \\ A_{3i} &= (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}})' \, \mathbf{C}_2 \, (\mathbf{x}_i - \widetilde{\boldsymbol{\mu}}). \end{aligned}$$

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

$$\begin{aligned} L(\widehat{\boldsymbol{\theta}}) &= c^n \left[\frac{(\hat{\sigma}^2)^p}{1-\hat{\rho}^2}\right]^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^n \sqrt{(\mathbf{x}_i - \hat{\boldsymbol{\mu}})' \frac{1}{\hat{\sigma}^2}[\mathbf{I}_p + \hat{\rho}^2 \mathbf{C}_1 - \hat{\rho}\mathbf{C}_2](\mathbf{x}_i - \hat{\boldsymbol{\mu}})}\right\} \\ &= c^n \left[\frac{(\hat{\sigma}^2)^p}{1-\hat{\rho}^2}\right]^{-\frac{n}{2}} \exp\left\{-\frac{1}{\hat{\sigma}} \sum_{i=1}^n \sqrt{B_{1i} + \hat{\rho}^2 B_{2i} - \hat{\rho} B_{3i}}\right\}, \end{aligned}$$

where

$$\begin{aligned} B_{1i} &= (\mathbf{x}_i - \hat{\boldsymbol{\mu}})' \, (\mathbf{x}_i - \hat{\boldsymbol{\mu}}), \\ B_{2i} &= (\mathbf{x}_i - \hat{\boldsymbol{\mu}})' \, \mathbf{C}_1 \, (\mathbf{x}_i - \hat{\boldsymbol{\mu}}), \\ B_{3i} &= (\mathbf{x}_i - \hat{\boldsymbol{\mu}})' \, \mathbf{C}_2 \, (\mathbf{x}_i - \hat{\boldsymbol{\mu}}). \end{aligned}$$

Then the likelihood ratio test of $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ rejects H_0 in favor of $H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$ if

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

that is, if

$$\Lambda = \left(\frac{\hat{\sigma}^2}{\widetilde{\sigma}^2}\right)^{\frac{np}{2}} \left(\frac{1-\widetilde{\rho}^2}{1-\hat{\rho}^2}\right)^{\frac{n}{2}}$$

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

where c is a suitable constant.

When the sample size n is large,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\right) \text{ is approximately distributed as } \chi_r^2$$

random variable, where the degrees of freedom, r = (dimension of Θ) - (dimension of Θ_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 $H_0 : \boldsymbol{\mu} = \mathbf{0}$ *vs.* $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$.

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

The test statistic $= -2\ ln\Lambda = 228.5621$. The P-value $= P[\chi^2_4 > 228.5621] < 0.0001$. Hence we reject H_0 and conclude that $\boldsymbol{\mu} \neq \mathbf{0}$.

(ii) Next, we want to test $H_0 : \mu_1 = \mu_2 = ... = \mu_4 = \gamma$ (say) vs. $H_1 : \boldsymbol{\mu}$ is arbitrary.

That is, $H_0 : \boldsymbol{\mu} = \gamma\mathbf{1}$ vs. $H_1 : \boldsymbol{\mu} \neq \gamma\mathbf{1}$, where $\mathbf{1} = (1, ..., 1)'$.

Under H_0, $\widetilde{\boldsymbol{\mu}} = \widetilde{\gamma}\mathbf{1}$, where $\widetilde{\gamma}$ is the MLE of γ.

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 $\boldsymbol{\mu}$ are zero. That is, $H_0 : \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix}$ vs. $H_1 : \boldsymbol{\mu}$ is arbitrary.

Under H_0, $\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1, \widetilde{\mu}_2, 0, 0)'$ 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 $\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in_i}$ is a random sample of size n_i from Kotz type population with the parameters $\boldsymbol{\mu}_i$ and Σ_i, $i = 1, ..., g$. The random samples from different g populations are assumed to be independent.

Let $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{ip})'$ and let $\boldsymbol{\sigma}_i = (\sigma_{i,11}, ..., \sigma_{i,1p}, ..., \sigma_{i,p-1,p}, \sigma_{i,pp})'$, $i = 1, ..., g$.

Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_g', \boldsymbol{\sigma}_1', ..., \boldsymbol{\sigma}_g')'$ be the vector of all unknown parameters.

(i) Test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$ under the assumption $\Sigma_1 = \Sigma_2 =$ $\cdots = \Sigma_g = \Sigma$, that is, $\boldsymbol{\sigma}_1 = \cdots = \boldsymbol{\sigma}_g = \boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = (\sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1,p}, \sigma_{pp})'$. Under H_0, let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\mu}}', \widetilde{\boldsymbol{\sigma}}')'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum likelihood function is given by

$$L(\widetilde{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} |\widetilde{\Sigma}|^{\frac{-\sum_1^g n_i}{2}} \exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\Sigma}^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})}\right\}.$$

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

$$L(\widehat{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} |\widehat{\Sigma}|^{\frac{-\sum_1^g n_i}{2}} \exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\Sigma}^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}\right\}.$$

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

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

that is, if

$$\Lambda = \left[\frac{|\widehat{\Sigma}|}{|\widetilde{\Sigma}|}\right]^{\frac{(\sum_1^g n_i)}{2}}$$
$$\exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\Sigma}^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})} - \sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\Sigma}^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}\right]\right\} < c.$$

(ii) Test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$ when Σ_i's are different, that is, when $\boldsymbol{\sigma}_i$'s are different.

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

$$L(\widetilde{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \prod_{i=1}^{g} |\widetilde{\Sigma}_i|^{-\frac{n_i}{2}} \exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\Sigma}_i^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})}\right\}.$$

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

$$L(\widehat{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \prod_{i=1}^{g} |\widehat{\Sigma}_i|^{-\frac{n_i}{2}} \exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\Sigma}_i^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}\right\}.$$

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

$$\Lambda = \frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})} < c,$$

that is, if

$$\Lambda = \prod_{i=1}^{g} \left[\frac{|\widehat{\boldsymbol{\Sigma}}_i|}{|\widetilde{\boldsymbol{\Sigma}}_i|}\right]^{\frac{n_i}{2}}$$
$$\exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[\sqrt{(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{x}_{ij}-\widetilde{\boldsymbol{\mu}})} - \sqrt{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)}\right]\right\} < c,$$

where c is a suitably chosen constant.

When the sample size n is large,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\right) \text{ is approximately distributed as } \chi_r^2$$

random variable, where the degrees of freedom, r = (dimension of Θ) - (dimension of Θ_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, $\mathbf{x}_1$ = eye-to-top-of-head measurement, $\mathbf{x}_2$ = ear-to-top-of-head measurement, and $\mathbf{x}_3$ = 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 $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}$.

The test statistic = $-2\ ln\Lambda$ = 13.054878. Using $-2\ ln\Lambda \sim \chi^2_{12}$, 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).

HS			C			NF		
x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
12.5	14	11	10.3	13.4	12.4	7.4	13	12
12	16	12	12.8	14.5	11.3	10.5	13.8	11.5
10	13	12	11.4	14.1	12.1	9.7	13.3	11.5
13.5	15	12	11	13.4	11	8.5	12	11.5
13	15.5	12	9.6	11.1	11.7	11.5	14.5	11.8
12	14	13	9.9	12.8	11.4	13	13.4	11.5
13.5	15.5	13	10.2	12.8	11.9	10.8	12.8	12.6
13	14	13	8.8	13	12.9	11.1	13.9	11.2
13.5	14.5	12.5	10.5	13.9	11.8	11.5	13.4	11.9
13	15	13	10.4	14.5	12	10.6	13.7	12.2
14	14.5	11.5	11.2	13.4	12.4	10.4	13.5	11.4
13	16	12.5	9.2	12.8	12.2	10	13.1	10.9
14	14.5	12	11.8	12.6	12.5	12	13.6	11.5
14	16	12	10.2	12.7	12.3	10.2	13.6	11.5
13.5	15	12	11.2	13.8	11.3	11.3	13.6	11.3
15	15	12	9.4	14.3	12.2	10.5	13.5	12.1
12	14.5	12	9.8	13.8	12.6	9.9	14	12.1
13	14	12.5	10.1	14.2	11.6	11	15.1	11.7
12.5	14	12.5	12	12.6	11.6	12.1	14.6	12.1
12	14	11	9.9	13.4	11.5	11.7	13.8	12.1
12	13	12	9.9	14.4	11.9	11.8	14.7	11.8
14.5	14.5	13	9.1	12.8	11.7	10.8	13.9	12
14	15.5	13.5	8.6	14.2	11.5	11.3	14	11.4
13	15.5	12.5	8.2	13	12.6	10.4	13.8	12.2
12	13	12.5	9.8	13.8	10.5	10.2	13.9	11.7
14.5	15.5	12.5	9.6	13	11.2	12.4	13.4	12.1
14.5	16.5	13.5	8.6	13.5	11.8	10.7	14.2	10.8
13	16	10.5	9.6	14.1	12.3	13.1	14.5	11.7
13.5	14	12	9	13.9	13.3	12.1	13	12.7
12.5	14.5	12.5	10.3	13.8	12.8	11.9	13.3	13.3

(ii) Next, we want to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}$, given $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}$, that is, test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}$, given $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2 = \boldsymbol{\sigma}_3 = \boldsymbol{\sigma}$,
where $\boldsymbol{\sigma} = (\sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1,p}, \sigma_{pp})'$.
Under H_0, let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\mu}}', \widetilde{\boldsymbol{\sigma}}')'$ be the MLE of $\boldsymbol{\theta}$. Then the maximum likelihood function is given by

$$L(\widetilde{\boldsymbol{\theta}}) = c^{(\Sigma_1^3 n_i)} |\widetilde{\boldsymbol{\Sigma}}|^{-\frac{\Sigma_1^3 n_i}{2}} \exp\left\{ -\sum_{i=1}^{3}\sum_{j=1}^{n_i} \sqrt{(\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})} \right\}.$$

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

$$L(\widehat{\boldsymbol{\theta}}) = c^{(\Sigma_1^3 n_i)} |\widehat{\boldsymbol{\Sigma}}|^{-\frac{\Sigma_1^3 n_i}{2}} \exp\left\{ -\sum_{i=1}^{3}\sum_{j=1}^{n_i} \sqrt{(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)} \right\}.$$

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

$$\Lambda = \left[\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widetilde{\boldsymbol{\Sigma}}|}\right]^{\frac{\Sigma_1^3 n_i}{2}}$$
$$\exp\left\{ -\sum_{i=1}^{3}\sum_{j=1}^{n_i}\left[\sqrt{(\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})'\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})} - \sqrt{(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)} \right] \right\} < c,$$

where c is a suitable constant.

When the sample size n is large,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\right) \text{ is approximately distributed as } \chi_6^2.$$

The test statistic $= -2\, ln\Lambda = 91.70311$. The P-value $= P[\chi_6^2 > 91.70311] < 0.0001$. Hence we reject H_0 and conclude that at least some $\boldsymbol{\mu}_i's$ are different.

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

Suppose $\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in_i}$, $i = 1..., g$ are random samples of size n_i from g independent Kotz type distributions with the parameters $\boldsymbol{\mu}_i$ and Σ_i, $i = 1, ..., g$. We assume that $\Sigma_i's$ have AR(1) structures with different variance and correlation parameters.

$\boldsymbol{\mu}_i = (\mu_{i1},\ \mu_{i2}, ...,\ \mu_{ip})'$, $i = 1, ..., g$. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_g', \sigma_1, ..., \sigma_g, \rho_1, ..., \rho_g)'$ be the vector of all unknown parameters

(i) Test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$ when $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$, that is $\sigma_1^2 = \cdots = \sigma_g^2 = \sigma^2$ and $\rho_1 = \cdots = \rho_g = \rho$.

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

$$L(\widehat{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \left[\frac{(\hat{\sigma}^2)^p}{1-\hat{\rho}^2}\right]^{-\frac{\sum_1^g n_i}{2}}$$

$$\exp\left\{-\frac{1}{\hat{\sigma}} \sum_{i=1}^g \sum_{j=1}^{n_i} \sqrt{D_{1,ij} + \hat{\rho}^2 D_{2,ij} - \hat{\rho} D_{3,ij}}\right\},$$

where

$$\begin{aligned} D_{1,ij} &= (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)'\ (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i), \\ D_{2,ij} &= (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)'\ \mathbf{C}_1\ (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i), \\ D_{3,ij} &= (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i)'\ \mathbf{C}_2\ (\mathbf{x}_{ij} - \hat{\boldsymbol{\mu}}_i), \end{aligned}$$

and $\mathbf{C}_1$ and $\mathbf{C}_2$ are defined in (2.5), (2.6).

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

$$L(\widetilde{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \left[\frac{(\widetilde{\sigma}^2)^p}{1-\widetilde{\rho}^2}\right]^{-\frac{\sum_1^g n_i}{2}}$$

$$\exp\left\{-\frac{1}{\widetilde{\sigma}} \sum_{i=1}^g \sum_{j=1}^{n_i} \sqrt{E_{1,ij} + \widetilde{\rho}^2 E_{2,ij} - \widetilde{\rho} E_{3,ij}}\right\},$$

where

$$\begin{aligned} E_{1,ij} &= (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})'\ (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}}), \\ E_{2,ij} &= (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})'\ \mathbf{C}_1\ (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}}), \\ E_{3,ij} &= (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}})'\ \mathbf{C}_2\ (\mathbf{x}_{ij} - \widetilde{\boldsymbol{\mu}}). \end{aligned}$$

The resulting likelihood ratio test rejects H_0 if

$$\Lambda = \left[\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right]^{\frac{(\sum n_i)p}{2}} \left[\frac{1-\tilde{\rho}^2}{1-\hat{\rho}^2}\right]^{\frac{\sum n_i}{2}}$$

$$\exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[\frac{1}{\tilde{\sigma}}\sqrt{E_{1,ij}+\tilde{\rho}^2E_{2,ij}-\tilde{\rho}E_{3,ij}}-\frac{1}{\hat{\sigma}}\sqrt{D_{1,ij}+\hat{\rho}^2D_{2,ij}-\hat{\rho}D_{3,ij}}\right]\right\} < c,$$

where c is a suitable constant.

(ii) Test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g = \boldsymbol{\mu}$ when σ_i^2's and ρ_i's are different.

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

$$L(\widehat{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \prod_{i=1}^{g}\left[\frac{(\hat{\sigma}_i^2)^p}{1-\hat{\rho}_i^2}\right]^{-\frac{n_i}{2}}$$

$$\exp\left\{-\sum_{i=1}^{g}\frac{1}{\hat{\sigma}_i}\sum_{j=1}^{n_i}\sqrt{D_{1,ij}+\hat{\rho}_i^2D_{2,ij}-\hat{\rho}_iD_{3,ij}}\right\}.$$

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

$$L(\widetilde{\boldsymbol{\theta}}) = c^{(\sum_1^g n_i)} \prod_{i=1}^{g}\left[\frac{(\widetilde{\sigma}_i^2)^p}{1-\widetilde{\rho}_i^2}\right]^{-\frac{n_i}{2}}$$

$$\exp\left\{-\sum_{i=1}^{g}\frac{1}{\widetilde{\sigma}_i}\sum_{j=1}^{n_i}\sqrt{E_{1,ij}+\widetilde{\rho}_i^{\,2}E_{2,ij}-\widetilde{\rho}_iE_{3,ij}}\right\}.$$

Now the resulting likelihood ratio test rejects H_0 if

$$\Lambda = \prod_{i=1}^{g}\left[\left(\frac{\hat{\sigma}_i^2}{\widetilde{\sigma}_i^2}\right)^p\left(\frac{1-\widetilde{\rho}_i^2}{1-\hat{\rho}_i^2}\right)\right]^{\frac{n_i}{2}}$$

$$\exp\left\{-\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left[\frac{1}{\widetilde{\sigma}_i}\sqrt{E_{1,ij}+\widetilde{\rho}_i^2E_{2,ij}-\widetilde{\rho}_iE_{3,ij}}-\frac{1}{\hat{\sigma}_i}\sqrt{D_{1,ij}+\hat{\rho}_i^2D_{2,ij}-\hat{\rho}_iD_{3,ij}}\right]\right\} < c,$$

where c is an appropriate constant.

When the sample size n is large,

$$-2\ln\Lambda = -2\ln\left(\frac{L(\widetilde{\boldsymbol{\theta}})}{L(\widehat{\boldsymbol{\theta}})}\right) \text{ is approximately distributed as } \chi_r^2$$

random variable, where the degrees of freedom, r = (dimension of Θ) - (dimension of Θ_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 $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}$, 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 $\boldsymbol{\mu}_i$ is different from the others.

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

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

Theorem 3.1 *Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be a random sample from Kotz type distribution with parameters $\boldsymbol{\mu}$ and Σ and $\hat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$ be the ML estimates of $\boldsymbol{\mu}$ and Σ. Then*

$$\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\boldsymbol{0},\ \widehat{\Sigma}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\widehat{\Sigma}),$$

where

$$\boldsymbol{B} = E\left[\frac{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'}{(\boldsymbol{x}-\boldsymbol{\mu})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}\right] \quad \textit{and}$$

$$\boldsymbol{A} = E\left[\frac{1}{\sqrt{(\boldsymbol{x}-\boldsymbol{\mu})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}\left(\widehat{\boldsymbol{\Sigma}} - \frac{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'}{(\boldsymbol{x}-\boldsymbol{\mu})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}\right)\right].$$

Further $\boldsymbol{B}$ *and* $\boldsymbol{A}$ *can be estimated by*

$$\widehat{\boldsymbol{B}} = \frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'}{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})},$$

$$\widehat{\boldsymbol{A}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{\sqrt{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})}}\left[\widehat{\boldsymbol{\Sigma}} - \frac{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'}{(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}})}\right]\right\}.$$

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

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

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

where $\hat{\tau}_i =$ *the* i^{th} *diagonal element of* $\hat{\boldsymbol{\tau}} = \widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{\Sigma}}$, *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 $\boldsymbol{\mu}$.

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

$$\left(\boldsymbol{a}_i'\hat{\boldsymbol{\mu}} - z_{\alpha/2m}\sqrt{\frac{\boldsymbol{a}_i'\hat{\boldsymbol{\tau}}\boldsymbol{a}_i}{n}},\ \boldsymbol{a}_i'\hat{\boldsymbol{\mu}} + z_{\alpha/2m}\sqrt{\frac{\boldsymbol{a}_i'\hat{\boldsymbol{\tau}}\boldsymbol{a}_i}{n}}\right),\ i = 1, ..., m \qquad (3.1)$$

where $\boldsymbol{a}_i$*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 $\mathbf{a}$, $\mathbf{a}'\boldsymbol{\mu}$ is a linear function of the $\mu_i's$ and using Theorem 3.1 we have the large sample distributon of $\mathbf{a}'\hat{\boldsymbol{\mu}}$ as:

$$\mathbf{a}'\hat{\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} N(\mathbf{a}'\boldsymbol{\mu},\ \mathbf{a}'\hat{\boldsymbol{\tau}}\mathbf{a}).$$

Then $100(1-\alpha)\%$ Bonferroni simultaneous confidence intervals for the linear functions $\mathbf{a}_1'\boldsymbol{\mu}, ..., \mathbf{a}_m'\boldsymbol{\mu}$ are clearly given by

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

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 $\mu_i, i = 1, ..., 4$ as given by (3.1) are:

$$\left(\hat{\mu}_i - z_{.05/2(4)}\sqrt{\frac{\hat{\tau}_i}{n}},\ \hat{\mu}_i + z_{.05/2(4)}\sqrt{\frac{\hat{\tau}_i}{n}}\right)$$

where $\hat{\tau}_i$ is the i^{th} diagonal element of $\boldsymbol{\tau}$ defined in Proposition 3.1. For our data, the 95% Bonferroni simultaneous confidence intervals for $\mu_i, i = 1, ..., 4$ are:

$$\begin{aligned}
\mu_1 &\in (1715.813,\ 1985.0188),\\
\mu_2 &\in (1565.0283,\ 1817.212),\\
\mu_3 &\in (1361.8197,\ 1600.7757),\\
\mu_4 &\in (1553.4839,\ 1817.3641).
\end{aligned}$$

(ii) Using Proposition 3.2, the 95% Bonferroni simultaneous confidence intervals for $\mu_l - \mu_k,\ l < k = 1, ..., 4$ 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

$$\begin{aligned}
\mu_1 - \mu_2 &\in (100.17677,\ 218.41464),\\
\mu_2 - \mu_3 &\in (148.81812,\ 270.82676),\\
\mu_3 - \mu_4 &\in (-265.3922, -142.8604).
\end{aligned}$$

(iii) The 95% Bonferroni simultaneous confidence intervals for μ_3, μ_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 μ_i's are given by

$$\left(\mathbf{a}_i'\hat{\boldsymbol{\mu}} - z_{\alpha/2m}\sqrt{\frac{\mathbf{a}_i'\hat{\boldsymbol{\tau}}\mathbf{a}_i}{n}},\ \mathbf{a}_i'\hat{\boldsymbol{\mu}} + z_{1-\alpha/2m}\sqrt{\frac{\mathbf{a}_i'\hat{\boldsymbol{\tau}}\mathbf{a}_i}{n}}\right),\ i = 1, ..., m, \tag{3.2}$$

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 $\hat{\boldsymbol{\tau}} = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}\widehat{\boldsymbol{\Sigma}}$ with $\hat{\boldsymbol{\Sigma}} = \frac{\hat{\sigma}^2}{1-\hat{\rho}^2}\mathbf{V}(\hat{\rho})$ and $\mathbf{V}(\hat{\rho}) = \hat{\rho}^{|j-k|},\ j,k = 1, ..., p.$

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

(i) The 95% Bonferroni simultaneous confidence intervals for the individual μ_i, $i = 1, ..., 4$ 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 $\mu_l - \mu_k$, $l < k = 1, ..., 4$ 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 $\mathbf{a}_1' = (0,0,1,0)$, and $\mathbf{a}_2' = (0,0,0,1)$ as

$$(1386.8905,\ 1583.3258),$$
$$(1576.4067,\ 1800.6901).$$

3.8 Simultaneous Confidence Intervals for Several Population Means

Let $\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in_i},\ i = 1, .., g,$ be g independent random samples of size n_i each from Kotz type distributions with parameters $\boldsymbol{\mu}_i$ and $\Sigma_i,\ i = 1, ..., g$. 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 $\mu_{ij}'s$.

(i) Assume that $\Sigma_1 = \cdots = \Sigma_g = \Sigma$, that is, $\boldsymbol{\sigma}_1 = \cdots = \boldsymbol{\sigma}_g = \boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = (\sigma_{11}, ..., \sigma_{1p}, ..., \sigma_{p-1,p}, \sigma_{pp})'$.

Proposition 3.3. *Let $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{in_i}$ be a random sample of size n_i from Kotz type distribution with parameters $\boldsymbol{\mu}_i$ and $\Sigma_i,\ i = 1, ..., g$. 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 $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'},\ l < l' = 1, ..., g$ are given by*

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

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

and

$$\widehat{\boldsymbol{B}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},$$

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

Proof. For a fixed $\mathbf{a}$, $\mathbf{a}'\boldsymbol{\mu}_i$ is a linear function of the $\mu_{ij}'s$ and using Theorem 3.1 and Proposition 3.2 we have

$$\mathbf{a}'(\hat{\boldsymbol{\mu}}_l-\hat{\boldsymbol{\mu}}_{l'}) \xrightarrow{\mathcal{D}} N\left(\mathbf{a}'(\boldsymbol{\mu}_l-\boldsymbol{\mu}_{l'}),\ \mathbf{a}'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l+\frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}\right)$$

Then $100(1-\alpha)\%$ Bonferroni simultaneous confidence intervals for the linear functions $\mathbf{a}_k'(\boldsymbol{\mu}_l-\boldsymbol{\mu}_{l'})$ are clearly given by

$$\mathbf{a}_k'(\hat{\boldsymbol{\mu}}_l-\hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m}\sqrt{\mathbf{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l+\frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}_k},\quad k=1,...,m,\ l<l'=1,...,g. \ \square$$

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

Proposition 3.4. *Let $\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},...,\boldsymbol{x}_{in_i}$ be a random sample of size n_i from Kotz type distribution with $\boldsymbol{\mu}_i$ and Σ_i, $i=1,...,g$. 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 $\boldsymbol{\mu}_l-\boldsymbol{\mu}_{l'}$, $l<l'=1,...,g$ are given by*

$$\boldsymbol{a}_k'(\hat{\boldsymbol{\mu}}_l-\hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m}\sqrt{\boldsymbol{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l+\frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\boldsymbol{a}_k},\quad k=1,...,m \tag{3.4}$$

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

$$\widehat{\boldsymbol{B}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\Sigma}_i^{-1}(\boldsymbol{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},$$

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

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 $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'}$, $l < l' = 1, ..., 3$. Let $\boldsymbol{\mu}_i = (\mu_{i1}, .., \mu_{ip})'$, $i = 1, ..., 3$. Then with the choices, $\mathbf{a}_1' = (1, 0, 0)$, $\mathbf{a}_2' = (0, 1, 0)$, and $\mathbf{a}_3' = (0, 0, 1)$ using (3.3), the 95% Bonferroni simultaneous confidence intervals for $\mu_{1j} - \mu_{2j}$, $j = 1, ..., 3$ are:

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

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

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

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

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

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 $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$. We assume that $\boldsymbol{\Sigma}_i$'s have AR(1) structures.

(i) Assume that $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$, 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 $\mathbf{a}_k'(\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'})$, $l < l' = 1, ..., g$, are given by

$$\mathbf{a}_k'(\hat{\boldsymbol{\mu}}_l - \hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m} \sqrt{\mathbf{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l + \frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}_k}, \quad k = 1, ..., m \tag{3.5}$$

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, $\widehat{\boldsymbol{\tau}}_i = \widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{A}}_i^{-1}\widehat{\mathbf{B}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\boldsymbol{\Sigma}}$, $i = l, l' = 1, ..., g$ and

$$\widehat{\mathbf{B}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},$$

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

(ii) Assume that $\Sigma_i's$ are different, that is σ_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 $\mathbf{a}_k'(\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'})$, $l < l' = 1, ..., g$, are given by

$$\mathbf{a}_k'(\hat{\boldsymbol{\mu}}_l - \hat{\boldsymbol{\mu}}_{l'}) \pm z_{\alpha/2m}\sqrt{\mathbf{a}_k'(\frac{1}{n_l}\hat{\boldsymbol{\tau}}_l + \frac{1}{n_{l'}}\hat{\boldsymbol{\tau}}_{l'})\mathbf{a}_k}, \quad k = 1, ..., m \tag{3.6}$$

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, $\widehat{\boldsymbol{\tau}}_i = \widehat{\boldsymbol{\Sigma}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\mathbf{B}}_i\widehat{\mathbf{A}}_i^{-1}\widehat{\boldsymbol{\Sigma}}_i$, $i = l, l' = 1, ..., g$ and

$$\widehat{\mathbf{B}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'}{(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{x}_{ij}-\hat{\boldsymbol{\mu}}_i)},$$

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

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 $\boldsymbol{\mu}_l - \boldsymbol{\mu}_{l'}$, $l < l' = 1, ..., 3$. 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 $\mu_{1j} - \mu_{2j}$, $j = 1, ..., 3$ 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:

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

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

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

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 $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ be the probability density functions of a $p \times 1$ random vector $\mathbf{X}$ under populations Π_1 and Π_2 respectively. Let Ω be the sample space and $R_1 \subset \Omega$ be the set of $\mathbf{x}$ values for which we classify objects into Π_1 and $R_2 = \Omega - R_1$ be the remaining $\mathbf{x}$ values for which we classify objects into Π_2.

The conditional probability $P(j/i)$ of classifying an object into Π_j when it is from Π_i is

$$\begin{aligned} P(j/i) &= P(\mathbf{X} \in R_j/\Pi_i) \\ &= \int_{R_j} f_i(\mathbf{x})d\mathbf{x}, \quad \text{with } P(i/i) = 1 - \sum_{j=1}^{g} P(j/i),\ i,j = 1,2 \text{ and } i \neq j. \end{aligned}$$

Let p_i be the prior probability of Π_i such that $p_1 + p_2 = 1$.

$$\begin{aligned} P(\text{the object is correctly classified into } \Pi_i) &= P(\text{the object comes from } \Pi_i \\ &\quad \text{and is correctly classified into } \Pi_i) \\ &= P(\mathbf{X} \in R_i/\Pi_i)P(\Pi_i) \\ &= P(i/i)p_i,\ i = 1,2. \\ P(\text{the object is misclassified to } \Pi_i) &= P(\text{the object comes from } \Pi_j \\ &\quad \text{and is misclassified to } \Pi_i) \\ &= P(\mathbf{X} \in R_i/\Pi_j)P(\Pi_j) \\ &= P(i/j)p_j,\ i,j = 1,2 \text{ and } i \neq j. \end{aligned}$$

The costs of misclassification can be defined by a cost matrix

		Classify into:	
		Π_1	Π_2
True Population:	Π_1	0	$c(2/1)$
	Π_2	$c(1/2)$	0

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 $\boldsymbol{x}$ for which the following inequalities hold, that is,*

$$R_1 : \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq \frac{c(1/2)}{c(2/1)}\frac{p_2}{p_1} \quad \textit{and } R_2 : \textit{ otherwise.}$$

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

$$R_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \quad \text{and } R_2 : \text{otherwise.} \tag{4.1}$$

4.2.1 Classification Rules Under Kotz Type Distributions and General Covariance Structure

We now assume that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are Kotz type densities with parameters $\boldsymbol{\mu}_1$, Σ_1 and $\boldsymbol{\mu}_2$, Σ_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 $\mathbf{X}$ for populations Π_1 and Π_2 are given by

$$f_i(\mathbf{x}) = c \mid \Sigma \mid^{-\frac{1}{2}} \exp\{-[(\mathbf{x}-\boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)]^{\frac{1}{2}}\},\ \boldsymbol{\mu}_i \in \Re^p,\ \Sigma\ p.\ d., i = 1,2,$$

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

Suppose $\mathbf{x}_{11}, \mathbf{x}_{12}, ..., \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, \mathbf{x}_{22}, ..., \mathbf{x}_{2n_2}$ are two random samples of sizes n_1 and n_2 from Kotz type populations, Π_1 and Π_2 with unknown parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ 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

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

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 Π_1 and Π_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 $\boldsymbol{x}_0$, a new observation vector to be classified, to Π_1 if

$$(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\widehat{\Sigma}^{-1}\boldsymbol{x}_0 - \frac{1}{2}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\widehat{\Sigma}^{-1}(\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2) \geq 0, \qquad (4.2)$$

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

Allocate $\boldsymbol{x}_0$ to Π_2 otherwise.

We note that this rule is similar to the rule based on the normal densities except that here we estimate $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ differently; that is, by the optimization method mentioned in Chapter 2.

Case 2. $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

Suppose the Kotz type densities of $\mathbf{X}$ for populations Π_1 and Π_2 are given by

$$f_i(\mathbf{x}) = c \mid \boldsymbol{\Sigma}_i \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)]^{\frac{1}{2}}\},\ \boldsymbol{\mu}_i \in \Re^p,\ \boldsymbol{\Sigma}_i\ p.\ d., i = 1, 2,$$

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

Let $\mathbf{x}_{11}, \mathbf{x}_{12}, ..., \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, \mathbf{x}_{22}, ..., \mathbf{x}_{2n_2}$ be two random samples of sizes n_1 and n_2 from Kotz type populations Π_1 and Π_2 with unknown parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_2$, 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

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

$\hat{R}_2$: 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 Π_1 and Π_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 $\boldsymbol{x}_0$ to Π_1 if*

$$\sqrt{(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_2)'\widehat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_2)} - \sqrt{(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_1)'\widehat{\boldsymbol{\Sigma}}_1^{-1}(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_1)} - \frac{1}{2}\ln\left(\frac{\mid \widehat{\boldsymbol{\Sigma}}_1 \mid}{\mid \widehat{\boldsymbol{\Sigma}}_2 \mid}\right) \ \geq\ 0, \tag{4.3}$$

that is, if

$$\sqrt{(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_1)'\widehat{\boldsymbol{\Sigma}}_1^{-1}(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_1)} + \frac{1}{2}\ln \mid \widehat{\boldsymbol{\Sigma}}_1 \mid \ \leq\ \sqrt{(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_2)'\widehat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{x}_0-\hat{\boldsymbol{\mu}}_2)} + \frac{1}{2}\ln \mid \widehat{\boldsymbol{\Sigma}}_2 \mid,$$

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

Allocate $\boldsymbol{x}_0$ to Π_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, $\mathbf{x}_1$ = sepal length, $\mathbf{x}_2$ = sepal width, $\mathbf{x}_3$ = petal length and $\mathbf{x}_4$ = 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 $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1$ and Σ_2 under Kotz type distributons. The results are as follows:

Π_1 (Iris Versicolor) : $n_1 = 50$,

$$\hat{\boldsymbol{\mu}}_1 = \left(5.9301907,\ 2.7846629,\ 4.2389524,\ 1.3164927\right)',$$

$$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.0651039 & 0.0220169 & 0.0426881 & 0.0135329 \\ 0.0220169 & 0.022449 & 0.0206865 & 0.0101087 \\ 0.0426881 & 0.0206865 & 0.0489653 & 0.0167524 \\ 0.0135329 & 0.0101087 & 0.0167524 & 0.0088665 \end{pmatrix},$$

$$\text{and } \widehat{\boldsymbol{\Sigma}}_1^{-1} = \begin{pmatrix} 39.706180 & -18.48802 & -37.56574 & 31.451537 \\ -18.48802 & 101.38476 & 8.9867497 & -104.3503 \\ -37.56574 & 8.9867497 & 94.079695 & -130.6641 \\ 31.451537 & -104.3503 & -130.6641 & 430.62776 \end{pmatrix}.$$

Π_2 (Iris Virginica) : $n_2 = 50$,

$$\hat{\boldsymbol{\mu}}_2 = \left(6.5506952,\ 2.9737207,\ 5.5073221,\ 2.0322709\right)',$$

$$\widehat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 0.0836885 & 0.0193137 & 0.0613979 & 0.0109602 \\ 0.0193137 & 0.021707 & 0.0147842 & 0.0108312 \\ 0.0613979 & 0.0147842 & 0.0625200 & 0.0110904 \\ 0.0109602 & 0.0108312 & 0.0110904 & 0.0175368 \end{pmatrix},$$

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

Iris Versicolor				Iris Virginica			
Sepal length	Sepal width	Petal Length	Petal width	Sepal length	Sepal width	Petal Length	Petal width
7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0
6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8

$$\text{and } \widehat{\Sigma}_2^{-1} = \begin{pmatrix} 45.953273 & -15.75067 & -42.84096 & 8.1010053 \\ -15.75067 & 77.428683 & 4.3874471 & -40.75300 \\ -42.84096 & 4.3874471 & 59.427269 & -13.51723 \\ 8.1010053 & -40.75300 & -13.51723 & 85.678729 \end{pmatrix}.$$

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

$$\sqrt{(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_1)'\widehat{\Sigma}_1^{-1}(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_1)}$$

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

and distance from $\mathbf{x}_0$ to $\hat{\boldsymbol{\mu}}_2$ as

$$\sqrt{(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_2)'\widehat{\Sigma}_2^{-1}(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_2)}$$

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

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by " allocate $\mathbf{x}_0$ to Iris Versicolor specie (Π_1) if the distance from $\mathbf{x}_0$ to $\hat{\boldsymbol{\mu}}_1$ plus $\frac{1}{2}\ln|\widehat{\Sigma}_1| = -8.496731$ is smaller than or equal to the distance from $\mathbf{x}_0$ to $\hat{\boldsymbol{\mu}}_2$ plus $\frac{1}{2}\ln|\widehat{\Sigma}_2| = -7.519708$." Otherwise, allocate $\mathbf{x}_0$ to Iris Virginica specie (Π_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 $\mathbf{X}$ for populations Π_1 and Π_2 are given by

$$f_i(\mathbf{x}) = c\left(\frac{(\sigma_i^2)^p}{1-\rho_i^2}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{\sigma_i}\sqrt{(\mathbf{x}-\boldsymbol{\mu}_i)'\left[\mathbf{I}_p + \rho_i^2\mathbf{C}_1 - \rho_i\mathbf{C}_2\right](\mathbf{x}-\boldsymbol{\mu}_i)}\right\},\ i = 1, 2,$$

where $c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are defined in (2.5), (2.6).

Case 1. $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and $\rho_1 = \rho_2 = \rho$.

Let $\mathbf{x}_{11}, \mathbf{x}_{12}, ..., \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, \mathbf{x}_{22}, ..., \mathbf{x}_{2n_2}$ be two random samples of sizes n_1 and n_2 from Kotz type populations Π_1 and Π_2 with unknown parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and a positive definite common covariance matrix $(p+1)\mathbf{\Sigma}$, 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

$$\begin{aligned} \hat{R}_1 \ &: \ \exp\ \{-\frac{1}{\hat{\sigma}}\sqrt{D_{11} + \hat{\rho}^2 D_{21} - \hat{\rho} D_{31}} + \frac{1}{\hat{\sigma}}\sqrt{D_{12} + \hat{\rho}^2 D_{22} - \hat{\rho} D_{32}}\} \geq 1 \\ \text{and } \hat{R}_2 \ &: \ \text{otherwise.} \end{aligned}$$

Here

$$D_{11} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_1)'\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_1), \tag{4.4}$$

$$D_{21} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_1)'\ \mathbf{C}_1\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_1), \tag{4.5}$$

$$D_{31} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_1)'\ \mathbf{C}_2\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_1), \tag{4.6}$$

$$D_{12} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_2)'\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_2), \tag{4.7}$$

$$D_{22} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_2)'\ \mathbf{C}_1\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_2), \tag{4.8}$$

$$D_{32} = (\mathbf{x} - \hat{\boldsymbol{\mu}}_2)'\ \mathbf{C}_2\ (\mathbf{x} - \hat{\boldsymbol{\mu}}_2). \tag{4.9}$$

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 Π_1 and Π_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 $\boldsymbol{x}_0$ to Π_1 if*

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

This classification function is equivalent to allocating $\boldsymbol{x}_0$ to Π_1 if

$$(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\boldsymbol{x}_0 + \hat{\rho}^2(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\boldsymbol{C}_1\boldsymbol{x}_0 - \hat{\rho}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\boldsymbol{C}_2\boldsymbol{x}_0 - k \ \geq \ 0, \tag{4.10}$$

where $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\mu}}_2$ and $\hat{\rho}$ are MLEs of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and ρ respectively and $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are as in (2.5) and (2.6) and

$$k = \frac{1}{2}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'(\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2) + \frac{1}{2}\hat{\rho}^2(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\boldsymbol{C}_1(\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2) + \frac{1}{2}\hat{\rho}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)'\boldsymbol{C}_2(\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2).$$

Allocate $\boldsymbol{x}_0$ to Π_2 otherwise.

Case 2. $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, that is, σ_i^2's and ρ_i's are different.

Let $\mathbf{x}_{11}, \mathbf{x}_{12}, ..., \mathbf{x}_{1n_1}$ and $\mathbf{x}_{21}, \mathbf{x}_{22}, ..., \mathbf{x}_{2n_2}$ be two random samples of sizes n_1 and n_2 from Kotz type populations, Π_1 and Π_2 with unknown parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_2$, 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

$$\begin{aligned} \hat{R}_1 \ &: \ \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\right)^{-\frac{p}{2}}\left(\frac{1-\hat{\rho}_2^2}{1-\hat{\rho}_1^2}\right)^{-\frac{1}{2}} \\ &\quad \exp\ \{-\frac{1}{\hat{\sigma}_1}\sqrt{D_{11} + \hat{\rho}_1^2 D_{21} - \hat{\rho}_1 D_{31}} + \frac{1}{\hat{\sigma}_2}\sqrt{D_{12} + \hat{\rho}_2^2 D_{22} - \hat{\rho}_2 D_{32}}\ \} \geq 1 \\ \text{and } \hat{R}_2 \ &: \ \text{otherwise.} \end{aligned}$$

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 Π_1 and Π_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 $\boldsymbol{x}_0$ to Π_1 if

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

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

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

Alternatively, allocate $\boldsymbol{x}_0$ to Π_1 if

$$\frac{1}{\hat{\sigma}_1}\sqrt{D_{11}+\hat{\rho}_1^2 D_{21}-\hat{\rho}_1 D_{31}}+\frac{1}{2}\ln\Big(\frac{(\hat{\sigma}_1^2)^p}{1-\hat{\rho}_1^2}\Big) \le \frac{1}{\hat{\sigma}_2}\sqrt{D_{12}+\hat{\rho}_2^2 D_{22}-\hat{\rho}_2 D_{32}}+\frac{1}{2}\ln\Big(\frac{(\hat{\sigma}_2^2)^p}{1-\hat{\rho}_2^2}\Big),$$

$$\frac{1}{\hat{\sigma}_1}\sqrt{D_{11}+\hat{\rho}_1^2 D_{21}-\hat{\rho}_1 D_{31}}+\frac{1}{2}\ln\Big(\frac{(\hat{\sigma}_1^2)^p}{1-\hat{\rho}_1^2}\Big) \le \frac{1}{\hat{\sigma}_2}\sqrt{D_{12}+\hat{\rho}_2^2 D_{22}-\hat{\rho}_2 D_{32}}+\frac{1}{2}\ln\Big(\frac{(\hat{\sigma}_2^2)^p}{1-\hat{\rho}_2^2}\Big)$$

otherwise, allocate $\boldsymbol{x}_0$ *to* Π_2.

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 $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, σ_1, σ_2, ρ_1 and ρ_2 under Kotz type distributions. The results are as follows:

Π_1(Iris Versicolor) : $n_1 = 50$,

$$\hat{\boldsymbol{\mu}}_1 = \Big(5.9048405,\ 2.7879834,\ 4.2520208,\ 1.3172935\Big)',$$
$$\hat{\sigma}_1 = 0.1523249 \text{ and } \hat{\rho}_1 = 0.5391383.$$

Π_2(Iris Virginica) : $n_2 = 50$,

$$\hat{\boldsymbol{\mu}}_2 = \Big(6.5320442,\ 2.9815112,\ 5.4915763,\ 2.04259659\Big)',$$
$$\hat{\sigma}_2 = 0.1909728 \text{ and } \hat{\rho}_2 = 0.389897.$$

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 $\mathbf{x}_0$ to Iris Versicolor specie (Π_1) if

$$\tfrac{1}{0.1909728}\Big(D_{12}+0.15201967D_{22}-0.389897D_{32}\Big)^{\frac{1}{2}} - \tfrac{1}{0.1523249}\Big(D_{11}+0.290670106D_{21}-0.5391383D_{31}\Big)^{\frac{1}{2}} + 0.815193 \ge 0,$$

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

Allocate $\mathbf{x}_0$ to Iris Virginica specie (Π_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

$$AER = p_1 \int_{\hat{R}_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{\hat{R}_1} f_2(\mathbf{x})d\mathbf{x},$$

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 $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$. 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 Π_1 and Π_2 respectively, the confusion matrix has the form

		Predicted membership Π_1	membership Π_2	
Actual membership	Π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1
	Π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

where

$$n_{iC} \;=\; \text{number of } \Pi_i \text{ items correctly classified into } \Pi_i \text{ items, } i = 1,\ 2$$

$$n_{iM} \;=\; \text{number of } \Pi_i \text{ items misclassified into } \Pi_j \text{ items, } i,\ j = 1,\ 2 \text{ and } i \neq j.$$

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: (Π_1: Iris Versicolor, Π_2: Iris Virginica).

		Predicted membership Π_1	Predicted membership Π_2	
Actual membership	Π_1	46	4	50
	Π_2	1	49	50

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:

		Predicted membership Π_1	Predicted membership Π_2	
Actual membership	Π_1	47	3	50
	Π_2	1	49	50

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 Π_1 and Π_2 with the following means and variance covariance matrices; where

$$\boldsymbol{\mu}_1 = \left(5.9301910,\ 2.7846634,\ 4.2389532,\ 1.3164930\right)',$$

$$\Sigma_1 = \begin{pmatrix} 0.0651017 & 0.0220158 & 0.042686 & 0.0135321 \\ 0.0220158 & 0.0224482 & 0.0206853 & 0.0101082 \\ 0.042686 & 0.0206853 & 0.0489631 & 0.0167515 \\ 0.0135321 & 0.0101082 & 0.0167515 & 0.0088661 \end{pmatrix},$$

$$\boldsymbol{\mu}_2 = \left(6.5506954,\ 2.9737208,\ 5.5073221,\ 2.0322710\right)',$$

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

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:

		Predicted membership		
		Π_1	Π_2	
Actual membership	Π_1	30	0	30
	Π_2	1	29	30

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:

		Predicted membership Π_1	Π_2	
Actual membership	Π_1	29	1	30
	Π_2	1	29	30

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 $f_i(\mathbf{x})$ be the probability density function of a $p \times 1$ random vector $\mathbf{X}$ under population Π_i, $i = 1, ..., g$.

$$\begin{aligned} \text{Let } p_i &= \text{the prior probability of population } \Pi_i, \\ c(j/i) &= \text{the cost of allocating an object into } \Pi_j \text{ when it comes from } \Pi_i, \\ R_j &= \text{the set of } \mathbf{x}\text{'s classified into } \Pi_j \\ \text{and } P(j/i) &= P(\text{classifying an object into } \Pi_j \text{ when it is from } \Pi_i) \\ &= \int_{R_j} f_i(\mathbf{x}) d\mathbf{x} \quad \text{with } P(i/i) = 1 - \sum_{j=1,\ j\neq i}^{g} P(j/i). \end{aligned}$$

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 $\boldsymbol{x}$ to population $\Pi_j, j = 1, ..., g$ 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 $\mathbf{x}_0$, a new observation, into population Π_j if

$$f_j(\mathbf{x}_0) > f_i(\mathbf{x}_0),\ \text{for } i = 1, ..., g,\ i \neq j.$$

Alternatively, allocate $\mathbf{x}_0$ into population Π_j if

$$\ln f_j(\mathbf{x}_0) > \ln f_i(\mathbf{x}_0),\ \text{for } i = 1, ..., g,\ i \neq j.$$

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 $f_i(\mathbf{x})$ is a Kotz type density with parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, $i = 1, ..., g$. We consider the two cases, of equal and unequal covariance assumptions.

Case 1. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = ... = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$.

The Kotz type densities of $\mathbf{X}$ for population Π_i with parameters $\boldsymbol{\mu}_i$ and a common variance covariance matrix $(p+1)\boldsymbol{\Sigma}$, $i = 1, ..., g$ are given by

$$f_i(\mathbf{x}) = c \mid \boldsymbol{\Sigma} \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)]^{\frac{1}{2}}\},\ \boldsymbol{\mu}_i \in \Re^p,\ \boldsymbol{\Sigma}\ p.\ d.,\ i = 1, ..., g,$$

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

Let $\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in_i}$ 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 Π_i, $i = 1, ..., g$ 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 $\boldsymbol{x}_0$ *to* Π_j *if*

$$\hat{d}_j(\boldsymbol{x}_0) = \max_{1\le i\le g} \hat{d}_i(\boldsymbol{x}_0),$$

where

$$\hat{d}_i(\boldsymbol{x}_0) = \ln f_i(\boldsymbol{x}_0) = -\frac{1}{2}\ln|\widehat{\boldsymbol{\Sigma}}| - \sqrt{(x_0-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(x_0-\hat{\boldsymbol{\mu}}_i)}.$$

This classification function is equivalent to allocating $\boldsymbol{x}_0$ *to* Π_j *if*

$$\begin{aligned}\hat{d}_{ji}(\boldsymbol{x}_0) &= d_j(\boldsymbol{x}_0) - d_i(\boldsymbol{x}_0)\\ &= (\hat{\boldsymbol{\mu}}_j-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{x}_0 - \frac{1}{2}(\hat{\boldsymbol{\mu}}_j-\hat{\boldsymbol{\mu}}_i)'\widehat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_j+\hat{\boldsymbol{\mu}}_i) \ge 0, \ \textit{for } i=1,...g, \ i\ne j,\end{aligned} \tag{4.12}$$

where $\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\mu}}_j$ *and* $\widehat{\boldsymbol{\Sigma}}$ *are MLEs of* $\boldsymbol{\mu}_i, \boldsymbol{\mu}_j$ *and* $\boldsymbol{\Sigma}$ *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 $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$ by the optimization method mentioned in Chapter 2.

Case 2. $\boldsymbol{\Sigma}_i$'s are different

The Kotz type densities of $\mathbf{X}$ for population Π_i with parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$, $i=1,...,g$ are given by

$$f_i(\mathbf{x}) = c \mid \boldsymbol{\Sigma}_i \mid^{-\frac{1}{2}} \exp\ \{-[(\mathbf{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)]^{\frac{1}{2}}\},\ \boldsymbol{\mu}_i \in \Re^p,\ \boldsymbol{\Sigma}_i\ p.\ d.,\ i=1,...,g,$$

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

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

Result 4.8 *Let the populations* $\Pi_i,\ i=1,...,g$ *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 $\boldsymbol{x}_0$ *to* Π_j *if*

$$\hat{d}_j(\boldsymbol{x}_0) = \max_{1\le i\le g} \hat{d}_i(\boldsymbol{x}_0),$$

where $\hat{d}_i(\boldsymbol{x}_0) = \ln f_i(\boldsymbol{x}_0) = -\frac{1}{2}\ln|\hat{\Sigma}_i| - \sqrt{(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)'\,\widehat{\Sigma}_i^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)},\ i = 1, ..., g.$

This classification function is equivalent to allocating $\boldsymbol{x}_0$ *to* Π_j *if*

$$\begin{aligned} \hat{d}_{ji}(\boldsymbol{x}_0) &= d_j(\boldsymbol{x}_0) - d_i(\boldsymbol{x}_0) \\ &= \sqrt{(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)'\widehat{\Sigma}_i^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)} - \sqrt{(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_j)'\widehat{\Sigma}_j^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_j)} - \frac{1}{2}\ln\Big(\frac{|\widehat{\Sigma}_j|}{|\widehat{\Sigma}_i|}\Big) \geq 0, \end{aligned} \tag{4.13}$$

that is, if

$\sqrt{(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_j)'\widehat{\Sigma}_j^{-1}(\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_j)} + \frac{1}{2}\ln|\Sigma_j|$ *is smallest, (that is, if* $\ln f_j(\boldsymbol{x}_0)$ *is largest)*

where $\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\mu}}_j, \widehat{\Sigma}_i$ *and* $\widehat{\Sigma}_j$ *are MLEs of* $\boldsymbol{\mu}_i, \boldsymbol{\mu}_j, \Sigma_i$ *and* Σ_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 (Π_2), Iris Virginica (Π_3) displayed in Table 4.1 and Iris Sesota (Π_1) displayed in Table 4.2 respectively. Data on four variables, $\mathbf{x}_1$ = sepal length, $\mathbf{x}_2$ = sepal width, $\mathbf{x}_3$ = petal length and $\mathbf{x}_4$ = 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and the discrimination rule for the new group that we included in this example (as compared to Example 4.1).

Π_1 (Iris Setosa) : $n_1 = 50$,

$$\hat{\boldsymbol{\mu}}_1 = \Big(5.0003585,\ 3.4047069,\ 1.4585232,\ 0.2283304\Big)',$$

$$\widehat{\Sigma}_1 = \begin{pmatrix} 0.0276767 & 0.0213725 & 0.0033927 & 0.0020632 \\ 0.0213725 & 0.0301198 & 0.0023413 & 0.0022393 \\ 0.0033927 & 0.0023413 & 0.0057416 & 0.0009577 \\ 0.0020632 & 0.0022393 & 0.0009577 & 0.0021534 \end{pmatrix},$$

Table 4.2: Data on Iris Setosa.

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

$$\text{and } \widehat{\Sigma}_1^{-1} = \begin{pmatrix} 83.726660 & -56.73314 & -24.62621 & -10.27257 \\ -56.73314 & 74.871435 & 7.4663227 & -26.82105 \\ -24.62621 & 7.4663227 & 197.70168 & -72.09592 \\ -10.27257 & -26.82105 & -72.09592 & 534.18177 \end{pmatrix}.$$

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

$$\sqrt{(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_1)'\hat{\Sigma}_1^{-1}(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_1)} + \tfrac{1}{2}\ln \mid \widehat{\Sigma}_1 \mid$$

$$= \left[(x_{01} - 5.0003585, ..., x_{04} - 0.2283304)' \begin{pmatrix} 83.72666 & \dots & -10.27257 \\ \vdots & \ddots & \vdots \\ -10.27257 & \dots & 534.18177 \end{pmatrix} \begin{pmatrix} x_{01} - 5.0003585 \\ \vdots \\ x_{04} - 0.2283304 \end{pmatrix}\right]^{\frac{1}{2}}$$

$$- 9.700359,$$

$$\sqrt{(x_0 - \hat{\boldsymbol{\mu}}_2)'\hat{\Sigma}_2^{-1}(x_0 - \hat{\boldsymbol{\mu}}_2)} + \tfrac{1}{2}\ln \mid \widehat{\Sigma}_2 \mid$$

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

$$- 8.496731,$$

$$\text{and } \sqrt{(x_0 - \hat{\boldsymbol{\mu}}_3)'\hat{\Sigma}_3^{-1}(x_0 - \hat{\boldsymbol{\mu}}_3)} + \tfrac{1}{2}\ln \mid \widehat{\Sigma}_3 \mid$$

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

$$- 7.5197089.$$

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by " allocate $\mathbf{x}_0$ to Π_j if the distance from $\mathbf{x}_0$ to $\hat{\boldsymbol{\mu}}_j$ plus $\frac{1}{2}\ln \mid \widehat{\Sigma}_j \mid$, $j = 1,\ 2,\ 3$ is smallest."

4.3.2 Under AR(1) Structure

The Kotz type densities of $\mathbf{X}$ for populations Π_i with parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i, i = 1, ..., g$, are given by

$$f_i(\mathbf{x}) = c\Big(\frac{(\sigma_i^2)^p}{1-\rho_i^2}\Big)^{-\frac{1}{2}} \exp\ \{-\frac{1}{\sigma_i}\sqrt{(\mathbf{x}-\boldsymbol{\mu}_i)'\ [\mathbf{I}_p + \rho_i^2\mathbf{C}_1 - \rho_i\mathbf{C}_2]\ (\mathbf{x}-\boldsymbol{\mu}_i)}\}, i = 1, ..., g,$$

where $c = \frac{\Gamma(\frac{p}{2})}{2\pi^{\frac{p}{2}}\Gamma(p)}$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are as in (2.5), (2.6).

Case 1. $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = ... = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$, that is, $\sigma_1^2 = \sigma_2^2 = ... = \sigma_g^2 = \sigma^2$ and $\rho_1 = \rho_2 = ... = \rho_g = \rho$.

Suppose $\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{in_i}$ 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 Π_i, $i = 1, ..., g$ 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 $\boldsymbol{x}_0$ to Π_j if

$$\hat{d}_j(\boldsymbol{x}_0) = \max_{1\leq i\leq g}\ \hat{d}_i(\boldsymbol{x}_0),$$

where $\hat{d}_i(\boldsymbol{x}_0) = -\frac{1}{\hat{\sigma}}\sqrt{D_{1i} + \hat{\rho}^2 D_{2i} - \hat{\rho} D_{3i}}$ and

$$\begin{aligned} D_{1i} &= (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)'\ (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i), \\ D_{2i} &= (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)'\ C_1\ (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i), \\ D_{3i} &= (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i)'\ C_2\ (\boldsymbol{x}_0 - \hat{\boldsymbol{\mu}}_i). \end{aligned}$$

This classification function is equivalent to allocating $\boldsymbol{x}_0$ to Π_j if

$$\hat{d}_{ji}(\boldsymbol{x}_0) = (\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}_i)'\boldsymbol{x}_0 + \hat{\rho}^2(\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}_i)' C_1\boldsymbol{x}_0 - \hat{\rho}(\hat{\boldsymbol{\mu}}_j - \hat{\boldsymbol{\mu}}_i)' C_2\boldsymbol{x}_0 - k_1 \geq 0, \quad (4.14)$$

where $\hat{\boldsymbol{\mu}}_j$, $\hat{\boldsymbol{\mu}}_i$, $\hat{\sigma}$ and $\hat{\rho}$ are MLEs of $\boldsymbol{\mu}_j$, $\boldsymbol{\mu}_i$, σ and ρ respectively and

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

Case 2. **$\boldsymbol{\Sigma}_i$'s are different**, that is, σ_i^2's and ρ_i's are different.

Result 4.10 *Let the populations* Π_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 $\boldsymbol{x}_0$ *to* Π_j *if*

$$\hat{d}_j(\boldsymbol{x}_0) = \max_{1\leq i\leq g} \hat{d}_i(\boldsymbol{x}_0),$$

where $\hat{d}_i(\boldsymbol{x}_0) = -\frac{1}{2}\ln\left(\frac{(\sigma_i^2)^p}{1-\rho_i^2}\right) - \frac{1}{\hat{\sigma}_i}\sqrt{D_{1i} + \hat{\rho}_i^2 D_{2i} - \hat{\rho}_i D_{3i}}.$

This classification function is equivalent to allocating $\boldsymbol{x}_0$ *to* Π_j *if*

$$\hat{d}_{ji}(\boldsymbol{x}_0) = \frac{1}{\hat{\sigma}_i}\sqrt{D_{1i} + \hat{\rho}_i^2 D_{2i} - \hat{\rho}_i D_{3i}} - \frac{1}{\hat{\sigma}_j}\sqrt{D_{1j} + \hat{\rho}_j^2 D_{2j} - \hat{\rho}_j D_{3j}} - k_1 \geq 0$$

where $k_1 = \frac{p}{2}\ln\left(\frac{\hat{\sigma}_j^2}{\hat{\sigma}_i^2}\right) + \frac{1}{2}\ln\left(\frac{1-\hat{\rho}_i^2}{1-\hat{\rho}_j^2}\right),$

that is, if

$$\frac{1}{\hat{\sigma}_j}\sqrt{D_{1j} + \hat{\rho}_j^2 D_{2j} - \hat{\rho}_j D_{3j}} + \frac{1}{2}\ln\left(\frac{(\hat{\sigma}_j^2)^p}{1-\hat{\rho}_j^2}\right) \text{ is smallest.} \tag{4.15}$$

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.

Π_1 (Iris Sesota) : $n_1 = 50$,

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

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

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

$$\frac{1}{\hat{\sigma}_1}\left(D_{11} + \hat{\rho}_1^2 D_{21} - \hat{\rho}_1 D_{31}\right)^{\frac{1}{2}} + \frac{1}{2}\ln\left(\frac{(\hat{\sigma}_1^2)^p}{1-\hat{\rho}_1^2}\right)$$
$$= \frac{1}{0.1077129}\left(D_{11} + 0.208516436 D_{21} - 0.456636 D_{31}\right)^{\frac{1}{2}} - 8.796221$$

$$\frac{1}{\hat{\sigma}_2}\left(D_{12} + \hat{\rho}_2^2 D_{22} - \hat{\rho}_2 D_{32}\right)^{\frac{1}{2}} + \frac{1}{2}\ln\left(\frac{(\hat{\sigma}_2^2)^p}{1-\hat{\rho}_2^2}\right)$$
$$= \frac{1}{0.1523249}\left(D_{12} + 0.290670106 D_{22} - 0.5391383 D_{32}\right)^{\frac{1}{2}} - 7.355241$$

$$\frac{1}{\hat\sigma_3}\left(D_{13}+\hat\rho_3^2 D_{23}-\hat\rho_3 D_{33}\right)^{\frac{1}{2}}+\frac{1}{2}\ln\left(\frac{(\hat\sigma_3^2)^p}{1-\hat\rho_3^2}\right)$$
$$=\frac{1}{0.1909728}\left(D_{13}+0.152019436D_{23}-0.3898967D_{33}\right)^{\frac{1}{2}}-6.540048$$

The rule that minimizes the ECM under equal prior probabilities and equal misclassification costs is given by
allocate $\mathbf{x}_0$ to Π_j if $\frac{1}{\hat\sigma_j}\left(D_{1j}+\hat\rho_j^2 D_{2j}-\hat\rho_j D_{3j}\right)^{\frac{1}{2}}+\frac{1}{2}\ln\left(\frac{(\hat\sigma_j^2)^p}{1-\hat\rho_j^2}\right)$, $j=1,\ 2,\ 3$ 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, $i=1,...,g$ 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 (Π_1: Iris Setosa, Π_2: Iris Versicolor, Π_3: Iris Virginica).

		Predicted membership			
		Π_1	Π_2	Π_3	
	Π_1	50	0	0	50
Actual membership	Π_2	0	46	4	50
	Π_3	0	1	49	50

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:

		Predicted Π_1	membership Π_2	Π_3	
	Π_1	50	0	0	50
Actual membership	Π_2	0	47	3	50
	Π_3	0	1	49	50

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 Π_1, Π_2 and Π_3 with the following means and variance covariance matrices respectively where

$$\boldsymbol{\mu}_1 = \left(5.006,\ 3.428,\ 1.462\right)',$$

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

$$\boldsymbol{\mu}_2 = \left(5.936,\ 2.77,\ 4.26\right)',$$

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

$$\boldsymbol{\mu}_3 = \left(6.588,\ 2.974,\ 5.552\right)',$$

$$\Sigma_3 = \begin{pmatrix} 0.396256 & 0.091888 & 0.297224 \\ 0.091888 & 0.101924 & 0.069952 \\ 0.297224 & 0.069952 & 0.298496 \end{pmatrix}.$$

Now for these data the Lachenbruch's holdout procedure using Kotz type densities provides the following confusion matrix.

		Predicted membership			
		Π_1	Π_2	Π_3	
	Π_1	30	0	0	30
Actual membership	Π_2	1	29	5	35
	Π_3	0	7	33	40

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

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

		Predicted membership			
		Π_1	Π_2	Π_3	
	Π_1	29	1	0	30
Actual membership	Π_2	1	28	6	35
	Π_3	0	9	31	40

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

Ahn, S. K. (1992). F-probability plot and its application to multivariate normality. *Commun. Statist. -Theory Meth.* **21**, 997–1023.

Baringhaus, L. and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* **35**, 339–348.

Baringhaus, L. and Henze, N. (1992). Limit distributions for Mardia's measure of multivariate skewness. *Ann. Statist.* **20**, 1889–1902.

Beirlant, J., Mason, D. M. and Vynckier, C. (1999). Goodness-of-fit analysis for multivariate normality based on generalized quantiles. *Comp. Statist. Data Anal.* **30**, 119–142.

Bera, A. and John, S. (1983). Tests for multivariate normality with Pearson alternatives. *Commun. Statist. -Theory Meth.* **12**, 103–117.

Bowman, A. W. and Foster, P. J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.* **88**, 529–537.

Brown, B. M. (1983). Statistical use of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45**, 25–30.

Cox, D. R. and Small, N. J. H. (1978). Testing multivariate normality. *Biometrika* **65**, 263–272.

Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.* **14**, 708–723.

D' Agostino, R. B. (1986). Test for the normal distribution. In *Goodness-of-fit Techniques* (Edited by R. B. D'Agostino and M. A. Stephen), 367–413. Marcel Dekker, Inc., New York.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag, New York.

Ducharme, G. R. and Milasevic, P. (1987). Spatial median and directional data. *Biometrika* **74**, 212–215.

Epps, T. W. and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70**, 723–726.

Fang, K. T., and Anderson, T. W. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions.* Allerton Press, New York.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London.

Fang, K. T., Li, R. Z. and Liang, J. J. (1998). A multivariate version of Ghosh's T_3-plot to detect non-multinormality. *Comp. Statist. Data Anal.* **28**, 371–386.

Fattorini, L. (1986). Remarks on the use of the Shapiro-Wilk statistic for testing multivariate normality. *Statistica* **46**, 209–217.

Ghosh, S. and Ruymgaart, F. H. (1992). Applications of empirical characteristic functions in some multivariate problems. *Canad. J. Statist.* **20**, 429–440.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28**, 81–124.

Gómez, E., Gómez-Villegas, M. A. and Marín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Commun. Statist. -Theory Meth.* **27**, 589–600.

Gower, J.S. (1974). The mediancentre. *App. Statist.* **23**, 466–470.

Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35**, 414-415.

Hartley, A. M. and Naik, D. N. (2001). *Estimation of familial correlations under autoregressive circular covariance Commun. Statist. -Theory Meth.* **30**, 1811–1828.

Hasofer, A. M. and Stein, G. Z. (1990). Testing for multivariate normality after coordinate transformation. *Commun. Statist. -Theory Meth.* **19**, 1403–1481.

Hawkins, D. M. (1981). A new test for multivariate normality and homoscedasticity. *Technometrics* **23**, 105–110.

Healy, M. J. R. (1968). Multivariate normal plotting. *App. Statist.* **17**, 157–161.

Henze, N. (1994). On Mardia's kurtosis test for multivariate normality. *Commun. Statist. -Theory Meth.* **23**, 1031–1045.

Henze, N. (2002). Invariant tests for multivariate normality: A critical review. *Statist. Papers.* **43**, 467–506.

Henze, N. and Zirkler, B. (1990). A class of invariance consistent tests for multivariate normality. *Commun. Statist. -Theory Meth.* **19**, 3595–3617.

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

Isogai, T. (1982). On a measure of multivariate skewness and a test for multivariate normality. *Ann. Statist.* **34**, 531–541.

Isogai, T. (1989). On using influence functions for testing multivariate normality. *Ann. Statist.* **41**, 169–186.

Jarque, C. M. and McKenzie, C. R. (1995). Testing for multivariate normality in simultaneous equations models. *Commun. Statist. Comput. Simulation* **39**, 323–328.

Johnson, M. E. (1987). *Multivariate Statistical Simulation.* Wiley, New York.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

Johnson, R. A. and Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis.* 4th edition. Prentice Hall, Inc., New Jersey.

Kariya, T. and George, E. I. (1995). Locally best invariant tests for multivariate normality. *Sankhyā* **57**, 440–451.

Kariya, T., Tsay, R. S., Terui, N. and Li, H. (1999). Tests for multinormality with applications to time series. *Commun. Statist. -Theory Meth.* **28**, 519–536.

Khattree, R. and Naik, D. N. (1994). *Estimation of interclass correlation under circular covariance Biometrika* **81**, 612–617.

Khattree, R. and Naik, D. N. (1999). *Applied Multivariate Statistics with SAS Software.* SAS Institute, Inc., Cary North Carolina.

Koziol, J. A. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika* **69**, 423–427.

Koziol, J. A. (1983). On assessing multivariate normality. *J. Roy. Statist. Soc. Ser. B* **45**, 358–361.

Koziol, J. A. (1986). A note on the asymptotic distribution of Mardia's measure of multivariate kurtosis. *Commun. Statist. -Theory Meth.* **15**, 1507–1513.

Koziol, J. A. (1989). A note on measures of multivariate kurtosis. *Biometrika* **15**, 619–624.

Kuwana, Y. and Kariya T. (1991). LBI tests for multivariate normality in exponential power distributions. *J. Multiv. Anal.* **39**, 117–134.

Lachenbruch, P. A. and Mickey, M. R. (1963). Hierarchical grouping to optimize an objective function. In *J. Amer. Statist. Assoc.* (Edited by J. H. Ward Jr.), **58**, 236–244.

Liang, J. and Bentler, P. M. (1999). A t-distribution plot to detect non-multinormality. *Comp. Statist. Data Anal.* **30**, 31–44.

Liang, J., Li, R., Fang, H. and Fang, K. T. (2000). Testing multinormality based on low-dimensional projection. *J. Statist. Plann. Inference* **86**, 129–141.

Lin, C. T. and Mudholkar, G. S. (1980). A simple test for normality against asymmetric alternatives. *Biometrika* **67**, 455–461.

Loh, W. (1986). Testing multivariate normality by simulation. *Commun. Statist. Comput. Simulation* **26**, 243–253.

Machado, S. G (1983). Two statistics for testing for multivariate normality. *Biometrika* **70**, 713–718.

Malkovich, J. F. and Afify, A. A. (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68**, 176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519-530.

Mardia, K. and Foster, K. (1983). Omnibus tests of multinormality based on skewness and kurtosis. *Comp. Statist. Data Anal.* **12**, 207–221.

Mardia, K. and Kent, J. T. (1991). Rao score tests for goodness of fit and independence. *Biometrika* **78**, 355–363.

Mecklin, C. J. and Mundfrom, D. J. (2002). A Monte Carlo comparison of the type II error rates of tests of multivariate normality. *Commun. Statist. Comput. Simulation* Review (to appear).

Mecklin, C. J. and Mundfrom, D. J. (2003a). On using asymptotic critical values in testing for multivariate normality. *InterStat.* available on-line at http://interstat.stat.vt.edu/Interstat/ARTICLES/2003/articles/J03001.pdf.

Mecklin, C. J. and Mundfrom, D. J. (2003b). An appraisal and bibliography of tests for multivariate normality. *InterStat.* Review (to appear).

Moore, D. S. and Stubblebine, J. B. (1981). Chi-square tests for multivariate normality with application to common stock prices. *Commun. Statist. -Theory Meth.* **10**, 713–738.

Móri, T. F., Rohatgi, V. K. and Székely, G. J. (1993). On multivariate skewness and kurtosis. *Th. Prob. Appl.* **38**, 675–679.

Mudholkar, G. S., McDermott, M. and Srivastava, D. K. (1992). A test of p-variate normality. *Biometrika* **79**, 850–854.

Mudholkar, G. S., Srivastava, D. K. and Lin, C. T. (1995). Some p-variate adaptations of the Shapiro-Wilk test of normality. *Commun. Statist. -Theory Meth.* **24**, 953–985.

Naik, D. N. (1993). Multivariate medians: a review. In *Probability and Statistics.* (Edited by S. K. Basu and B. K. Sinha), 80–90. Publishing House, New Delhi, India.

Naik, D. N. and Patwardhan, G. R. (1991). A note on testing for correlation in a certain bivariate distribution. *J. Quan. Econ.* **7**, 295–302.

Naito, K. (1996). On weighting the studentized empirical characteristic function for testing normality. *Commun. Statist. Comput. Simulation* **25**, 201–213.

Paulson, A. S., Roohan, P. and Sullo, P. (1987). Some empirical distribution function tests for multivariate normality. *Commun. Statist. Comput. Simulation* **28**, 15–30.

Quiroz, A. J., Dudley, R. M. (1991). Some new tests for multivariate normality. *Prob. Theory Related Fields* **87**, 521–546 .

Rao. C. R. (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā Ser. A* **50**, 289–313.

Rencher, A. C. (2002). *Methods of Multivariate Analysis.* 2nd edition. Wiley, New York.

Romeu, J. L. and Ozturk, A. (1993). A comparative study of goodness-of-fit tests for multivariate normality. *J. Multiv. Anal.* **46**, 309–334.

Romeu, J. L. and Ozturk, A. (1996). A new graphical test for multivariate normality. *Amer. J. Manage. Sci.* **16**, 5–48.

Ronning, G. (1977). A simple scheme for generating multivariate gamma distributions with non-negative covariance matrix. *Technometrics* **19**, 179–183.

Royston, J. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk. *App. Statist.* **32**, 121–133.

Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc.* **63**, 1343–1372.

Simoni, S. de (1968). Miscellaneous real multivariate distributions. In *Distributions in Statistics: Continuous Multivariate Distributions.* (Edited by N. L. Johnson and S. Kotz), 298–299. Wiley, New York.

Singh, A. (1993). Omnibus robust procedures for assessment of multivariate normality and detecting of multivariate outliers. In *Multivariate Environmental Statistics.* (Edited by G. P. Patil and C. R. Rao), 445–488. Noth-Holland, Amsterdam.

Slate, E. H. (1999). Assessing multivariate normality using univariate distributions. *Biometrika* **86**, 191–202.

Small, N. (1980). Marginal skewness and kurtosis in testing multivariate normality. *App. Statist.* **29**, 85–87.

Smith, T. and Richardson, M. (1993). A test for multivariate normality in stock returns. *J. Business.* **66**, 295–321.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statist. Prob. Lett.* **2**, 263–267.

Tsai, K. T. and Koziol, J. A. (1988). A correlation type procedure for assessing multivariate normality. *Commun. Statist. Comput. Simulation* **17**, 623–651.

Ward, P. J. (1988). A comparison of the powers of some tests for multivariate normality. *ASA Proc. Statist. Comput.* 197–201.

Weiss, L. (1958). A test of fit for multivariate distribution. *Ann. Math. Statist.* **29**, 595–599.

Zhu, L., Fang, K. T. and Bhatti, M. I. (1997). On estimated proection pursuit Cramér-von Mises statistics. *J. Multiv. Anal.* **63**, 1–14.

Zhu, L. X., Wong, H. L. and Fang, K. T. (1995). A test for multivariate normality based on sample entropy and projection pursuit. *J. Statist. Plann. Inference* **45**, 373–385.

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;
kappa1=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;
    beta1hat=( sum(g_matrix#g_matrix#g_matrix)  )/(n*n);
    beta2hat=trace( g_matrix#g_matrix )/n;
    kappa1[k]=n*beta1hat/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_b1p=count1/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 $\boldsymbol{\mu}$ and general covariance $\boldsymbol{\Sigma}$.

```
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,p,0);
a_hat=j(p,p,0);    b_hat=j(p,p,0);    asym_mu=j(p,p,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,p,0); h1=j(p-1,1,0);
```

```
asy_lmu3=j(3,3,0); h3=j(3,1,0);
asym1=j(p,1,0);   low1=j(p,1,0);  up1=j(p,1,0);
asym2=j(p,1,0);   low2=j(p-1,1,0);  up2=j(p-1,1,0);
asym3=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 nlpnra(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 $APER$ and $\hat{E}(AER)$ using Kotz type densities for unequal general covariances for p=4 and g=2.

```
options ls=70;
proc iml;
/* Log-likelihood function:f_ha_1 */
start f_ha_1(x)
global(data1,data2,optn,
       con,p,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;
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,invsig1,invsig2,
               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)*invsig1*diff )  + 0.5*log(detsig1);
diff = data[,i] - muopt[,g2];
distsq2 = sqrt( t(diff)*invsig2*diff )  + 0.5*log(detsig2);
if (distsq2 < distsq1) then n12_m = 1;
finish check_dist;
start get_xbar_sigma(muopt_ha,invsig1,invsig2,
                     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);
vv1=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;
* 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,

          . . . .
          . . . .

          .   .   .   .
          .   .   .
          .   .
          .
```

```
          . . . .
          . . .
          . .
          .};

call nlpnra(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];
```

```
                  vv2[j,i]=xopt_ha[p2+k2];
                  k2=k2+1;
                  end;
  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,
               6.1000764 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,
```

```
          5.686537 2.7037714 4.3215518 1.3376282,
         5.5398378 2.9877456 4.1977785 1.4275273,
         6.5792779 2.8897342 4.3448149 1.2068549,
         5.7580576 2.6495915 4.3011515 1.2496377,
         5.7195076 2.7513127 3.8838301 1.2760272,
         6.2880095 2.7209261 4.5542073 1.4095898,
         5.6827247 2.9219009 4.2515982 1.4155149,
         5.9629738 2.8588019 4.3858777 1.3956366,
         5.2534427 2.8480558 4.1071094 1.2321512,
         5.7880453 2.7655088 4.1234011 1.2536415,
         6.4859478 3.2043381 4.2138811 1.3666392,
         5.9656021 3.0616776 4.1505009  1.312617,
         5.8308631 2.8428823 4.0287503 1.2345323,
         6.3202378 2.9035702 4.2867455 0.9928669,
          6.246117 3.2109154 4.0325683 1.3407778,
          5.284534 2.5278402 3.9808005 1.1471642};
tdata2={ 6.8462112  3.046601 5.9178479 1.8842269,
         7.1805166 3.3370394 5.9864498 2.2277146,
         6.2938092 3.0646918 5.3058226 2.0478006,
         6.0491458 2.8742568 5.4588787 2.4409555,
         5.5687688 2.9170615 4.7702703 1.8806772,
         5.5103595 3.1343121 5.6372549  2.315888,
         5.4857636 3.3183748 4.9128641 2.2565673,
         6.4811978 3.3335411 5.3803565 1.8987372,
         6.4331137  2.874984 5.3741692  1.758083,
         8.4994476 3.2233767 7.3671073  2.534746,
         6.4880198 3.1290922 5.2651431 2.1074788,
          6.356663 2.8471859 5.8390458 2.4549109,
         6.6537091 3.3155046 5.5364078 2.2072308,
         6.5688772 3.0201839 5.7733679 2.3116469,
          7.357926 2.6389793 6.7314072 1.5106304,
```

```
                     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,invsig2,invsig1,
                  detsig2,detsig1,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=no1 + no2 ;
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.
	1991 - 1992 Invited Lecturer, College of Nursing, Rachaburi, Thailand.
	1986 - 1991 Government Official, Graduate School, Silpakorn University, Bangkok, Thailand.
	1983 - 1984 Statistician, Office of Agricultural Economics, Ministry of Agriculture and Cooperatives, Bangkok, Thailand.

This document was prepared by the author using LaTeX.