ON SCALE-SCALE CURVES FOR MULTIVARIATE DATA BASED ON RANK REGIONS

by

PRITHA GUHA

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
College of Engineering and Physical Sciences
The University of Birmingham
May 2012

Abstract

Quantile-quantile plots are in use to compare univariate distributions for a long time, but as there is no ordering in higher dimension, there is no straight forward generalisation of quantiles for the multivariate data and hence there is no visual tool which can be considered as a generalisation of quantile-quantile plots to compare multivariate distributions. In this work we have considered some notions of multivariate ranks, quantiles and data depths. Based on spatial rank, we have constructed central rank regions and some measures of scale. We proposed a scale-scale plot, which can be used to compare multivariate distributions. Under spherical symmetry, our scale curves have some nice closed form formula, however they are not equivariant under affine transformations. We discussed this issue with illustrations and proposed an affine equivariant version based on data-driven transformations. We established some characterisation results for the proposed affine equivariant scale curves under elliptic symmetry and used the fact to propose some visual test of location and scale in the family of elliptically symmetric distributions. Our proposed scale-scale plot is based on volume functionals of central rank region. We gave some asymptotic results regarding the distribution of the volume functional and constructed a test statistic based on the volume functional. We proposed some asymptotic results regarding the distribution of the test statistic and also studied the power of the proposed test of multivariate normality. As further applications to our scale-scale plots, we discuss the behaviour of our proposed scale-scale plots when the distribution is not elliptically symmetric with illustrations and study the power of the test of for skew elliptic and g and h distribution based on the previously defines test statistic. Among other application of the scale-scale plots, we propose a kurtosis plot, which can be used to study the peakedness and tail behaviour of the multivariate distributions, a visual test of location and scale.

To Thamma,

Tattu, Baba, Ma, Apratim

and Rudranee

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Biman Chakraborty whose encouragement enabled me to develop an understanding of the subject. This thesis would not have been possible without his invaluable guidance. I would like to thank Dr. Hui Li for her continuous support. I would also like to thank Dr. Christian Hennig for his valuable comments on my qualifier thesis which led to various new ideas. I would like to thank all my friends in the School of Mathematics for four wonderful years.

I would like to express my deepest gratitude to my husband for his endless support throughout the time of my study. A special thanks to my mother and to the rest of my family for their support.

Finally a big thanks to my baby daughter for being an angel while I was writing my thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Characterising and describing the underlying distribution of a data set is the basis of statistical analysis. In many statistical analyses some quantitative measures are used to describe various features of the data and summarise them. Suppose that $X_1, X_2, \cdots, X_n$ are independently and identically distributed (i.i.d) observations in $\mathbb{R}$ with a distribution, which is continuous with respect to the Lebesgue measure and symmetric about some unknown $\theta \in \mathbb{R}$, i.e. $f(x-\theta) = f(\theta - x)$ for all $x \in \mathbb{R}$, where f is the common density function. The distribution can be characterised in many ways, for example, it can be characterised by a parametric model or by a parametric model which is contaminated by a small non-parametric mixture (see Bickel and Lehmann (1975a)). One of the important characteristic of the underlying distribution would be some measure of location of the centre. When the distribution is symmetric about a parameter in some sense, a natural location parameter would be the centre of symmetry. Bickel and Lehmann (1975b) described the required properties of a measure of location. For example, a natural choice for the centre of symmetry for a symmetric univariate distribution is the median of the distribution.

The notions of symmetry for multivariate distributions can have many variations: it could be based on the properties of the distribution function, of the characteristic function or of the density function. One may alternatively demand invariance in relation to specific classes of transformations. In the following section we introduce a few of the most popular

concepts of symmetry for multivariate distributions.

1.1 Concepts of Symmetry for Multivariate Distributions

Many of the examples of multivariate symmetry of a random vector $\mathbf{X}$ about some location parameter $\boldsymbol{\theta}$ can be expressed in terms of some suitable transformations of the centred vector $\mathbf{X} - \boldsymbol{\theta}$. We now discuss a few of such concepts.

$\mathbf{X}$ is said to have a *spherically symmetric distribution* about a point $\boldsymbol{\theta}$, in $\mathbb{R}^d$, if the distribution of $\mathbf{X} - \boldsymbol{\theta}$ remains unchanged under any orthogonal transformation. Some of the oldest references on spherically symmetric distributions are Maxwell (1860), Bartlett (1934) and Hartman and Wintner (1940). An excellent review of spherical symmetry, can be found in Chmielewski (1981).

For a spherically symmetric distribution, the length of the random vector $||\mathbf{X}-\boldsymbol{\theta}||$ and its direction vector $\frac{\mathbf{X}-\boldsymbol{\theta}}{||\mathbf{X}-\boldsymbol{\theta}||}$ are distributed independently, see Dempster (1969) for a proof. The characteristic function of $\mathbf{X}$, derived by Lord (1954), is of the form $e^{i\mathbf{t}^T\boldsymbol{\theta}}h(\mathbf{t}^T\mathbf{t})$, where $h(\cdot)$ is some suitable real-valued function and $\mathbf{t} \in \mathbb{R}^d$. Moreover the density $f(\mathbf{x})$ of $\mathbf{X}$, if exists, is then a function of $(\mathbf{x} - \boldsymbol{\theta})^T(\mathbf{x} - \boldsymbol{\theta})$.

A generalisation to the concept of spherical symmetry is that of elliptical symmetry. A random vector $\mathbf{X}$ defined on $\mathbb{R}^d$ with distribution function F is said to have an *elliptically symmetric distribution* about $\boldsymbol{\theta} \in \mathbb{R}^d$ if there exists $d \times d$ matrix $\mathbf{A}$ such that $\mathbf{A}(\mathbf{X}-\boldsymbol{\theta})$ has a spherically symmetric distribution about $\mathbf{0}$. In other words, an affine transformation of a spherically symmetric random variable generates an elliptically symmetric random variable. One may note that the family of elliptically symmetric distributions is closed under affine transformations.

From our discussion of the spherical symmetry and the definition of the elliptically symmetry, it follows that if $\mathbf{X}$ is elliptically symmetric about $\boldsymbol{\theta}$ with scale matrix Σ, then

its characteristic function takes the form $e^{i\mathbf{t}^T\boldsymbol{\theta}}h(\mathbf{t}^T\Sigma\mathbf{t})$, where $h(\cdot)$ is some suitable real-valued function and $\mathbf{t} \in \mathbb{R}^d$. Moreover the density $f(\mathbf{x})$ of $\mathbf{X}$, if it exists, is a function of $(\mathbf{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})$. An alternative characterisation in terms of the matrix $\mathbf{A}$ was considered by Beran (1979), who also developed one of the earliest tests of elliptical symmetry. A detailed discussion of the properties of the elliptically symmetric distributions can be found in Fang, Kotz and Ng (1990).

We will again come back to spherical and elliptically symmetric distributions in Chapter 2 where we will discuss some characterisations of the rank functions for these distributions. Subsequently we will develop a test of elliptical symmetry in Chapter 6.

Perhaps the most immediate extension of the concept of univariate symmetry to the multivariate regime for a random vector $\mathbf{X}$ defined on $\mathbb{R}^d$ is to be *centrally symmetric* about $\boldsymbol{\theta} \in \mathbb{R}^d$, i.e. to have $(\mathbf{X}-\boldsymbol{\theta})$ and $(\boldsymbol{\theta}-\mathbf{X})$ to have the same distribution. This concept of symmetry is more general to either of the situations we described above, for example, the uniform distribution on the d-dimensional hypercube $[-1,1]^d$ is centrally symmetric but not spherically or elliptically symmetric. Zuo and Serfling (2000) established that this symmetry may alternatively be characterised by the requirement that $\mathbf{u}^T(\mathbf{X}-\boldsymbol{\theta})$ and $\mathbf{u}^T(\boldsymbol{\theta}-\mathbf{X})$ have the same univariate distribution for every unit vector $\mathbf{u} \in \mathbb{R}^d$.

Liu (1990) introduced a more general idea of symmetry: a random vector $\mathbf{X} \in \mathbb{R}^d$ is *angularly symmetric* about $\boldsymbol{\theta} \in \mathbb{R}^d$ if $\frac{\mathbf{X}-\boldsymbol{\theta}}{||\mathbf{X}-\boldsymbol{\theta}||}$ and $\frac{\boldsymbol{\theta}-\mathbf{X}}{||\boldsymbol{\theta}-\mathbf{X}||}$ have the same distribution, or equivalently, if $\frac{\mathbf{X}-\boldsymbol{\theta}}{||\mathbf{X}-\boldsymbol{\theta}||}$ is centrally symmetric. Liu (1990) noted that the point of angular symmetry $\boldsymbol{\theta}$ of $\mathbf{X}$, if exists, is unique as long as the distribution of $\mathbf{X}$ is not concentrated on a line in $\mathbb{R}^d$ with more than one median. For some more characterisations of this symmetry, see Serfling (2006).

Clearly, the four concepts of symmetry we have discussed so far increasingly generalise the concept of multivariate symmetry in the sense that spherical symmetry implies elliptical symmetry, which in turn implies central symmetry, and further, central symmetry implies angular symmetry. It is interesting to note that all of the above concepts reduce

to the same notion of univariate symmetry defined before (Serfling (2006)).

Some other generalisations of the concept of central symmetry exist: $\mathbf{X} \in \mathbb{R}^d$ is defined to be *sign-symmetric* about $\boldsymbol{\theta} \in \mathbb{R}^d$ if $(\mathbf{X} - \boldsymbol{\theta}) = (X_1 - \theta_1, X_2 - \theta_2, \cdots, X_d - \theta_d)$ have the same distribution as $(\pm(X_1 - \theta_1), \pm(X_2 - \theta_2), \cdots, \pm(X_d - \theta_d))$ for all 2^d choices of $+/-$ combinations. Blough (1989) defined $d \times d$ matrices $\mathbf{B}_k$ by

$$\mathbf{B}_k = ((b_{ij}))_k = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \neq k \\ -1 \text{ if } i = j = k, \end{cases}$$

and correspondingly defined $\mathbf{X} \in \mathbb{R}^d$ to be *symmetric of degree* m if there exists a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m, 0, \cdots, 0)^T$ in $\mathbb{R}^d$ and an orthogonal transformation $\mathbf{T}$ such that $\mathbf{T}(\mathbf{X} - \boldsymbol{\theta})$ have the same distribution as $\mathbf{B}_1\mathbf{B}_2 \cdots \mathbf{B}_m\mathbf{T}(\mathbf{X} - \boldsymbol{\theta})$. Symmetry of degree d is therefore the regular central symmetry. In general symmetry of order m means symmetry about m mutually orthogonal $(d-1)$- dimensional hyperplanes.

Beran and Millar (1997) looked at the various ideas of symmetry from a characterisation concept in terms of probabilities of half-spaces. For $\mathcal{A}_0$, a compact subgroup of all orthogonal transformations on $\mathbb{R}^d$, define a random vector $\mathbf{X} \in \mathbb{R}^d$ to be $\mathcal{A}_0$-*symmetric* about $\boldsymbol{\theta} \in \mathbb{R}^d$ if the distribution of $\mathbf{X} - \boldsymbol{\theta}$ remains unchanged under any orthogonal transformation from $\mathcal{A}_0$. Clearly, when $\mathcal{A}_0$ is the space of all orthogonal transformations, this symmetry reduces to the spherical symmetry, the central symmetry corresponds to $\mathcal{A}_0$ consisting of the identity transformation and its negative, and sign- symmetry is achieved when $\mathcal{A}_0$ is the group of all the d-dimensional transformations of the type $(\pm 1, \pm 1, \cdots, \pm 1)^T$. Hence one may think that the sign-symmetry is somewhere in between spherical and central symmetry in terms of generality.

Cambanis, Keener and Simons (1983) attempted a different generalisation of the idea of spherical symmetry. They define a random vector $\mathbf{X} \in \mathbb{R}^d$ to be α-*symmetric* about $\boldsymbol{\theta} =$

$(\theta_1, \theta_2, \cdots, \theta_d)^T$, if the characteristic function of $\mathbf{X}$ is of the form $e^{i\mathbf{t}^T\boldsymbol{\theta}}h(|t_1|^\alpha, |t_2|^\alpha, \cdots, |t_d|^\alpha)$ for some $\alpha > 0$. The spherical symmetry is obtained as a special case for $\alpha = 2$. For $d = 1$, this idea reduces to the usual symmetry regardless of the value of α.

Finally, we discuss a concept of symmetry, due to Zuo and Serfling (2000), based on the idea of a symmetry over *halfspaces*. The halfspaces over $\mathbb{R}^d$ are defined by

$$H(\mathbf{s}, t) = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{s}^T\mathbf{x} \leq t\}, \;\; \mathbf{s} \in \mathbf{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : ||\mathbf{u}|| = 1\}, t \in \mathbb{R}.$$

A random vector $\mathbf{X} \in \mathbb{R}^d$ is then *halfspace symmetric* about $\boldsymbol{\theta} \in \mathbb{R}^d$ if

$$P(\mathbf{X} \in H) \geq \frac{1}{2},$$

for every closed halfspace H with $\boldsymbol{\theta}$ on the boundary. Among the notable characteristics of halfspace symmetry, one may note that the angular symmetry about a point implies the halfspace symmetry about that point, which means that this is the most general form of symmetry that we discuss. Furthermore, similar to the case of angular symmetry, it may be shown that the point of halfspace symmetry $\boldsymbol{\theta}$ of $\mathbf{X}$, if exists, is unique as long as the distribution of $\mathbf{X}$ is not concentrated on a line in $\mathbb{R}^d$ with more than one median. For proofs of these results and further discussions, one may refer to Zuo and Serfling (2000).

With the various notions of symmetry in hand, one may wish to see a few examples. In the following section we introduce some popular multivariate distributions, all of which are elliptically symmetric and will be used throughout the thesis.

1.2 Some Multivariate Distributions

Multivariate Normal Distribution : Suppose $\mathbf{X}$ is a d-dimensional random vector with multivariate normal distribution $N_d(\boldsymbol{\theta}, \Sigma)$ and $f(\mathbf{x})$ is the density function of $\mathbf{X}$.

Then

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma^{\frac{1}{2}}|} exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})\right).$$

Here $\boldsymbol{\theta}$ is the location vector and Σ is the covariance matrix of $\mathbf{X}$. We say that $\mathbf{X}$ is a standard d-variate normal distribution if $\boldsymbol{\theta} = \mathbf{0}$ and $\Sigma = \mathbf{I}_d$.

It may be noted that the standard multivariate normal distribution is symmetric in every sense described in Section 1.1, whereas a general multivariate normal distribution is not necessarily spherically symmetric, sign-symmetric or $\mathcal{A}_0$ symmetric but satisfies the conditions of the other concepts of symmetry defined above.

Multivariate Laplace Distribution : Suppose $\mathbf{X}$ is a d-variate random variable with multivariate Laplace distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Then the density $f(\mathbf{x})$ of $\mathbf{X}$ is given by

$$f(\mathbf{x}) \propto \frac{1}{|\Sigma|^{\frac{1}{2}}} exp\left(-\sqrt{(\mathbf{x}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\theta})}\right).$$

By standard d-variate Laplace distribution we denote the multivariate Laplace distribution which corresponds to $\boldsymbol{\theta} = \mathbf{0}$ and $\Sigma = \mathbf{I}_d$.

Like the case of the multivariate normal distribution, the standard multivariate Laplace distribution is symmetric in every sense described in Section 1.1. Similarly, a general multivariate Laplace distribution is not necessarily spherically symmetric, sign-symmetric or $\mathcal{A}_0$ symmetric but satisfies the conditions of the other concepts of symmetry defined above.

Multivariate t Distribution: The density of the standard d-variate t-distribution with k degrees of freedom is obtained from $\frac{\mathbf{X}}{\sqrt{\mathbf{Z}/k}}$ where $\mathbf{X}$ has multivariate normal distribution $N_d(\mathbf{0}, \mathbf{I})$, $\mathbf{Z}$ has χ^2_k distribution and $\mathbf{X}$ and $\mathbf{Z}$ are independent of each other. If $\mathbf{Y}$ is a d-variate t distribution with k degrees of freedom location vector $\boldsymbol{\theta}$ and scale matrix Σ, the density $f(\mathbf{y})$ of $\mathbf{Y}$ is given by

$$f(\mathbf{y}) = \frac{\Gamma\left(\frac{k+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) k^{\frac{d}{2}}\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}[1+\frac{1}{d}(\mathbf{y}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{y}-\boldsymbol{\theta})]^{\frac{k+d}{2}}}.$$

The multivariate t distribution is another example of a distribution which is symmetric in every sense described in Section 1.1.

1.3 Notions of Data Depth

The idea of data depth came from a location theory problem considered by Weber (1909). Here one tries to locate a point $\mathbf{x} \in \mathbb{R}^d$ such that it has the minimum distance from all the points in some sense. To solve this problem Gini and Galvani (1929) introduced a concept of spatial median. Let us assume that $\mathbf{X}$ is a random vector in $\mathbb{R}^d$ with absolutely continuous probability distribution F. The L_2 depth and the L_2 depth median or the spatial median is defined as follows.

Definition 1.3.1 *The L_2 depth at $\mathbf{x}$ with respect to F is,*

$$L_2D(F;\mathbf{x}) = (1 + E(||\mathbf{X} - \mathbf{x}||_2))^{-1} \tag{1.1}$$

The corresponding L_2-depth median is defined as,

$$Med_{L_2D} = \arg\max_{\mathbf{x}\in\mathbb{R}^d} L_2D(F;\mathbf{x}). \tag{1.2}$$

The geometric idea behind this median is to minimise the sum of Euclidean distances to the observations. Brown (1983) introduced the term *spatial median* for the L_2 median and showed that its empirical distribution is asymptotically normal. Spatial median is equivariant under orthogonal transformations but it is not equivariant under general non-singular transformation of the data (see Chaudhuri (1992)).The spatial median has a 50% breakdown point. Brown (1983) studied the properties of the spatial median extensively and introduced a sign test analogue, the angle test, based on the angles between the spatial median and the observations.

Other than the L_2 depth, the Mahalanobis depth function (see Mahalanobis (1936)) is one of the oldest and well known depth function. The Mahalanobis depth function is

defined as follows:

Definition 1.3.2 *The Mahalanobis depth at* $\mathbf{x}$ *with respect to* F *is,*

$$MD(F;\mathbf{x}) = [1 + (\mathbf{x} - \boldsymbol{\mu}_F)^T \Sigma_F^{-1} (\mathbf{x} - \boldsymbol{\mu}_F)]^{-1} \tag{1.3}$$

where $\boldsymbol{\mu}_F$ *and* Σ_F *are the mean vector and dispersion matrix of* F *respectively.*

The median associated with the Mahalanobis depth is obtained by minimising the standardised Euclidean norm. As Mahalanobis depth is affine invariant, this median is affine equivariant. This median was also mentioned by Haldane (1948) and is sometimes referred to as Haldane's multivariate median (Isogai (1985)). Isogai (1985) obtained various measures of multivariate skewness based on this median.

Tukey (1975) suggested the following notion of data depth, the half space depth:

Definition 1.3.3 *The Half-Space depth at* $\mathbf{x}$ *with respect to* F

$$HD(F;\mathbf{x}) = \inf_H \left\{ P(H) : H \textit{ is a closed half-space in } \mathbb{R}^d \textit{ and } \mathbf{x} \in H \right\}. \tag{1.4}$$

Based on this depth the median can be defined as

$$Med_{HD} = \arg \max_{\mathbf{x} \in \mathbb{R}^d} HD(F;\mathbf{x}). \tag{1.5}$$

The sample version of the half-space depth for the one dimensional case can be described in the following way: consider a one dimensional data set, $X = \{X_1, X_2, \cdots, X_n\}$, the depth of θ would be the minimum number of data points on the either side of θ. In numerical terms, this can be expressed as

$$D(\theta) = \min(\#\{i : X_i \leq \theta\}, \#\{i : X_i \geq \theta\}) \tag{1.6}$$

In d-dimension, to find the depth of the point $\boldsymbol{\theta}$ with respect to the multivariate data set $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ one would draw a plane through the point $\boldsymbol{\theta}$. Then rotate this plane

and keep track of the smallest number of points on one side of the plane. Then depth of $\boldsymbol{\theta}$, $D(\boldsymbol{\theta})$ = Smallest proportion of points in a closed half-space containing $\boldsymbol{\theta}$.

For $d = 2$ Rousseeuw and Ruts (1996) gave an exact algorithm to compute the depth contours of a bivariate data set with respect to the half space depth in $O(n \log n)$ time. Later Rousseeuw and Struyf (1998) obtained an algorithm to compute the depth contours with respect to the half-space depth in three dimensions in $O(n^2 \log n)$ time. Johnson and Preparata (1978) provided an algorithm of order $O(n^{d-1} \log n)$ for the computation of the half-space depth of a given point, but they did not consider it practical for $d > 4$. Struyf and Rousseeuw (2000) gave an algorithm named DEEPLOC, which can approximate the maximum depth in higher dimensions. DEEPLOC has a time complexity of $O(kmn \log n + kdn + md^3 + mdn)$, where k is the number of steps taken by the program and m is the number of directions.

The half-space median, as defined in (1.5), is affine equivariant. The breakdown point for the sample version of this median is nearly $\frac{1}{3}$ and this property has been studied extensively by Donoho (1982), Donoho and Huber (1983), Donoho and Gasko (1987) and Chen (1995). Bai and He (1999) showed that it is also asymptotically normal. This median may not be a unique point. In two dimensions this median can be computed in $O(n^2 log^2 n)$ time by an algorithm HALFMED suggested by Rousseeuw and Ruts (1998), but computation gets increasingly difficult as the dimension increases. Rousseeuw and Struyf (1998) discusses a fast algorithm to obtain the spatial median in higher dimensions.

It may be noted that the Hodges sign test statistic, to be defined formally later in Section 1.5, is equivalent to the half-space depth of the origin.

Liu (1988) suggested another notion of data depth named simplicial depth using random simplices.

Definition 1.3.4 *The simplicial depth at* $\mathbf{x}$ *with respect to* F *is defined as,*

$$SD(F; \mathbf{x}) = P_F \{\mathbf{x} \in S[X_1, \cdots, X_{d+1}]\} \tag{1.7}$$

where $S[X_1, \cdots, X_{d+1}]$ *is a closed simplex formed by* $(d+1)$ *random observations from* F*, as* $(d+1)$ *vertices.*

The median corresponding to this depth is defined as,

$$Med_{SD} = \arg \max_{\mathbf{x} \in \mathbb{R}^d} SD(F; \mathbf{x}) \tag{1.8}$$

In one dimension, this depth function gives the proportion of intervals with two observations as end points containing $x \in \mathbb{R}$, *i.e.*, $SD(F; x) = P(x \in \overline{X_1 X_2})$, where $\overline{X_1 X_2}$ is the line segment joining observations X_1 and X_2. In dimension $d > 1$, this depth gives the probability of the point $\mathbf{x}$ is inside the simplices formed by $(d+1)$ random observations from F. The simplicial depth median is equivariant under general affine transformations. For two dimensions Rousseeuw and Ruts (1996) algorithm converges in $O(n \log n)$ time for simplical depth as well. This attains the lower bound, see Aloupis, Cortes, Gomez, Soss and Toussaint (2002). Cheng and Ouyang (2001) suggested an algorithm to compute the simplical depth with a time complexity of $O(n^4)$ for $d = 4$. Rousseeuw and Ruts proposed an $O(n^3)$ algorithm for $d = 3$. For dimensions more than 4, there is no known efficient algorithm.

Liu (1988, 1990) showed that the sample version of this median is consistent. Arcones, Chen and Gine (1994) showed that it is asymptotically normal using U-statistics. For convergence properties one can see Liu (1990) and D*ü*mbgen (1992). One problem with this median is again its computation. Aloupis, Langerman, Soss and Toussaint (2003) provides an algorithm to compute this median in dimension $d = 2$ in $O(n^4)$ time. No known algorithm exists for higher dimensions.

Oja (1983) also defined a depth function based on simplices. This depth is related with the volume of the simplices in d-dimension. In the univariate case, the volume would be the distance between two points in $\mathbb{R}$. The formal definition for the d-dimensional case follows:

Definition 1.3.5 *The Oja depth at* $\boldsymbol{x}$ *with respect to* F *is defined as,*

$$OD(F; \boldsymbol{x}) = [1 + E_F \{volume\ (S[\mathbf{x}, X_1, \cdots, X_d])\}]^{-1} \tag{1.9}$$

where $S[\mathbf{x}, X_1, \cdots, X_d]$ *is the closed simplex with vertices* $\mathbf{x}$ *and* d *random observations, say* $X_1, \cdots, X_d$*, from* F*.*

The median corresponding to this depth, known as Oja's median is defined as,

$$Med_{OD} = \arg \max_{\mathbf{x} \in \mathbb{R}^d} OD(F; \mathbf{x}). \tag{1.10}$$

This median is also equivariant under general affine transformations. Oja and Niinimaa (1985) showed that the sample version of this median is asymptotically as efficient as the spatial median when the underlying distribution is multivariate normal. Niinima, Oja and Tableman (1990) showed that this median has 0% breakdown. Arcones, Chen and Gine (1994) showed that the sample version of this median is asymptotically normal using U-statistic type representation. Ronkainen, Oja and Orponen (2003) obtained an exact algorithm to compute the Oja median in d dimensions in $O(dn^d \log n)$ time.

There are many other defintions of depths, such as, convex hull peeling depth by Barnett (1976), majority depth by Singh (1991), projection depth (see Donoho and Gasko (1992), Liu (1992)), likelihood depth by Fraiman and Meloche (1996), zonoid depth by Koshevoy and Mosler (1997). The value of depth is usually different for different definitions but all of them basically indicate that a larger value of depth of a point x implies a deeper or a more central $\mathbf{x}$ with respect to F. Every notion of depth gives rise to a partial ordering of multivariate data. This ordering helps one to rank the sample points from the centre to the outward way. It also gives us the nested contours depending upon depth and these depth contours form a sequence of nested convex sets.

There are a few desirable properties for the depth functions, which are listed below following Zuo and Serfling (2000).

1. **Affine invariance:** $D(F; \mathbf{x})$ is independent of the underlying coordinate system

i.e. for any random vector $\mathbf{X} \in \mathbb{R}^d$, for any non singular $d \times d$ matrix $\mathbf{A}$ and any d-dimensional vector $\mathbf{b}$,

$$D(F_{\mathbf{AX}+\mathbf{b}}; \mathbf{Ax} + \mathbf{b}) = D(F; \mathbf{x}) \tag{1.11}$$

2. **Maximality at centre:** As the depth function gives us a centre outward ordering, so we expect that the maximum depth would be obtained at the centre of symmetry *i.e.,* if F is symmetric about a point $\boldsymbol{\theta}$ in some sense, then $D(F; \mathbf{x})$ is maximal at $\boldsymbol{\theta}$.

3. **Decreasing along rays:** This property can also be described as the property of monotonicity relative to the deepest point. As we expect to have the greatest depth at the centre of symmetry, if any point $\mathbf{x} \in \mathbb{R}^d$ moves away from the deepest point along any fixed ray through the centre of symmetry, then the depth function decreases monotonically *i.e.,* the depth $D(F; \mathbf{x})$ decreases along each ray from the deepest point.

4. **Vanishing at infinity:** When a point $\mathbf{x} \in \mathbb{R}^d$ moves further away from the centre of symmetry, which is the deepest point, the depth of the point decreases and it approaches zero as the point $\mathbf{x} \in \mathbb{R}^d$ moves to infinity, *i.e.,* $D(F; \mathbf{x}) \to 0$, as $||\mathbf{x}|| \to \infty$.

5. **Symmetry:** When the distribution is symmetric about a point, then the greatest depth is achieved there and the depth function should decrease in the same way along any ray which gives a symmetric depth function *i.e.,* if F is symmetric about a point $\boldsymbol{\theta}$ in some sense, then $D(F; \mathbf{x})$ is also symmetric.

6. **Continuity with respect to x and continuity with respect to F:** Given a fixed continuous multivariate distribution function F, $D(F; \mathbf{x})$ is continuous as a function of $\mathbf{x}$ and given a fixed $\mathbf{x}$, $D(F; \mathbf{x})$ is continuous as a function of F.

7. **Quasi-concavity as a function of x:** The depth contours form sequence of nested convex set *i.e.,* the sets $\{\mathbf{x} : D(F;\mathbf{x}) \geq c\}$ is convex for each real c and $\{\mathbf{x} : D(F;\mathbf{x}) \geq c_1\} \subseteq \{\mathbf{x} : D(F;\mathbf{x}) \geq c_2\}$ if $c_2 \leq c_1$.

The properties $1, 2, 3, 4$ are the most important ones as they help us to define a statistical depth function. Almost all of the above mentioned depth based medians are affine equivariant with the exception of the L_2 depth median. It is desirable that a depth function is affine equivariant or at least a modified version should be affine equivariant. For example, Rao (1988) suggested a modification to the L_2D in the following way

$$\widetilde{L_2D}(F;\mathbf{x}) = (1 + E(||\mathbf{X} - \mathbf{x}||_{\Sigma^{-1}}))^{-1} \tag{1.12}$$

where Σ is the covariance matrix of F and $||\mathbf{X} - \mathbf{x}||_{\Sigma^{-1}} = \sqrt{(\mathbf{X} - \mathbf{x})^T \Sigma^{-1} (\mathbf{X} - \mathbf{x})}$. This modification makes L_2D affine invariant. Liu (1990) showed that the simplicial depth $SD(F;\mathbf{x})$ defined in (1.7) satisfies properties $1, 2, 3$ and 4. Zuo and Serfling (2000) discussed how these four properties are satisfied by various depth functions.

From the point of view of satisfying the above four properties, the half space depth seems to be the most elligible depth.

1.4 Multivariate Quantiles

In the univariate setup, quantiles characterise the underlying distribution. Quantiles are used to construct quantile-quantile (Q-Q) plots to visually compare the shape of the distribution with some hypothesised distribution. Let Z be a univariate random variable such that $E|Z| < \infty$. Then the p-th quantile $Q(p)$ of Z for $0 < p < 1$ is defined as

$$Q(p) = \inf \{z : P(Z \leq z) \geq p\},$$

which can also be characterised as

$$\arg \min_{\theta \in \mathbb{R}} E\left\{|Z-\theta| + (2p-1)(Z-\theta)\right\} \text{ (Ferguson (1967)).}$$

When the data are multivariate, it is quite natural for the quantiles to have both magnitude and direction as suggested by Brown and Hettmansperger (1987). Babu and Rao (1988) defined the vector of quantiles of the marginal distributions as the multivariate quantile vector.

Even though this vector of quantiles is quite easy to compute, as marginal distributions do not characterise the joint distribution, this definition of multivariate quantiles can not characterise a general multivariate distribution.

Chaudhuri (1996) extended the definition of the univariate quantiles to higher dimensions by defining the $\mathbf{u}$-th quantile $\mathbf{Q}(\mathbf{u})$ of $\mathbf{X} \in \mathbb{R}^d$ as any minimiser of the function

$$E\left\{\Phi(\mathbf{u}, \mathbf{X}-\boldsymbol{\theta}) - \Phi(\mathbf{u}, \mathbf{X})\right\}$$

where $||\mathbf{u}|| < 1$ and $\Phi(\mathbf{u}, \mathbf{t}) = ||\mathbf{t}|| + \mathbf{u}^T\mathbf{t}$, and $||.||$ is the usual Euclidean norm. This definition of multivariate quantiles was popularised as *geometric quantiles* or *spatial quantiles*. Koltchinskii (1997) showed that these quantiles characterise the multivariate distribution. Further, the geometric quantiles are also equivariant under rotations of the data cloud. However, they are not equivariant under general affine transformations.

Suppose $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ are random variables in $\mathbb{R}^d$ and the sample version of the geometric quantile, $\mathbf{Q}_n(\mathbf{u})$ is defined as the minimiser of

$$\sum_{i=1}^{n} \left\{||\mathbf{X}_i - \boldsymbol{\theta}|| + \mathbf{u}^T(\mathbf{X}_i - \boldsymbol{\theta})\right\}.$$

A few properties of the sample version of the geometric quantile are as follows.

1. The sample version of the geometric quantile $\mathbf{Q}_n(\mathbf{u})$ would always exist for any given $\mathbf{u} \in \mathbb{B}^d$ where $\mathbb{B}^d$ is the open unit ball in d- dimension and it would be unique for

$d \geq 2$ unless $\mathbf{X}_i$'s are all along a straight line in $\mathbb{R}^d$.

2. Suppose $\mathbf{A}$ is a $d \times d$ orthogonal matrix and $\mathbf{b}$ is a fixed d-dimensional vector. Let $\mathbf{Y}_i = \mathbf{A}\mathbf{X}_i + \mathbf{b}$ and set $\mathbf{v} = \mathbf{A}\mathbf{u}$. If $\mathbf{Q}_n^{\mathbf{Y}}(\mathbf{v})$ is the sample version of the geometric quantile corresponding to $\mathbf{v}$ and depending on $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$, then

$$\mathbf{Q}_n^{\mathbf{Y}}(\mathbf{v}) = Q_n^{\mathbf{Y}}(\mathbf{A}\mathbf{u}) = \mathbf{A}Q_n(\mathbf{u}) + \mathbf{b} \tag{1.13}$$

3. Suppose $c > 0$ is a fixed scalar and let $\mathbf{Y}_i = c\mathbf{X}_i$. Then the sample version of the geometric quantile based on $\mathbf{Y}_i$'s would be $c\mathbf{Q}_n(\mathbf{u})$.

Chaudhuri (1996) proposed an algorithm based on Newton-Raphson method for the computation of such quantiles. The geometric quantiles constructed in this way can be useful to extend quantile regression from univariate to multivariate model (see Chakraborty (2003)) and also to construct L estimates of univariate location. Chaudhuri (1996) showed that the joint asymptotic distribution of the centred and normalised geometric quantiles would be Gaussian with mean zero. These geometric quantiles are consistent estimates of the corresponding population quantiles and their rate of convergence is $n^{-\frac{1}{2}}$. In the univariate case, quantiles can be thought of as inverse of the cumulative density functions. Similarly in the multivariate case, the geometric quantile $\mathbf{Q}(\mathbf{u})$ can be obtained by inverting the equation

$$E\left(\frac{\mathbf{Q}(\mathbf{u}) - \mathbf{X}}{||\mathbf{Q}(\mathbf{u}) - \mathbf{X}||}\right) = \mathbf{u}.$$

Geometric quantiles also help us to find out how close a data point is to the centre of the multivariate data set.

Chaudhuri (1996) also remarked that geometric quantiles can be used to construct quantile balls $\{\mathbf{Q}_n(\mathbf{u}) : |\mathbf{u}| < r\}$ with radius $r \in (0, 1)$, which can be thought of multivariate analog of univariate quantile range. If we take $\lambda(r)$ as a d-dimensional Lebesgue measure of this set, then for a suitable r, $t \in (0, 1)$ with $r < t$, $\frac{\lambda(r)}{\lambda(t)}$ can be considered

as a measure of kurtosis in the multivariate case. In an alternative approach Avérous and Meste (1994, 1997) developed an affine invariant kurtosis function for the multivariate case using a similar concept of median balls obtained minimising a distance function based on arbitrary convex norm functions.

1.5 Sign and Rank Based Methods

In this section we discuss some sign and rank based methods that could be used to get some insight about various location parameters of the distribution.

A univariate sign test can be considered as a test for the median θ of the corresponding probability distribution generating the data. Here one will test $H_0 : \theta = \theta_0$ against, for example, $H_1 : \theta \neq \theta_0$. The test statistic would be $\sum_{i=1}^{n} I(X_i - \theta_0 \leq 0)$, which will have a binomial distribution with parameters n and $\frac{1}{2}$ under H_0. Hence under H_0 the distribution of the test statistic does not depend on F. For details we refer to Oja (1999). Alternatively, based on the Sign function defined as follows

$$\text{Sign}(x) = \begin{cases} 1 \text{ if } x > 0, \\ 0 \text{ if } x = 0, \\ -1 \text{ if } x < 0, \end{cases}$$

one can obtain

$$\sum_{i=1}^{n} \text{Sign}(X_i - \theta) = \sum_{i=1}^{n} I(X_i - \theta_0 \leq 0) - S_n = n - 2S_n$$

where $S_n = \sum_{i=1}^{n} I(X_i - \theta_0 < 0)$. Hence S_n can also be used instead of $\sum_{i=1}^{n} \text{Sign}(X_i - \theta)$ to test for H_0. As this test is simple and distribution free in the univariate case, a multivariate version of this test would be quite helpful. There are several ways of defining the median of a multivariate distribution thus there are many possibilities for multivariate sign tests.

Now suppose $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ are i.i.d. in $\mathbb{R}^d$ with a common absolutely continuous distribution function F with coordinate-wise median θ. To perform a sign test in the multivariate median case we need to specify the Sign function first. One easy choice would be $\text{Sign}(\mathbf{X}_i) = (\text{Sign}(X_{i1}), \text{Sign}(X_{i2}), \cdots, \text{Sign}(X_{id}))^T$ where X_{ij} is the j-th component of the random vector $\mathbf{X}_i$. The test based on $\text{Sign}(\mathbf{X}_1), \text{Sign}(\mathbf{X}_2), \cdots, \text{Sign}(\mathbf{X}_n)$ is a test for vector of the medians of its real valued components.

Several modifications of the classical sign test has been suggested in literature. Two of the most famous attempts were by Hodges (1955) and Blumen (1958). Hodges (1955) suggested a bivariate sign-test statistic

$$T_H = \sup_{\lambda \in \mathbb{R}^2 : ||\lambda|| = 1} \left| \sum_{i=1}^{n} \text{Sign}(\lambda^T \mathbf{X}_i) \right| \tag{1.14}$$

where $\lambda^T \mathbf{X}_i$ is the projection of $\mathbf{X}_i$ onto the 2-dimensional hyperplane which passes through the origin and λ is the normal vector. This test statistic T_H was used for bivariate sign test and based on T_H, H_0 was rejected for large values of T_H. This test is "association invariant"(see Puri and Sen (1971)) and distribution free, and has an obvious generalisation to higher dimensions (Chaudhury and Sengupta (1993)). Blumen (1958) computed a bivariate sign test statistic for a bivariate dataset (x_i, y_i) for $i = 1, \cdots, n$

$$v^2 = \frac{2}{n} \left(\left(\sum_{i=1}^{n} \text{Sign}(y_i/x_i) \cos \frac{\pi j}{n} \right)^2 + \left(\sum_{i=1}^{n} \text{Sign}(y_i/x_i) \sin \frac{\pi j}{n} \right)^2 \right), \tag{1.15}$$

rejecting for large values of v^2. Like Hodges' test, this test is also affine-invariant and distribution free. Blumen's test belongs to a class of strictly distribution-free locally most powerful affine invariant sign tests defined by Oja and Nyblom (1989). This test was generalised by Randles (1989), based on interdirections, to a multivariate sign test.

Among the other notable attempts at multivariate sign tests, Bennett (1962) suggested a multivariate sign test and later Chatterjee (1966) proposed a bivariate sign test which essentially were combinations of component sign tests. Bickel (1965) obtained a

generalisation of the Wilcoxon signed rank test.

More recently Oja and Nyblom (1989) proposed a bivariate sign test based on the angle created by the observations at a centre of location. Brown and Hettmansperger (1989) and Brown, Hettmansperger, Nyblom and Oja (1992) investigated a bivariate sign test based on the gradient of the *Oja objective function*

$$T(\theta) = \sum_{i<j} A(x_i, x_j, \theta)$$

where $A(a, b, c)$ is the area of a triangle whose vertices are a, b and c. All these recent tests are affine invariant. Chaudhuri and Sengupta (1993) generalised Randles' test to a broad class of affine invariant multivariate sign tests by transformation-retransformation method of which Hodges' test is also an example.

Based on the sign function, one can define the univariate *centred rank function* as follows:

$$R_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \text{Sign}(\mathbf{x} - \mathbf{X}_i). \tag{1.16}$$

This statistic could be used as an estimator of location. It has both magnitude (robust distances from the median) and direction (sign with respect to the median) information, and if $R(x) = 2F(x) - 1$, then $\sup |R_n(x) - R(x)|$ goes to 0 in probability, see Oja (1999). Let us discuss an example where this can be used. Suppose F is a continuous univariate symmetric distribution with density function $f(x - \theta)$ where θ as the unknown centre of symmetry. We wish to estimate θ and test the null hypothesis $H_0 : \theta = 0$. The univariate sample median $\hat{\theta}$ is the solution of the implicit equation $R_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \text{Sign}(\hat{\theta} - X_i) = 0$. The sign test statistics for testing the null hypothesis $H_0 : \theta = 0$ is $R_n(0)$. For univariate median and sign test under general assumptions the limiting distribution of $n^{\frac{1}{2}}\hat{\theta}$ is univariate normal $\left(0, \frac{1}{(2f(0))^2}\right)$ (Oja (1999)).

Rank tests were originally developed to provide exact tests for nonparametric hypothesis. These tests could be thought of as a subfamily of permutation tests, though results

known for permutation tests are not very useful for rank tests. Rank test procedures are simple, fast, distribution free and can be applied even when only ranking data are available.

1.6 Current Work

In the last two decades, there has been an extensive development in the proposals for multivariate versions of quantiles and their uses in describing various aspects of the multivariate data. Chaudhuri (1996) proposed geometric quantiles, which are multivariate quantities indexed by vector a $\mathbf{u}$ in the d-dimensional unit ball, based on minimising Euclidean distances. Chakraborty (2001) generalised the definition of geometric quantiles for other l_p distances and made the quantiles equivariant under affine transformations using a transformation retransformation technique. He also suggested a bivariate version of the Q-Q plot using affine equivariant quantiles. However that definition of Q-Q plot cannot be extended to higher dimensions in any natural way. Marden (1998) also proposed a bivariate version of the Q-Q plots based on geometric quantiles. Serfling (2010) provides a detailed discussion on latest developments and several properties of the geometric quantiles and other definitions of multivariate quantiles primarily based on geometric quantiles.

In a different development Balanda and MacGillivray (1990) proposed spread-spread plots to compare univariate distributions where the spread functionals preserve the spread ordering of Bickel and Lehmann (1979). The spread-spread plots are quite useful in detecting the changes in the shape of the distributions at the peak or at the tails and it displays a growth pattern. Liu, Parelius and Singh (1999) gave another concept of data plot using the concept of data depth. They proposed depth-depth plots using which they discussed characteristics of a multivariate distribution, such as location, scale, skewness, kurtosis etc. Singh, Tyler, Zhang and Mukherjee (2009) defined quantile scale curves which are based on the volumes of regions with respect to the simplicial depth functions.

The quantile scale curves suggested by them have been utilised as a graphical test for finding out linear or nonlinear association between two groups of variables, measuring heavy tailedness of a distribution and testing for multivariate location and scale.

There are some more graphical tools available in the literature to detect non-normality for high dimensional data; see, for example, Liang, Pan and Yang (2004), Fang, Li, and Liang (1998), Liang and Ng (2009). However all of these methods heavily depend on different characterisations of multivariate normal distribution and cannot be extended to a large class of multivariate distributions.

In this thesis, recent developments in multivariate quantiles and ranks and the concept of univariate spread-spread plots are combined to propose graphical methods of comparing multivariate distributions. In Chapter 2, we define a notion of multivariate rank vector and propose a scale-scale plot motivated by the idea of Liu, Parelius and Singh (1999) and study its properties under the assumptions of elliptical symmetry. In Chapter 3, we discuss the asymptotic properties of the central rank region based volume functional. In Chapter 4, we discuss the uses of scale-scale plot as visual tests with examples in the one sample and two sample problems and propose some theoretical test procedures. In Chapter 5, we propose a test statistic based on the volume functional and perform a test of goodness of fit of multivariate normality and multivariate Laplace distribution. In Chapter 6, we discuss scale-scale plot for some non-elliptical distributions and a test for elliptical symmetry. These scale-scale plots are also quite useful in detecting the deviations in regard to peakedness or tail behaviour which we have proposed in Chapter 7. In Chapter 7 we also propose some visual tests of multivariate location and scale as some further applications of our scale-scale plots. Chapter 8 concludes and discusses some further extensions.

CHAPTER 2

SCALE CURVES BASED ON MULTIVARIATE RANKS

We have seen before the usefulness of sign and rank functions in construction of several described measures of the distribution and for tests of location. In this chapter we introduce a notion of multivariate rank vectors and use that to construct some measures of scale for the multivariate distributions.

2.1 Definitions and Some Properties

To begin with, we define a *multivariate sign function* followed by rank and rank regions.

Definition 2.1.1 *For* $\mathbf{x} \in \mathbb{R}^d$*, the* **multivariate sign function** *is defined as*

$$Sign(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{||\mathbf{x}||}, \text{ if } \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, \text{ if } \mathbf{x} = \mathbf{0}, \end{cases} \tag{2.1}$$

where $||\mathbf{x}|| = \sqrt{x_1^2 + \cdots + x_d^2}$, $\mathbf{x} = (x_1, \cdots, x_d)^T$.

The sign function $Sign(\mathbf{x})$ for $\mathbf{x} \neq 0$ gives the unit direction vector for $\mathbf{x}$.

Definition 2.1.2 *Suppose* $\mathbf{X} \in \mathbb{R}^d$ *is a random vector with distribution function* F*, which*

is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. *Then the* **multivariate spatial rank** *for a d-dimensional vector* $\mathbf{x}$ *with respect to* $\mathbf{X}$ *is defined as,*

$$R_F(\mathbf{x}) = E_F\left(\frac{\mathbf{x}-\mathbf{X}}{||\mathbf{x}-\mathbf{X}||}\right). \tag{2.2}$$

Note that the *spatial median*, $\boldsymbol{\theta}$, can be obtained by solving

$$R_F(\boldsymbol{\theta}) = 0. \tag{2.3}$$

We can also see that this rank function $R_F(\mathbf{x})$ is the inverse function of the multivariate geometric quantile function $Q(\mathbf{u})$ defined by Chaudhuri (1996) in the sense that $R_F(\mathbf{x}) = \mathbf{u}$ implies that $Q(\mathbf{u}) = \mathbf{x}$ and vice-versa. As the relation between data depth $D(.)$ and the multivariate spatial rank $R_F(.)$ would be given by $D(\mathbf{x}) = 1 - ||R_F(\mathbf{x})||$, a measure of outlyingness can be defined using $||R_F(\mathbf{x})||$. It can be verified that this outlyingness function is invariant under orthogonal and homogeneous scale transformations. For related discussion one can see Serfling (2004, 2006b). This rank orders the multivariate data in central outward way.

We now define formally the sample version of the rank function in the univariate set up.

Definition 2.1.3 *The (sample) centred rank of* x *with respect to a random sample* $X_1, X_2, \cdots, X_n \in \mathbb{R}$ *with distribution* F *on* $\mathbb{R}$ *is given by*

$$R_{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n} Sign(x - X_i). \tag{2.4}$$

This centred rank function has the following properties,

1. $-1 \leq R_{F_n}(x) \leq 1$.

2. $R_{F_n}(x) = 0$ implies that x is the median of $X_1, \cdots, X_n$.

3. $R_{F_n}(x) = -1$ implies $x \leq \min\{X_1, X_2, \cdots, X_n\}$ and $R_{F_n}(x) = +1$ implies $x \geq \max\{X_1, X_2, \cdots, X_n\}$.

4. $E(R_{F_n}(x)) = 2F(x) - 1$.

In a similar way we can define a multivariate version of rank using the definition of sign vector:

Definition 2.1.4 *Suppose* $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n \in \mathbb{R}^d$ *is a random sample with common distribution function* F *on* $\mathbb{R}^d$, $d \geq 1$. *The* ***centred rank*** *of* $\mathbf{x} \in \mathbb{R}^d$ *with respect to* $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ *is*

$$R_{F_n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} Sign(\mathbf{x} - \mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{x} - \mathbf{X}_i}{||\mathbf{x} - \mathbf{X}_i||}. \tag{2.5}$$

This multivariate rank has similar properties to the univariate rank function:

1. $||R_{F_n}(\mathbf{x})|| < 1$ for all $\mathbf{x} \in \mathbb{R}^d$.

2. $||R_{F_n}(\mathbf{x})|| = 0$ means that $\mathbf{x}$ is the spatial median of the data $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ in $\mathbb{R}^d$.

3. Smaller values of $||R_{F_n}(\mathbf{x})||$ imply that $\mathbf{x}$ is located more centrally with respect to the data points and larger values of $||R_{F_n}(\mathbf{x})||$ imply that $\mathbf{x}$ is an extreme point with respect to the data cloud. The direction of the vector $R_{F_n}(\mathbf{x})$ suggests the direction in which $\mathbf{x}$ is extreme compared to the data cloud.

4. $R_F(\mathbf{x}) = E(R_{F_n}(\mathbf{x}))$ is an injective function of the multivariate distribution function F, see Koltchinskii (1997). Hence $R_F(\mathbf{x})$ characterises a multivariate distribution.

Based on this notion of multivariate rank, we now define the *multivariate central rank region*.

Definition 2.1.5 *Suppose* $0 < p < 1$. *Let* $r_F(p)$ *be the* p*-th quantile of the distribution of* $||R_F(\mathbf{X})||$ *with* $\mathbf{X} \in \mathbb{R}^d$. *The* **central rank region** *denoted by* $C_F(p)$ *is defined as*

$$C_F(p) = \{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_F(p)\}\,. \tag{2.6}$$

The volume of the multivariate central rank region is given by

$$V_F(p) = \textit{volume of } C_F(p), 0 < p < 1. \tag{2.7}$$

Here $C_F(p)$ is the central region around the median, which has probability mass p, *i.e.* $P(C_F(p)) = p$, if F is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$. In the univariate case $C_F(p)$ is the central interval around the median, which has a probability mass p and $V_F(p)$ is the length of the interval. If the length of the interval is short, then the spread of the distribution is also small hence the measure of spread follows that satisfies properties in Bickel and Lehmann (1979). Thus when p varies from 0 to 1, one can study the changes of spread with the tails of the distribution. $V_F(p)$ gives a measure of change of scale in univariate case, it does a similar thing for the multivariate case.

Let us now go back to the sample version and denote the p-th quantile of $||R_{F_n}(\mathbf{X}_1)||, \cdots,$ $||R_{F_n}(\mathbf{X}_n)||$ by $r_{F_n}(p)$. Based on this we can define the sample central rank region:

Definition 2.1.6 *The sample central rank regions are defined as*

$$C_{F_n}(p) = \{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq r_{F_n}(p)\}\,, 0 < p < 1.$$

We denote the volume of $C_{F_n}(p)$ by V_{F_n}.

One can consider a plot of $V_F(p)$ against p, named scale curve, to visualise the measure of spread for varying p (see Liu, Parelius and Singh (1999) in the context of simplicial depth based on scale curves). Figure 2.1 gives us the plot of $V_F(p)$ against p *i.e.* scale plot for standard bivariate normal distribution, standard bivariate t distribution with 3 degrees of freedom and standard bivariate Laplace distribution. The definitions of these distributions are given in Section 1.2. For smaller values of p, the scales of these three distributions are

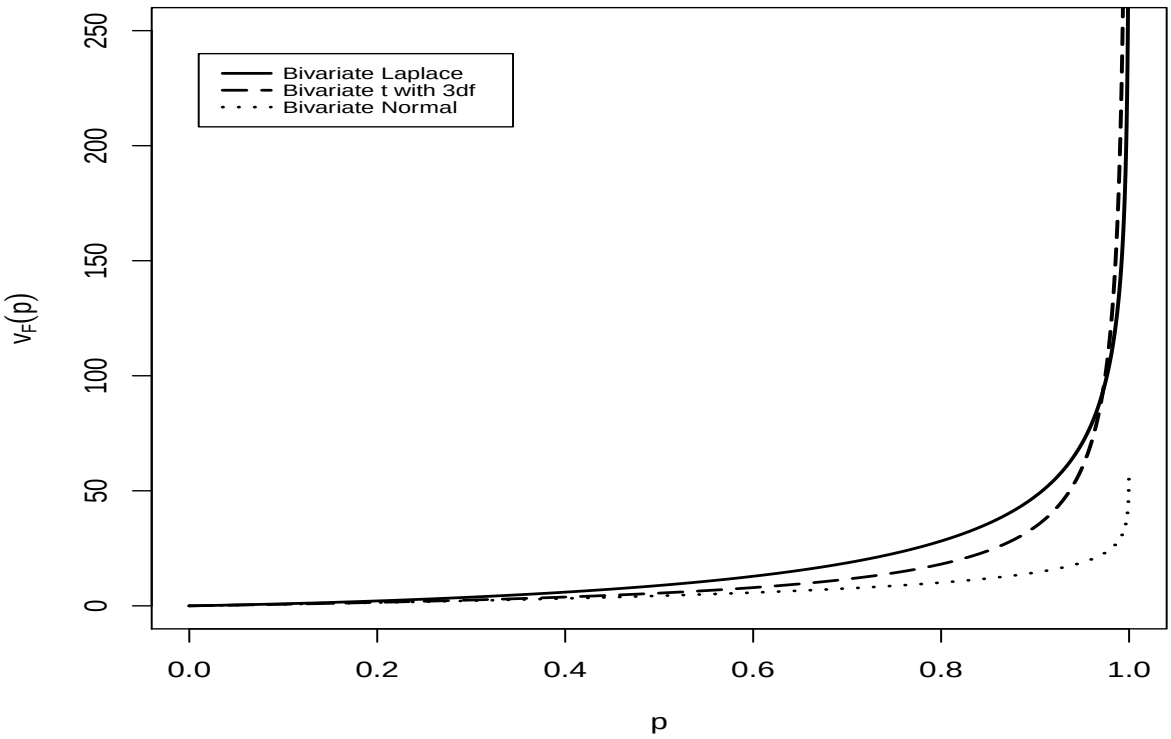

Figure 2.1: Scale plot for Bivariate normal , bivariate t distribution with 3 degrees of freedom and bivariate Laplace distribution, all with location $\boldsymbol{\theta} = (0, 0)^T$ and scale matrix $\Sigma = \mathbf{I}_2$.

similar, but the scale for standard bivariate Laplace and standard t distribution with 3 degrees of freedom increases at a faster rate than standard bivariate normal distribution. This suggests that standard bivariate Laplace distribution and standard t distribution with 3 degrees of freedom have larger tails compared to standard bivariate normal.

Figure 2.2 gives scale curves for the standard trivariate normal, standard trivariate Laplace and standard trivariate t distribution with 3 degrees of freedom and we see similar features.

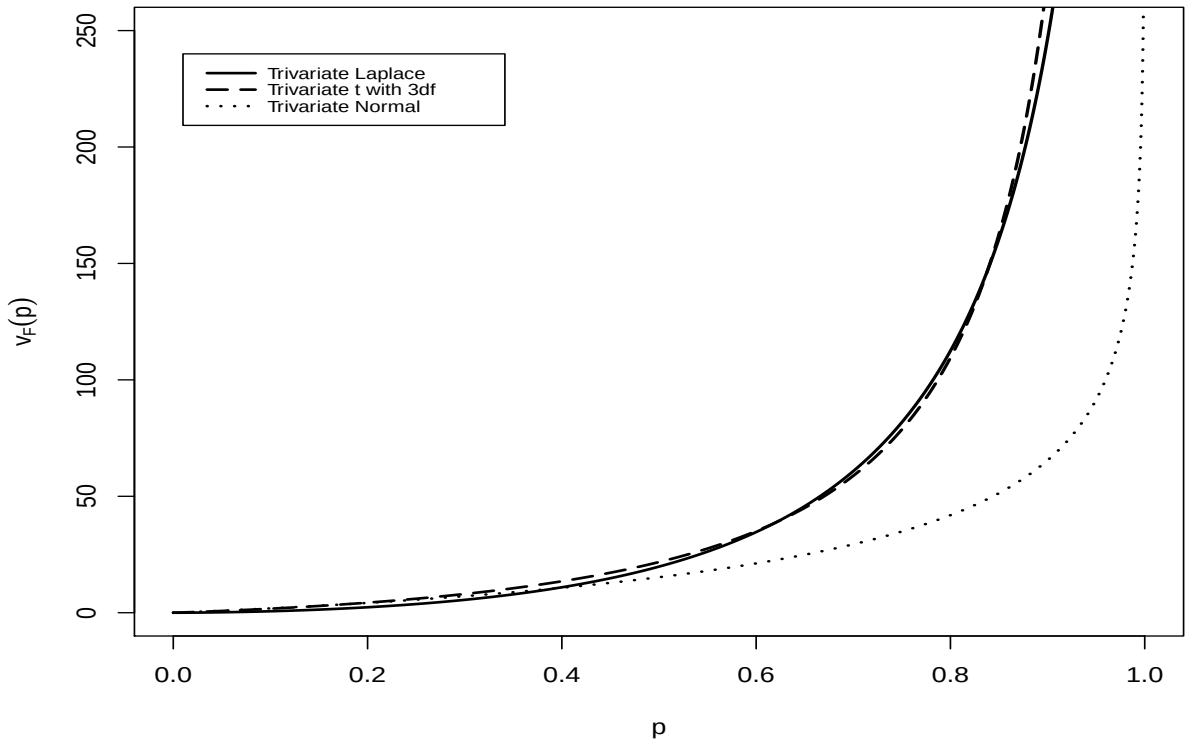

Figure 2.2: Scale plot for Trivariate normal, trivariate t distribution with 3 degrees of freedom and trivariate Laplace distribution all with location $\boldsymbol{\theta} = (0, 0, 0)^T$ and scale matrix $\Sigma = \mathbf{I}_3$.

2.2 Spherically and Elliptically Symmetric Distributions

We noted earlier that for a spherically symmetric distribution, the length of the random vector $||\mathbf{X} - \boldsymbol{\theta}||$ and its direction $\frac{\mathbf{X}-\boldsymbol{\theta}}{||\mathbf{X}-\boldsymbol{\theta}||}$ are distributed independently. The following theorem asserts that the length of the rank vector $R_F(\mathbf{x})$ depends only on the length of the vector $\mathbf{x} - \boldsymbol{\theta}$ and not on the direction of the vector $\mathbf{x} - \boldsymbol{\theta}$ where $\boldsymbol{\theta}$ is the centre of the spherical symmetry of F.

Theorem 2.2.1 *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution function F, which is spherically symmetric about $\boldsymbol{\theta} \in \mathbb{R}^d$. Then the spatial rank vector of $\mathbf{x}$ can be written as*

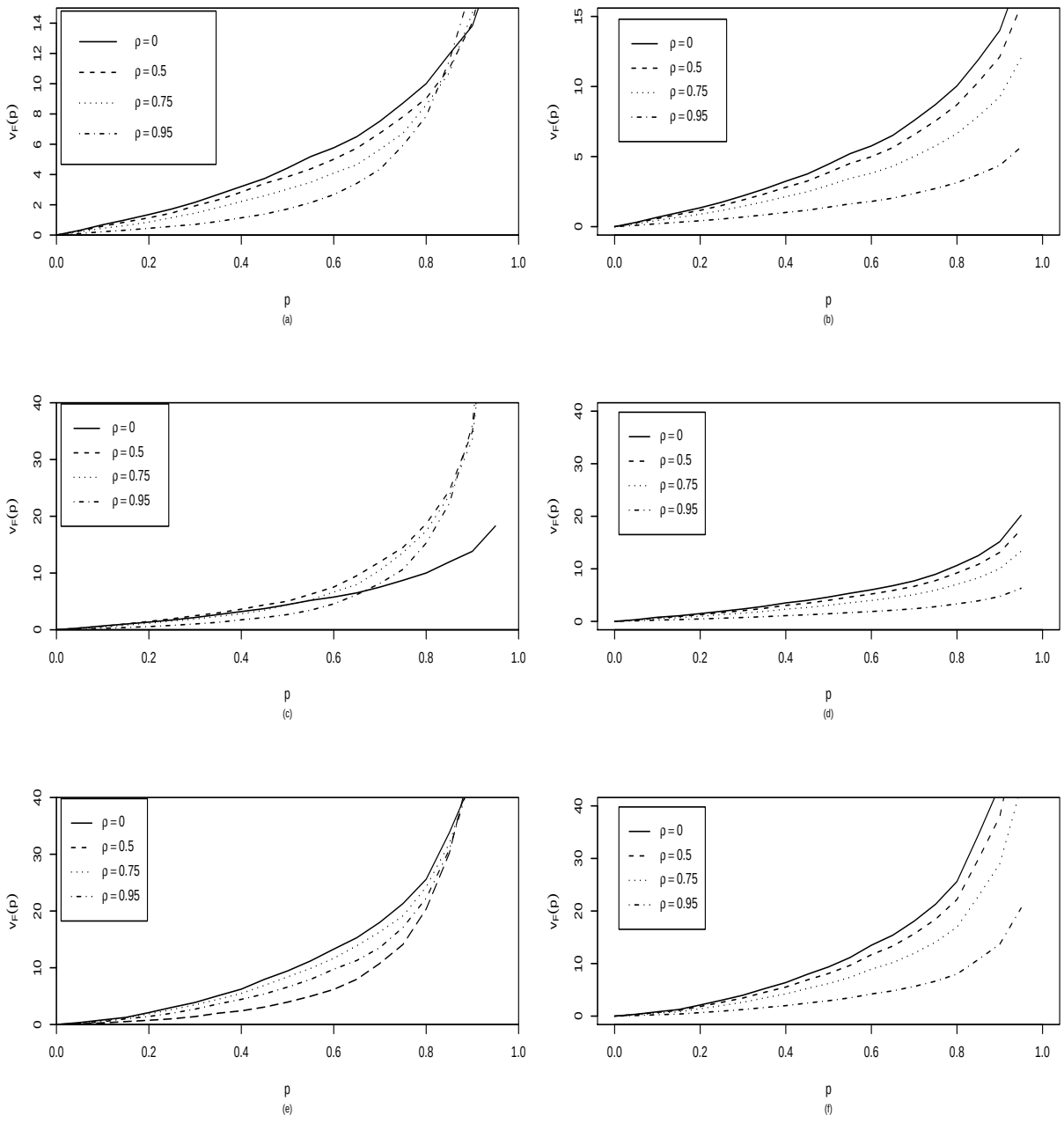

Figure 2.3: The first column of plots represent non-affine equivariant version of scale curves for bivariate normal, Laplace and t-distribution with 3 degrees of freedom. The second column of plots present the affine equivariant scale curves for the same distributions.

$$R_F(\mathbf{x}) = h(||\mathbf{x} - \boldsymbol{\theta}||)\frac{\mathbf{x} - \boldsymbol{\theta}}{||\mathbf{x} - \boldsymbol{\theta}||} \tag{2.8}$$

where $h(.)$ is an increasing function.

Proof : Without loss of generality let us take $\boldsymbol{\theta} = \mathbf{0}$. Let us define a function $S_F(\mathbf{x})$ as follows,

$$S_F(\mathbf{x}) = E_F\left\{||\mathbf{x} - \mathbf{X}|| - ||\mathbf{X}||\right\}. \tag{2.9}$$

such that $E_F\left\{||\mathbf{x} - \mathbf{X}|| - ||\mathbf{X}||\right\} < \infty$.

Let us consider a d-dimensional orthogonal matrix $\mathbf{A}$. Then we have,

$$\begin{aligned} S_F(\mathbf{Ax}) &= E_F\left\{||\mathbf{Ax} - \mathbf{X}|| - ||\mathbf{X}||\right\} \\ &= E_F\left\{\sqrt{(\mathbf{Ax} - \mathbf{X})^T(\mathbf{Ax} - \mathbf{X})} - \sqrt{\mathbf{X}^T\mathbf{X}}\right\} \\ &= E_F\left\{\sqrt{(\mathbf{x} - \mathbf{A}^T\mathbf{X})^T\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{A}^T\mathbf{X})} - \sqrt{(\mathbf{AX})^T\mathbf{AX}}\right\} \\ &= E_F\left\{||\mathbf{x} - \mathbf{A}^T\mathbf{X}|| - ||\mathbf{A}^T\mathbf{X}||\right\} \\ &= S_F(\mathbf{x}). \end{aligned} \tag{2.10}$$

The last equality holds as $\mathbf{X}$ and $\mathbf{A}^T\mathbf{X}$ have the same distribution for any orthogonal matrix $\mathbf{A}$. Hence if we rotate $\mathbf{x}$, then $S_F(\mathbf{x})$ remains unchanged. So $S_F(\mathbf{x})$ does not depend on $\frac{\mathbf{x}}{||\mathbf{x}||}$ and is a function of $||\mathbf{x}||$ only. Let $S_F(\mathbf{x}) = g(||\mathbf{x}||)$, then

$$S'_F(\mathbf{x}) = g'(||\mathbf{x}||).\frac{\mathbf{x}}{||\mathbf{x}||}. \tag{2.11}$$

Also from (2.9) we have,

$$S'_F(\mathbf{x}) = E_F\left(\frac{\mathbf{x} - \mathbf{X}}{||\mathbf{x} - \mathbf{X}||}\right) = R_F(\mathbf{x}). \tag{2.12}$$

Hence from (2.11) and (2.12) we have,

$$R_F(\mathbf{x}) = g'(||\mathbf{x}||).\frac{\mathbf{x}}{||\mathbf{x}||}.$$

Letting $g'(||\mathbf{x}||) = h(||\mathbf{x}||)$ we get (2.8). Now we need to show that $h(||\mathbf{x}||)$ is an increasing function. We can see that $||R_F(\mathbf{x})|| = h(||\mathbf{x}||)$. Suppose $||\mathbf{x}|| = r$. As h is a function of $||\mathbf{x}||$, so without loss of generality we can take $\mathbf{x} = (r, 0, \cdots, 0)^T$. Then $h(r) = ||R_F(\mathbf{x})||$ implies

$$R_F(\mathbf{x}) = h(r)(1, 0, \cdots, 0)^T. \tag{2.13}$$

Also,

$$R_F(\mathbf{x}) = \left(E_F\left(\frac{r - X_1}{||\mathbf{x} - \mathbf{X}||}\right), E_F\left(\frac{-X_2}{||\mathbf{x} - \mathbf{X}||}\right), \cdots, E_F\left(\frac{-X_d}{||\mathbf{x} - \mathbf{X}||}\right)\right)^T, \tag{2.14}$$

where $\mathbf{X} = (X_1, \cdots, X_d)^T$. From (2.13) and (2.14) we get,

$$h(r) = E_F\left(\frac{r - X_1}{\sqrt{(r - X_1)^2 + X_2^2 + \cdots + X_d^2}}\right). \tag{2.15}$$

Differentiating (2.15) we get,

$$h'(r) = E_F\left(\frac{X_2^2 + \cdots + X_d^2}{((r - X_1)^2 + X_2^2 + \cdots + X_d^2)^{\frac{3}{2}}}\right) \geq 0. \tag{2.16}$$

Thus h is an increasing function and that completes the proof. □

From Theorem 2.2.1, the central rank region $C_F(p)$ for the spherically symmetric case can be written as

$$C_F(p) = \{\mathbf{x} : ||\mathbf{x}|| \leq r_F(p)\} \tag{2.17}$$

where $r_F(p)$ is the p-th quantile of $||\mathbf{X} - \boldsymbol{\theta}||$, and hence we can write down the formula for the volume functional for the scale curve of a spherically symmetric distribution in a

nice closed form.

Corollary 2.2.1 *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector with distribution function F, which is spherically symmetric about $\boldsymbol{\theta}$. Also suppose that the multivariate central rank regions are denoted by $C_F(p)$. If $V_F(p)$ is the volume of the central rank region $C_F(p)$, then,*

$$V_F(p) = \frac{\pi^{\frac{d}{2}}(r_F(p))^d}{\Gamma(\frac{d}{2}+1)}, \tag{2.18}$$

where $r_F(p)$ is the p-th quantile of $||\mathbf{X} - \boldsymbol{\theta}||$.

Proof : From (2.17) we can say that $C_F(p)$ describes a sphere S^d when F is spherically symmetric. Thus $V_F(p)$ is the volume of the sphere S^d with radius $r_F(p)$. Hence, the volume of the central rank region for spherically symmetric distributions is obtained as in equation (2.18). □

We note that for the standard multivariate normal distribution in $\mathbb{R}^d$, $r_F^2(p)$ will be the p-th quantile of the χ_d^2 distribution; correspondingly, for the standard multivariate Laplace distribution in $\mathbb{R}^d$, $r_F(p)$ will be the p-th quantile of the $\Gamma(d,1)$ distribution and for the standard multivariate t distribution with ν degrees of freedom in $\mathbb{R}^d$, $r_F^2(p)/d$ will be the p-th quantile of the $F_{d,\nu}$ distribution.

A main deficiency of the definition of multivariate rank vector in (2.2) is that although it is invariant under orthogonal transformations, it is not invariant under a general affine transformation of the data. Thus the scale curve based on the central rank regions are not affine equivariant. We now take a look at the Figure 2.3. Let us take a closer look at the curve for the bivariate normal distribution in Figure 2.3(a). In Figure 2.3(a) we have plotted p against $V_F(p)$ taking different values of the correlation ρ for bivariate normal distribution with mean vector $\boldsymbol{\theta} = (0,0)^T$ and the scale matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

As the correlation ρ increases, the spread of the data points should become more and more elliptical and so the spread $V_F(p)$ should decrease with increasing value of ρ. But we can see that the scale curves are not in any specific order of ρ. The reason is that though the distributions are taking more and more elliptical form, the volume is being estimated with respect to a sphere as the volume here is based on a rank which is not affine invariant. So it requires that we modify the definition of the multivariate rank to be affine invariant.

To make the multivariate rank invariant under affine transformations, we utilise the transformation-retransformation method of Chakraborty (2001):

Definition 2.2.1 *Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be data points on $\mathbb{R}^d$. Suppose $\mathbf{X}_{i_0}, \mathbf{X}_{i_1}, \cdots, \mathbf{X}_{i_d}$ are $d+1$ data points in $\mathbb{R}^d$. Let $\alpha = \{i_0, i_1, \cdots, i_d\}$ denote the set of $(d+1)$ indices. Then we can define a data-driven coordinate system with $\mathbf{X}_{i_0}$ as the origin and the coordinate axes given by the vectors $\mathbf{X}_{i_1} - \mathbf{X}_{i_0}, \ldots, \mathbf{X}_{i_d} - \mathbf{X}_{i_0}$. Transforming all the observations into the new coordinate system and computing the multivariate rank vector, we can define*

$$R_{F_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1, i \notin \alpha}^{n} \frac{\{\mathbf{X}(\alpha)\}^{-1} (\mathbf{x} - \mathbf{X}_i)}{|| \{\mathbf{X}(\alpha)\}^{-1} (\mathbf{x} - \mathbf{X}_i)||} \tag{2.19}$$

under the assumption that $\{\mathbf{X}(\alpha)\}^{-1}$ exists, where $\mathbf{X}(\alpha)$ is the $d \times d$ transformation matrix, whose columns are $\mathbf{X}_{i_1} - \mathbf{X}_{i_0}, \mathbf{X}_{i_2} - \mathbf{X}_{i_0}, \cdots, \mathbf{X}_{i_d} - \mathbf{X}_{i_0}$.

The transformation matrix $\mathbf{X}(\alpha)$ is chosen in such a way that $\{\mathbf{X}(\alpha)\}^T \Sigma^{-1} \mathbf{X}(\alpha)$ becomes as close as possible to a diagonal matrix with equal diagonal elements where Σ is the scatter matrix associated with the underlying distribution of the $\mathbf{X}_i$'s. As the parameter Σ is unknown in practice, an affine equivariant consistent estimate of Σ, say $\hat{\Sigma}$, is used. After estimating $\hat{\Sigma}$, the transformation matrix $\mathbf{X}(\alpha)$ is chosen in such a way that the eigen values of the positive definite matrix $\{\mathbf{X}(\alpha)\}^T \hat{\Sigma}^{-1} \mathbf{X}(\alpha)$ becomes as equal as possible. To achieve this one can minimise the ratio between the arithmetic mean and the geometric mean of the eigen values of the matrix. The trace of a symmetric matrix gives the sum of its eigen values and the determinant of the symmetric matrix gives the

product of the eigen values. Thus a quantity $A(\alpha)$ is minimised over $\alpha = \{i_0, i_1, \cdots, i_d\}$ to achieve the goal, where $A(\alpha)$ is defined as

$$A(\alpha) = \frac{\text{trace}[\{\mathbf{X}(\alpha)\}^T \hat{\Sigma}^{-1} \mathbf{X}(\alpha)]/d}{[\det \ \{\mathbf{X}(\alpha)\}^T \hat{\Sigma}^{-1} \mathbf{X}(\alpha)]^{1/d}}.$$

For a detailed discussion on the optimality of this selection procedure see Chakraborty (2001). One can define $R_F(\mathbf{x})$ to be the population version of $R_{F_n}(\mathbf{x})$ defined above. The transformation-retransformation spatial rank $R_F(\mathbf{x})$ is affine invariant in the sense that if the distribution of $\mathbf{X}$ is denoted by F and the distribution of $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ is denoted by G for some non-singular matrix $\mathbf{A}$ and d-dimensional vector $\mathbf{b}$, then $R_G(\mathbf{A}\mathbf{x} + \mathbf{b}) = R_F(\mathbf{x})$. Using this affine invariant version of spatial rank, we can have similar central rank regions and scale curves as before. There are also other versions of affine invariant ranks available and these can also be used to form affine invariant central rank regions and scale curves, see Serfling (2010).

In Figure 2.3(b) we plotted the affine equivariant version of the scale curves based on the above definition of affine invariant rank vectors for the bivariate normal distribution with mean vector $\boldsymbol{\theta}$ and variance matrix Σ as before for different correlation coefficients ρ and we observe that the ordering between the scale curves with higher values of the correlation coefficient showing smaller scales. From now on, we restrict our discussion to the affine invariant versions of the multivariate ranks and corresponding scale curves.

We now consider a more general case of the spherically symmetric case, the elliptically symmetric distributions. The following theorem states a closed form formula for the affine equivariant scale curve when the underlying distribution is elliptically symmetric.

Theorem 2.2.2 *If the distribution of the random vector* $\mathbf{X}$ *is elliptically symmetric, that is, it has a density of the form*

$$f(\mathbf{x}) = |\Sigma|^{-\frac{1}{2}} h((\mathbf{x} - \boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\theta})) \tag{2.20}$$

for some strictly decreasing, continuous, non-negative scalar function h and positive definite matrix Σ *and* $R_F(\mathbf{x})$ *is the affine invariant spatial rank function as defined before, we have*

$$V_F(p) = \frac{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\zeta_p^d}{\Gamma(\frac{d}{2}+1)} \tag{2.21}$$

where $P((\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta}) \leq \zeta_p^2) = p$.

Proof: Let $\mathbf{Y} = \Sigma^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\theta})$, then $\mathbf{Y}$ has a spherically symmetric distribution about $\mathbf{0}$. Let F_0 denote the distribution of $\mathbf{Y}$. Then by the affine equivariance of the rank function $R_F(\mathbf{x})$, we have $R_{F_0}(\mathbf{Y}) = R_{F_0}(\Sigma^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\theta})) = R_F(\mathbf{X})$. This implies that $r_F(p) = r_{F_0}(p)$, where $r_F(p)$ and $r_{F_0}(p)$ are the p-th quantiles of the distributions of $||R_F(\mathbf{X})||$ and $||R_{F_0}(\mathbf{Y})||$ respectively. Thus,

$$\begin{aligned} C_F(p) &= \{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_F(p)\} \\ &= \left\{\mathbf{x} : ||R_{F_0}(\Sigma^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\theta}))|| \leq r_F(p)\right\} \\ &= \left\{\Sigma^{\frac{1}{2}}\mathbf{y}+\boldsymbol{\theta} : ||R_{F_0}(\mathbf{y})|| \leq r_{F_0}(p)\right\} \\ &= \Sigma^{\frac{1}{2}}C_{F_0}(p) + \boldsymbol{\theta}. \end{aligned} \tag{2.22}$$

Therefore, $V_F(p) = |\Sigma|^{\frac{1}{2}}V_{F_0}(p)$. Now proceeding as in the proof of Corollary 2.2.1, it is easy to note that $||R_{F_0}(\mathbf{y})||$ is a non-negative increasing function of $||\mathbf{y}||$ only, which implies that

$$V_{F_0}(p) = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}\zeta_p^d \tag{2.23}$$

where ζ_p is given by,

$$P((\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta}) \leq \zeta_p^2) = P(\mathbf{Y}^T\mathbf{Y} \leq \zeta_p^2) = p. \tag{2.24}$$

This proves the theorem. □

2.3 Scale-Scale Plot

We have seen earlier that verifying the distributional assumptions of the multivariate data and comparing samples of multivariate data are not easy tasks. In this section we propose a scale-scale plot to compare multivariate distributions as a generalisation of the univariate quantile quantile plot. If F and G are two d-dimensional distributions, we define a *scale-scale plot* as follows.

Definition 2.3.1 *Let F and G denote two d-dimensional distribution functions and $V_F(p)$ and $V_G(p)$ be the volumes of the affine equivariant central rank regions $C_F(p)$ and $C_G(p)$ respectively for $0 < p < 1$. Then a* **scale-scale plot** *is a plot of $V_F(p)$ against $V_G(p)$.*

If $F = G$, then the scale-scale plot will be a 45° line passing through the origin. Since $V_F(0) = 0$ for all continuous distributions F, the scale-scale plot will always pass through the origin and we cannot detect a change in origin with the scale-scale plot. However, if we can detect that the two multivariate distributions are same up to a location shift, quite often it is not difficult to estimate the location shift efficiently. For elliptically symmetric distributions F and G, we have the following characterisation:

Theorem 2.3.1 *Assume that $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^d$ have distributions F and G, respectively, which are elliptically symmetric. Then $\mathbf{Y} \stackrel{d}{=} \mathbf{A}\mathbf{X} + \mathbf{b}$ for some $d \times d$ matrix $\mathbf{A}$ and d-dimensional vector $\mathbf{b}$ if and only if $V_G(p) = k.V_F(p)$, $0 < p < 1$ for some $k > 0$.*

Proof : Suppose that the probability density functions of $\mathbf{X}$ and $\mathbf{Y}$ are given by

$$f(\mathbf{x}) = |\Sigma_{\mathbf{X}}|^{-\frac{1}{2}} h_{\mathbf{X}}((\mathbf{x} - \boldsymbol{\theta}_{\mathbf{X}})^T \Sigma_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\theta}_{\mathbf{X}}))$$

and

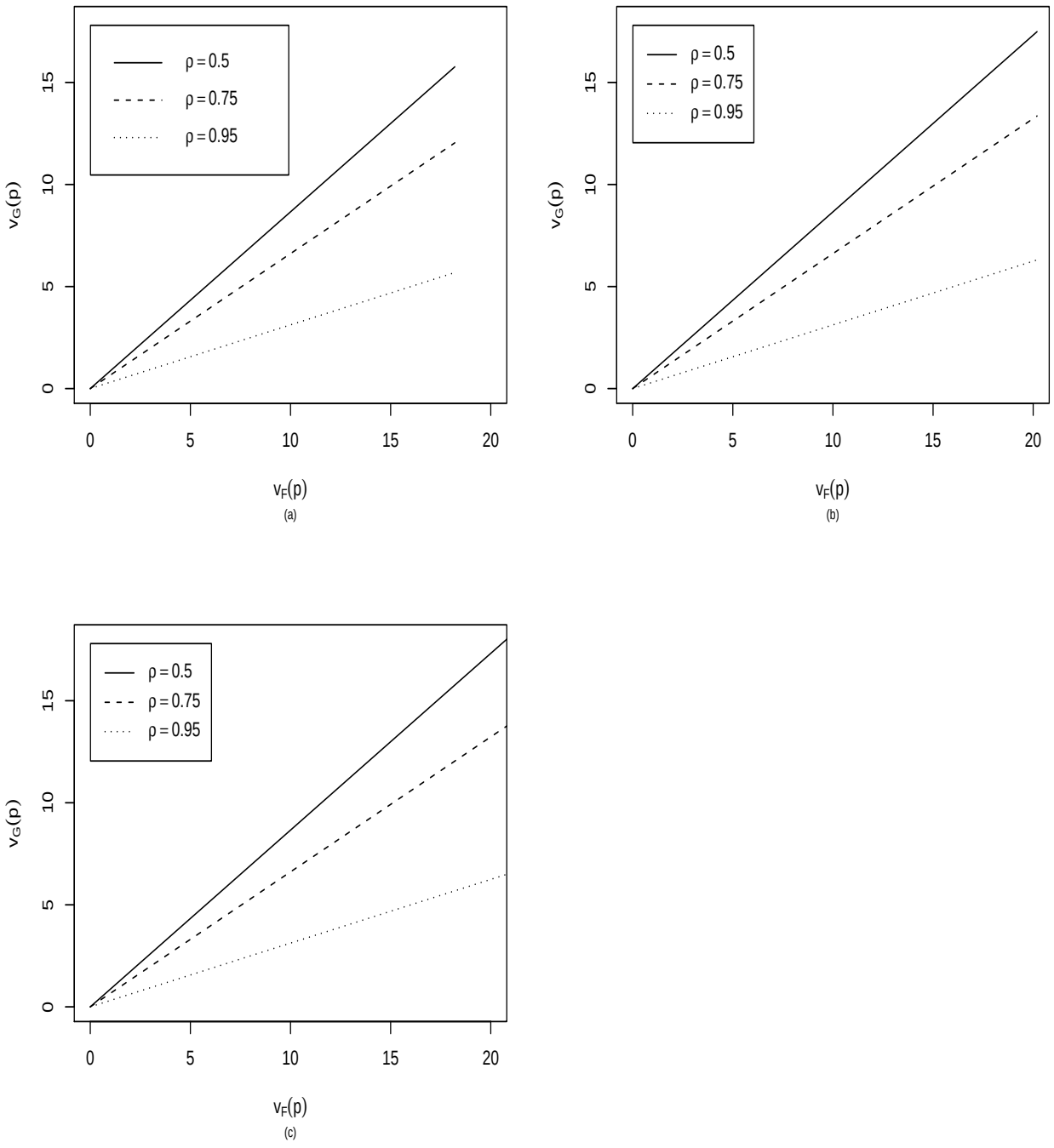

Figure 2.4: Affine equivariant version of Scale-Scale plots for (a) bivariate normal distribution (b)bivariate t distribution with 3 df (c)bivariate Laplace distribution comparing the standard distributions ($\rho = 0$) with different values of ρ.

$$g(\mathbf{y}) = |\Sigma_{\mathbf{Y}}|^{-\frac{1}{2}} g_{\mathbf{Y}}((\mathbf{y}-\boldsymbol{\theta}_{\mathbf{Y}})^T \Sigma_{\mathbf{Y}}^{-1}(\mathbf{y}-\boldsymbol{\theta}_{\mathbf{Y}}))$$

respectively. Then by Theorem 2.2.2, we have

$$V_F(p) = \frac{\pi^{\frac{d}{2}} |\Sigma_{\mathbf{X}}|^{\frac{1}{2}} \zeta_{\mathbf{X},p}^d}{\Gamma(\frac{d}{2}+1)} \tag{2.25}$$

and

$$V_G(p) = \frac{\pi^{\frac{d}{2}} |\Sigma_{\mathbf{Y}}|^{\frac{1}{2}} \zeta_{\mathbf{Y},p}^d}{\Gamma(\frac{d}{2}+1)} \tag{2.26}$$

where $\zeta_{\mathbf{X},p}^d$ and $\zeta_{\mathbf{Y},p}^d$ are the p-th quantiles of the distributions of $\sqrt{(\mathbf{X}-\boldsymbol{\theta}_{\mathbf{X}})^T \Sigma_{\mathbf{X}}^{-1}(\mathbf{X}-\boldsymbol{\theta}_{\mathbf{X}})}$ and $\sqrt{(\mathbf{Y}-\boldsymbol{\theta}_{\mathbf{Y}})^T \Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y}-\boldsymbol{\theta}_{\mathbf{Y}})}$ respectively.

If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, we have $h_{\mathbf{X}} = g_{\mathbf{Y}}$, $\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma\mathbf{A}^T$ and $\theta_{\mathbf{Y}} = \mathbf{A}\theta_{\mathbf{X}} + \mathbf{b}$, which implies

$$(\mathbf{X}-\boldsymbol{\theta}_{\mathbf{X}})^T \Sigma_{\mathbf{X}}^{-1}(\mathbf{X}-\boldsymbol{\theta}_{\mathbf{X}}) = (\mathbf{Y}-\boldsymbol{\theta}_{\mathbf{Y}})^T \Sigma_{\mathbf{Y}}^{-1}(\mathbf{Y}-\boldsymbol{\theta}_{\mathbf{Y}})$$

and thus

$$\zeta_{\mathbf{X},p} = \zeta_{\mathbf{Y},p} \text{ for all } p \in [0,1).$$

Therefore,

$$V_G(p) = |\mathbf{A}| V_F(p).$$

Now to prove the converse, let us assume that $V_G(p) = k.V_F(p)$, then again by Theorem 2.2.2, we have

$$\zeta_{\mathbf{Y},p} = k^* \zeta_{\mathbf{X},p}, \text{ for some } k^* > 0 \text{ and for all } p \in [0,1).$$

Then by elliptic symmetry of the distributions of $\mathbf{X}$ and $\mathbf{Y}$, we have

$$\Sigma_{\mathbf{Y}}^{-\frac{1}{2}}(\mathbf{Y}-\boldsymbol{\theta}_{\mathbf{Y}}) = k^* \Sigma_{\mathbf{X}}^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\theta}_{\mathbf{X}})$$

and therefore

$$\mathbf{Y} = k^* \Sigma_{\mathbf{Y}}^{\frac{1}{2}} \Sigma_{\mathbf{X}}^{-\frac{1}{2}} \mathbf{X} - k^* \Sigma_{\mathbf{Y}}^{\frac{1}{2}} \Sigma_{\mathbf{X}}^{-\frac{1}{2}} \boldsymbol{\theta}_{\mathbf{X}} + \boldsymbol{\theta}_{\mathbf{Y}}$$

which proves the theorem with

$$\mathbf{A} = k^* \Sigma_{\mathbf{Y}}^{\frac{1}{2}} \Sigma_{\mathbf{X}}^{-\frac{1}{2}} \text{ and } \mathbf{b} = -k^* \Sigma_{\mathbf{Y}}^{\frac{1}{2}} \Sigma_{\mathbf{X}}^{-\frac{1}{2}} \boldsymbol{\theta}_{\mathbf{X}} + \boldsymbol{\theta}_{\mathbf{Y}}. \quad \square$$

Theorem 2.3.1 suggests that if $\mathbf{X}$ and $\mathbf{Y}$ are in the same elliptically symmetric family of distributions but possibly differ in the location parameter $\boldsymbol{\theta}$ and the scale matrix Σ, the scale-scale plot will be a straight line. The slope of that straight line is determined by the determinants of the scale matrices associated with them. Figure 2.4(a) shows scale-scale plots for bivariate normal distributions with the mean vector $\boldsymbol{\theta} = (0, 0)^T$ and the scale matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

for $\rho = 0.5,\ 0.75$ and 0.95, where ρ is the correlation coefficient between $\mathbf{X}$ and $\mathbf{Y}$, compared with the standard bivariate normal distribution.

We can see that the slopes of the straight lines in these scale-scale plots are decreasing with the increasing value of the correlation coefficient ρ. Figures 2.4(b) and (c) show similar scale-scale plots for bivariate Laplace distribution and bivariate t distribution with 3 degrees of freedom respectively and we observe the similar decrease in slopes for the scale curves.

CHAPTER 3

ASYMPTOTIC DISTRIBUTION OF V_{F_N}

In the previous chapter, we have defined the volume functional based on the rank regions. Here in this chapter we will be discussing the asymptotic properties of the corresponding sample volume functional $\{V_{F_n}(p), 0 < p < 1\}$. Wang and Serfling (2006) discussed a similar idea for a general data depth, where they showed the convergence of the scale curve to a Brownian bridge under some very restrictive set up for a univariate case. In their works they considered a general depth function $D(x, F)$ and its corresponding central region $C_{F,D}(p)$ for a distribution F, with volume $V_{F,D}(p)$. Following the concepts of Einmahl and Mason (1990), the $V_{F,D}(p)$ is a generalised quantile function. For the sample version they established a weak convergence of the quantile process. Wang and Serfling (2006) modified the definition of the scale curve by restricting the central regions to some specific form and showed that the modified volume functional converges weakly to the Brownian bridge over a closed set.

In the next section we will show that the scale curve converges to a Brownian bridge under some very general conditions, without the need to modify the central regions.

3.1 Asymptotic Results

For the rest of this thesis, unless stated otherwise, $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n \in \mathbb{R}^d$ are independent with common distribution function F which is continuous with respect to the Lebesgue

measure in $\mathbb{R}^d$. We now show that $\{V_{F_n}(p), 0 < p < 1\}$ converges to a Brownian bridge under some regularity conditions which are stated as follows:

A1 The probability measure P of the data in $\mathbb{R}^d$ has corresponding cdf F possessing a density $f(\mathbf{x})$, which is positive for all $\mathbf{x} \in \text{supp}(F)$.

A2 $||R_F(\mathbf{x})||$ is continuous in $\mathbf{x}$.

A3 The set $\{\mathbf{x} : ||R_F(\mathbf{x})|| = r\}$ is nonempty for all $0 < r < r_F^* = \sup_{\mathbf{x}} ||R_F(\mathbf{x})||$.

A4 As a function of p, with $r_F(p) \in (0, r_F^*)$, $V_F(p)$ is finite, strictly increasing and possesses a continuous derivative.

A5 V_F is twice differentiable at p. v_F' is bounded in a neighbourhood of p, $0 < p < 1$, where $v_F(p) = \frac{d}{dp} V_F(p)$. Moreover $v_F(p)$ is bounded away from 0 in a neighbourhood of p. The bounds may be functions of p.

As p increases, $r_F(p)$ also increases and by assumptions A1, A2 and A3 we can say that the monotonicity is strict. Assumptions A4 and A5 signify the existence of the scaling factor $v_F(p)$. These assumptions, as we will see later in Section 3.2, are satisfied by some of the most common elliptically symmetric distributions, eg. multivariate normal, multivariate Laplace and multivariate t distributions. Similar assumptions can also be seen in Serfling (2002b), as well as in Bahadur (1966) in the context of quantile functions.

We require the following inverse volume functionals as follows before we state the main theorem:

Definition 3.1.1

$$V_F^{-1}(y) = F\text{-probability of the smallest central region } C_F(.) \text{ having volume } \geq y.$$
$$V_{F_n}^{-1}(y) = F_n\text{-probability of the smallest central region } C_{F_n}(.) \text{ having volume } \geq y.$$

Note that if A5 is true, then V_F^{-1} is well defined. Moreover

$$\frac{d}{dy}V_F^{-1}(y)\big|_{y=V_F(p)} = \frac{1}{v_F(V_F^{-1}(V_F(p)))} = \frac{1}{v_F(p)} > 0,$$

which implies that V_F^{-1} is strictly increasing in a neighbourhood of $V_F(p)$.

We are now ready to state the main theorem of this chapter, which describes the asymptotic distribution of the volume functional for the rank based regions.

Theorem 3.1.1 *Suppose the assumptions A1-A5 hold. Then*

$$\left\{\sqrt{n}\left\{v_F(p)\right\}^{-1}\left(V_{F_n}(p) - V_F(p)\right), 0 < p < 1\right\} \tag{3.1}$$

converges to a Brownian Bridge with covariance kernel $\Gamma(p_1, p_2) = p_1(1 - p_2)$ *where* $0 < p_1 \leq p_2 < 1$ *and* $n \to \infty$.

Note that the covariance kernel can alternatively written as,

$$\Gamma(p_1, p_2) = \begin{cases} p(1-p), \text{ if } p_1 = p_2 = p \\ \min\left\{p_1, p_2\right\} - p_1 p_2, \text{ if } p_1 \neq p_2. \end{cases}$$

This theorem tells us that the volume functional for the rank based regions converges to a Brownian Bridge at a $\sqrt{n}$ rate under some general conditions. The proof for the above theorem is arrived at through a series of lemmas as proved below.

Lemma 3.1.1 *Let*

$$Y_{in} = \left|\left|\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{X}_j - \mathbf{X}_i}{||\mathbf{X}_j - \mathbf{X}_i||}\right|\right|, Z_{in} = \left|\left|E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{X}_i}{||\mathbf{X} - \mathbf{X}_i||}\right)\right|\right| \tag{3.2}$$

where $\mathbf{X} \sim F$. *Then under conditions A1 and A2,*

$$Y_{in} - Z_{in} = O_P\left(\frac{1}{\sqrt{n}}\right) \tag{3.3}$$

Proof : Let us condition on $\mathbf{X}_1 = \mathbf{x}_1$. Then,

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=2}^{n}\frac{\mathbf{X}_j - \mathbf{x}_1}{||\mathbf{X}_j - \mathbf{x}_1||} - E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{x}_1}{||\mathbf{X} - \mathbf{x}_1||}\right)\right)$$
$$\Rightarrow N_d\left(\mathbf{0}, E_F\left\{\left(\frac{\mathbf{X} - \mathbf{x}_1}{||\mathbf{X} - \mathbf{x}_1||}\right)\left(\frac{\mathbf{X} - \mathbf{x}_1}{||\mathbf{X} - \mathbf{x}_1||}\right)^T\right\}\right) \text{ as } n \to \infty, \forall\, \mathbf{x}_1 \in \mathbb{R}^d \quad (3.4)$$

where the density of the limiting multivariate normal is $\phi(\mathbf{x}|\mathbf{x}_1)$.

Then as $n \to \infty$, the limiting density function of

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=2}^{n}\frac{\mathbf{X}_j - \mathbf{X}_1}{||\mathbf{X}_j - \mathbf{X}_1||} - E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{X}_1}{||\mathbf{X} - \mathbf{X}_1||}\right)\right)$$

is given by $\phi(\mathbf{x}|\mathbf{x}_1).f_{\mathbf{X}_1}(\mathbf{x}_1)$. Hence

$$\frac{1}{n}\sum_{j=2}^{n}\frac{\mathbf{X}_j - \mathbf{X}_1}{||\mathbf{X}_j - \mathbf{X}_1||} - E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{X}_1}{||\mathbf{X} - \mathbf{X}_1||}\right) = O_P\left(\frac{1}{\sqrt{n}}\right) \quad (3.5)$$

It follows that,

$$\left|\left|\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{X}_j - \mathbf{X}_i}{||\mathbf{X}_j - \mathbf{X}_i||}\right|\right| - \left|\left|E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{X}_i}{||\mathbf{X} - \mathbf{X}_i||}\right)\right|\right| = O_P\left(\frac{1}{\sqrt{n}}\right) \quad (3.6)$$

Recall from (3.2),

$$Y_{in} = \left|\left|\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{X}_j - \mathbf{X}_i}{||\mathbf{X}_j - \mathbf{X}_i||}\right|\right|, Z_{in} = \left|\left|E_{\mathbf{X}}\left(\frac{\mathbf{X} - \mathbf{X}_i}{||\mathbf{X} - \mathbf{X}_i||}\right)\right|\right|$$

Hence we have our result,

$$Y_{in} - Z_{in} = O_P\left(\frac{1}{\sqrt{n}}\right).\square$$

Lemma 3.1.2 *Let Y_{in} and Z_{in} be as in (3.2). Define, $Y^*_{in} = I(Y_{in} \leq y), Z^*_{in} = I(Z_{in} \leq y)$ where $y \in \mathbb{R}$. Then if A1 and A2 holds,*

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y_{in}^{*} - \frac{1}{n}\sum_{i=1}^{n} Z_{in}^{*}\right) = o(e^{-n}) \ w.p. \ 1. \tag{3.7}$$

Proof:

$$\begin{aligned} E(exp(it(Y_{in}^{*} - Z_{in}^{*})e^{n})) &= 1.P[Y_{in}^{*} - Z_{in}^{*} = 0] + exp(ite^{n}).P[Y_{in}^{*} - Z_{in}^{*} = 1] \\ &+ exp(-ite^{n}).P[Y_{in}^{*} - Z_{in}^{*} = -1] \end{aligned} \tag{3.8}$$

When

$$Y_{in}^{*} - Z_{in}^{*} = 1 \Leftrightarrow Y_{in} \leq y < Z_{in}$$

and

$$Y_{in}^{*} - Z_{in}^{*} = -1 \Leftrightarrow Z_{in} \leq y < Y_{in}.$$

Now, from Lemma 3.1.1 we have for every $\epsilon > 0$, $\exists c > 0$ such that

$$P[\sqrt{n}|Y_{in} - Z_{in}| > c] < \epsilon, \forall n \tag{3.9}$$

where Y_{in} and Z_{in} are defined in (3.2). Let us fix an ϵ and choose a corresponding c such that (3.9) is satisfied.

$$\begin{aligned} & P[Y_{in}^{*} - Z_{in}^{*} = 1] = P[Y_{in} \leq y < Z_{in}] \\ = \ & P[Y_{in} \leq y < Z_{in}, \sqrt{n}|Y_{in} - Z_{in}| > c] \\ + \ & P[Y_{in} \leq y < Z_{in}, \sqrt{n}|Y_{in} - Z_{in}| \leq c] \\ \leq \ & \epsilon + P\left[Y_{in} \leq y < Z_{in}, |Y_{in} - Z_{in}| \leq \frac{c}{\sqrt{n}}\right] \text{ (from Lemma 3.1.1)} \\ \leq \ & \epsilon + P\left[Z_{in} - \frac{c}{\sqrt{n}} \leq y < Z_{in}\right] \\ = \ & \epsilon + P\left[y < Z_{in} \leq y + \frac{c}{\sqrt{n}}\right] \end{aligned} \tag{3.10}$$

As Z_{in} has a bounded continuous density at a compact neighbourhood of y by A1 and A2,

$$P\left(y < Z_{in} \leq y + \frac{c}{\sqrt{n}}\right) = \int_{y}^{y+\frac{c}{\sqrt{n}}} f_{Z_{in}}(x)dx \leq k\frac{c}{\sqrt{n}}$$

for some $k \in \mathbb{R}^+$. Hence from (3.10)

$$P[Y_{in}^* - Z_{in}^* = 1] \leq \epsilon + \frac{kc}{\sqrt{n}} \tag{3.11}$$

Similarly,

$$P[Y_{in}^* - Z_{in}^* = -1] \leq \epsilon + \frac{k_1 c_1}{\sqrt{n}} \text{ for some } k_1 \in \mathbb{R}^+. \tag{3.12}$$

So,

$$\begin{aligned}
& E(exp(it(Y_{in}^* - Z_{in}^*)e^n)) \\
= \; & 1.P[Y_{in}^* - Z_{in}^* = 0] + exp(ite^n).P[Y_{in}^* - Z_{in}^* = 1] \\
+ \; & exp(-ite^n).P[Y_{in}^* - Z_{in}^* = -1] \\
= \; & 1 - P[Y_{in}^* - Z_{in}^* = 1] - P[Y_{in}^* - Z_{in}^* = -1] \\
+ \; & exp(ite^n).P[Y_{in}^* - Z_{in}^* = 1] + exp(-ite^n).P[Y_{in}^* - Z_{in}^* = -1] \\
= \; & 1 + (exp(ite^n - 1).P[Y_{in}^* - Z_{in}^* = 1] \\
+ \; & (exp(-ite^n) - 1).P[Y_{in}^* - Z_{in}^* = -1]
\end{aligned} \tag{3.13}$$

Now,

$$
\begin{aligned}
& |(exp(ite^n - 1).P[Y^*_{in} - Z^*_{in} = 1] + (exp(-ite^n) - 1).P[Y^*_{in} - Z^*_{in} = -1]| \\
\leq\ & |(exp(ite^n - 1)|.P[Y^*_{in} - Z^*_{in} = 1] + |(exp(-ite^n) - 1)|.P[Y^*_{in} - Z^*_{in} = -1] \\
\leq\ & 2\left(\epsilon + \frac{kc}{\sqrt{n}} + \epsilon + \frac{k_1 c_1}{\sqrt{n}}\right), \forall \epsilon > 0 \\
\to\ & 0, \text{ as } n \to \infty. \qquad (3.14)
\end{aligned}
$$

Thus,

$$E(exp(it(Y^*_{in} - Z^*_{in})e^n)) \to 1, \text{ as } n \to \infty$$

and hence

$$(Y^*_{in} - Z^*_{in})e^n \xrightarrow{a.s.} 0, \text{ as } n \to \infty.$$

Hence,

$$e^n\left(\frac{1}{n}\sum_{i=1}^{n} Y^*_{in} - \frac{1}{n}\sum_{i=1}^{n} Z^*_{in}\right) \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \qquad (3.15)$$

So,

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y^*_{in} - \frac{1}{n}\sum_{i=1}^{n} Z^*_{in}\right) = o(e^{-n}) \text{ with probability 1.}$$

Thus the lemma is proved. □

We now recall from Section 2.1 that $r_F(p)$ is the p-th quantile of $||R_F(\mathbf{X})||$ and $r_{F_n}(p)$ is the sample version.

Lemma 3.1.3 *Under A1, A2 and A3,* $r_{F_n}(p) \xrightarrow{a.s.} r_F(p)$ *as* $n \to \infty$ *for all* $0 < p < 1$. *Moreover*

$$r_{F_n}(p) - r_F(p) = O_P\left(\frac{1}{\sqrt{n}}\right) \qquad (3.16)$$

.

Proof: From Lemma 3.1.1, for every $\epsilon > 0$, $\exists\ c > 0$ such that, $P\left[\sqrt{n}|Y_{in} - Z_{in}| > c\right] < \epsilon, \forall n$ where Y_{in} and Z_{in} are defined in (3.2).
Now $I\left(||E_{\mathbf{X}}\left(\frac{\mathbf{X}-\mathbf{X}_i}{||\mathbf{X}-\mathbf{X}_i||}\right)|| \leq y\right)$, $1 \leq i \leq n$, are independently and identically distributed random variables. Hence by the strong law of large numbers we can say,

$$\frac{1}{n}\sum_{i=1}^{n} I\left(||E_{\mathbf{X}}\left(\frac{\mathbf{X}-\mathbf{X}_i}{||\mathbf{X}-\mathbf{X}_i||}\right)|| \leq y\right) - P\left[||E_{\mathbf{X}}\left(\frac{\mathbf{X}-\mathbf{X}_i}{||\mathbf{X}-\mathbf{X}_i||}\right)|| \leq y\right] = o(1) \text{ w.p. 1.} \quad (3.17)$$

So, by Lemma 3.1.2 and (3.17), we get

$$\frac{1}{n}\sum_{i=1}^{n}\left(I\left(||\frac{1}{n}\sum_{j=1,j\neq i}^{n}\frac{\mathbf{X}_j-\mathbf{X}_i}{||\mathbf{X}_j-\mathbf{X}_i||}|| \leq y\right)\right) - P\left[||E_{\mathbf{X}}\left(\frac{\mathbf{X}-\mathbf{X}_i}{||\mathbf{X}-\mathbf{X}_i||}\right)|| \leq y\right] = o(1) \text{ w.p. 1.} \quad (3.18)$$

Let $F^*(y) = P\left(||E_{\mathbf{X}}\left(\frac{\mathbf{X}-\mathbf{X}_i}{||\mathbf{X}-\mathbf{X}_i||}\right)|| \leq y\right)$ and $F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} I\left(||\frac{1}{n}\sum_{j\neq i}^{n}\frac{\mathbf{X}_j-\mathbf{X}_i}{||\mathbf{X}_j-\mathbf{X}_i||}|| \leq y\right)$ then $F_n^*(y) \xrightarrow{a.s.} F^*(y)$.

Let $r_F(p)$ be the unique solution (existence ensured by A3) of $F^*(x-) \leq p \leq F^*(x)$. Then

$$F^*(r_F(p)-\epsilon) < p < F^*(r_F(p)+\epsilon), \text{ for all } \epsilon > 0.$$

Now $F_n^*(r_F(p)-\epsilon) \xrightarrow{a.s.} F^*(r_F(p)-\epsilon)$ and $F_n^*(r_F(p)+\epsilon) \xrightarrow{a.s.} F^*(r_F(p)+\epsilon)$ as $n \to \infty$. Hence

$$\begin{aligned}
P(F_m^*(r_F(p)-\epsilon) < p < F_m^*(r_F(p)+\epsilon), \forall m \geq n) &\to 1, \text{ as } n \to \infty \\
\Rightarrow P(r_F(p)-\epsilon < {F_m^*}^{-1}(p) < r_F(p)+\epsilon, \forall m \geq n) &\to 1, \text{ as } n \to \infty \\
\Rightarrow P\left[Sup_{m\geq n}|r_{F_m}(p) - r_F(p)| > \epsilon\right] &\to 0, n \to \infty \\
\Rightarrow r_{F_n}(p) &\xrightarrow{a.s.} r_F(p) \qquad (3.19)
\end{aligned}$$

by Hoeffding's lemma, following Theorem 2.3.2 of Serfling (1980).

Now we establish (3.16). We take $G_n(t)$ as

$$\begin{aligned} G_n(t) &= P(\sqrt{n}|r_{F_n}(p) - r_F(p)| \leq t) \\ &= P\left(r_F(p) - \frac{t}{\sqrt{n}} \leq r_{F_n}(p) \leq r_F(p) + \frac{t}{\sqrt{n}}\right) \\ &= P\left(p \leq F_n^*\left(r_F(p) + \frac{t}{\sqrt{n}}\right)\right) \\ &= P(np \leq z_n), \end{aligned}$$

where $z_n = \sum_{i=1}^n V_{in}$ with $V_{in} = I\left(||\frac{1}{n}\sum_{j\neq i}\frac{\mathbf{X}_j - \mathbf{X}_i}{||\mathbf{X}_j - \mathbf{X}_i||}|| \leq r_F(p) + \frac{t}{\sqrt{n}}\right)$. Now

$$\begin{aligned} \frac{z_n}{n} &= \frac{1}{n}\sum_{i=1}^n V_{in} \\ &= \frac{1}{n}\sum_{i=1}^n I\left(||\frac{1}{n}\sum_{j\neq i}\frac{\mathbf{X}_j - \mathbf{X}_i}{||\mathbf{X}_j - \mathbf{X}_i||}|| \leq r_F(p) + \frac{t}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_{i=1}^n I\left(\left|\left|E_X\left(\frac{\mathbf{X} - \mathbf{X}_i}{||\mathbf{X} - \mathbf{X}_i||}\right)\right|\right| \leq r_F(p) + \frac{t}{\sqrt{n}}\right) + o(e^{-n}) \text{ w.p. 1 by Lemma 3.1.2} \end{aligned}$$

Thus,

$$z_n = \sum_{i=1}^n I\left(\left|\left|E_X\left(\frac{\mathbf{X} - \mathbf{X}_i}{||\mathbf{X} - \mathbf{X}_i||}\right)\right|\right| \leq r_F(p) + \frac{t}{\sqrt{n}}\right) + o(ne^{-n}) \tag{3.20}$$

has a binomial distribution, $Bin\left(n, F^*\left(r_F(p) + \frac{t}{\sqrt{n}}\right)\right)$. Hence proceeding as the proof of Theorem A, Serfling (1980), (pp. 78), we can show that for some suitable c, $G_n(t) \approx \Phi(ct) + O\left(\frac{1}{\sqrt{n}}\right)$, implying (3.16), where Φ is the cdf of $N(0,1)$. □

Remark 1. The c used in the proof of Lemma 3.1.3 is given by $c = \frac{\sqrt{p(1-p)}}{F^*(r_F(p))}$ and does not depend on n.

Remark 2. We have essentially proved that $\sqrt{n}(r_{F_n}(p) - r_F(p)) \Rightarrow N(0, c^2)$ in distribution. However that result is not required for our main theorem, so we are not emphasizing on the proof for this.

We now shift our attention to $V_{F_n}(p)$ and $V_F(p)$ and establish a bound on its rate of convergence towards $V_F(p)$.

Lemma 3.1.4 *Under A1, A2 and A3,* $V_{F_n}(p) - V_F(p) = O_P\left(\frac{1}{\sqrt{n}}\right)$

Proof :

$$
\begin{aligned}
& |V_{F_n}(p) - V_F(p)| \\
= \; & |\text{volume}\,\{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq r_{F_n}(p)\} - \text{volume}\,\{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_F(p)\}\,| \\
\leq \; & |\text{volume}\,\{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq r_{F_n}(p)\} - \text{volume}\,\{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_{F_n}(p)\}\,| \\
+ \; & |\text{volume}\,\{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_{F_n}(p)\} - \text{volume}\,\{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_F(p)\}\,| \qquad (3.21)
\end{aligned}
$$

Now by central limit theorem for independent and identical random variables $R_{F_n}(\mathbf{x}) - R_F(\mathbf{x}) = O_P\left(\frac{1}{\sqrt{n}}\right)$. Thus $||R_{F_n}(\mathbf{x})|| - ||R_F(\mathbf{x})|| = O_P\left(\frac{1}{\sqrt{n}}\right)$. Hence

$$
\begin{aligned}
& |\text{volume}\,\{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq r_{F_n}(p)\} \quad - \quad \text{volume}\,\{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_{F_n}(p)\}\,| \\
= \text{volume}\Bigg\{ & \mathbf{x} : \min(||R_{F_n}(\mathbf{x})||, ||R_F(\mathbf{x})||) \quad \leq \quad r_{F_n}(p) \leq \max(||R_{F_n}(\mathbf{x})||, ||R_F(\mathbf{x})||) \Bigg\} \\
= \text{volume}\Bigg\{ & \mathbf{x} : \min(||R_{F_n}(\mathbf{x})||, ||R_F(\mathbf{x})||) \quad \leq \quad r_{F_n}(p) \\
& \qquad\qquad\qquad\qquad\qquad\qquad \leq \quad \min(||R_{F_n}(\mathbf{x})||, ||R_F(\mathbf{x})||) + O_P\left(\frac{1}{\sqrt{n}}\right) \Bigg\} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left(\text{as } ||R_{F_n}(\mathbf{x})|| = ||R_F(\mathbf{x})|| + O_P\left(\frac{1}{\sqrt{n}}\right)\right) \\
& \qquad\qquad = O_P\left(\frac{1}{\sqrt{n}}\right) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.22)
\end{aligned}
$$

as the region is a dimension-1 or higher torus with difference of inner diameter and outer diameter of order $O_P\left(\frac{1}{\sqrt{n}}\right)$. Also

$$\begin{aligned}
& |\text{volume } \{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_{F_n}(p)\} - \text{volume } \{\mathbf{x} : ||R_F(\mathbf{x})|| \leq r_F(p)\}| \\
= & \text{ volume } \{\mathbf{x} : \min(r_{F_n}(p), r_F(p)) \leq ||R_F(\mathbf{x})|| \leq \max(r_{F_n}(p), r_F(p))\} \\
= & \text{ volume } \left\{\mathbf{x} : \min(r_{F_n}(p), r_F(p)) \leq ||R_F(\mathbf{x})|| \leq \min(r_{F_n}(p), r_F(p)) + O_P\left(\frac{1}{\sqrt{n}}\right)\right\} \\
& \text{(by } A1 \text{ and } A2) \\
= & O_P\left(\frac{1}{n^{\frac{d}{2}}}\right) \qquad (3.23)
\end{aligned}$$

since the region is a hyper-torus with difference of inner and outer diameter being $O_P\left(\frac{1}{\sqrt{n}}\right)$. Thus,

$$\begin{aligned}
V_{F_n}(p) - V_F(p) &= O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{n^{\frac{d}{2}}}\right) \\
\Rightarrow V_{F_n}(p) - V_F(p) &= O_P\left(\frac{1}{\sqrt{n}}\right). \qquad (3.24)
\end{aligned}$$

Hence the lemma is proved. □

The following result will be required to establish a useful approximation of $V_{F_n}^{-1}(V_F(p))$.

Lemma 3.1.5 *Suppose* $\{a_n\}$ *and* $\{b_n\}$ *are two sequences of real numbers such that for* $a \in \mathbb{R}$ *taking values on* $[0, \infty]$, $a_n \to a$ *and* $b_n \to a$, *as* $n \to \infty$. *Then under the following conditions:*

1. $\sup_S \left| ||R_{F_n}(\mathbf{x})|| - ||R_F(\mathbf{x})|| \right| \to 0$ *almost surely for any bounded set* $S \in \mathbb{R}^d$,
2. $P\{\mathbf{x} : ||R_F(\mathbf{x})|| = c\} = 0$ *for any* $c \in \mathbb{R}$,

we have,

$$P(C_{F_n}^{a_n}(p) \Delta C_F^{b_n}(p)) \to 0 \qquad (3.25)$$

almost surely on the set S, *where* $A\Delta B = (A \cup B) \backslash (A \cap B)$, $C_{F_n}^{a_n}(p) = \{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq a_n\}$, $C_F^{b_n}(p) = \{\mathbf{x} : ||R_F(\mathbf{x})|| \leq b_n\}$.

Proof:

$$\begin{aligned}
&\bigcap_{n=1}^{\infty}\bigcup_{N=n}^{\infty} \quad \left(C_{F_n}^{a_n}(p)\Delta C_F^{b_n}(p)\right) \\
&= \{\mathbf{x} \ : \ \mathbf{x} \in C_{F_n}^{a_n}(p)\Delta C_F^{b_n}(p) \text{ infinitely often}\} \\
&= \{\mathbf{x} \ : \ (||R_{F_n}(\mathbf{x})|| \leq a_n) \ \triangle \ (||R_F(\mathbf{x})|| \leq b_n) \text{ infinitely often}\} \\
&= \{\mathbf{x} \ : \ (||R_{F_n}(\mathbf{x})|| \leq a_n, ||R_F(\mathbf{x})|| > b_n) \\
&\qquad\qquad \cup(||R_{F_n}(\mathbf{x})|| > a_n, ||R_F(\mathbf{x})|| \leq b_n) \text{ infinitely often}\} \\
&= \{\mathbf{x} \ : \ ||R_F(\mathbf{x})|| = a\}
\end{aligned}$$

for almost all $\omega \in S$. Hence, $P(\cup_{N=n}^{\infty} C_{F_n}^{a_n}(p)\Delta C_F^{b_n}(p)) \to 0$ on S almost surely by conditions 1 and 2 of the lemma. $\square$.

Lemma 3.1.6 *Under A1-A4,*

$$V_{F_n}^{-1}(V_F(p)) - \frac{1}{n}\sum_{j=1}^{n} I(\mathbf{X}_j \in C_F(p)) = o(e^{-n}) \ w.p. \ 1. \tag{3.26}$$

Proof : Let $C_{F_n}^*(p) =$ Smallest $C_{F_n}(.)$ with volume $\geq V_F(p)$. Then

$$\begin{aligned}
&|P(\mathbf{X}_1 \in \text{ Smallest } C_{F_n} \text{ with volume } \geq V_F(p)) - P(\mathbf{X}_1 \in C_F(p))| \\
\leq \ & P(\mathbf{X}_1 \in C_{F_n}^*(p)\Delta C_F(p))
\end{aligned}$$

Now

$$\begin{aligned}
&\mathbf{X}_1 \in C_F(p) \Leftrightarrow ||R_F(\mathbf{X}_1)|| \leq r_F(p) \\
&\Leftrightarrow ||E_X\left(\frac{\mathbf{X}_1 - \mathbf{X}}{||\mathbf{X}_1 - \mathbf{X}||}\right)|| \leq r_F(p).
\end{aligned}$$

Let p_n be the smallest number such that $C_{F_n}(p_n)$ has volume $\geq V_F(p)$. Hence,

$$C_{F_n}^*(p) = C_{F_n}(p_n) = \{\mathbf{x} : ||R_{F_n}(\mathbf{x})|| \leq r_{F_n}(p_n)\}.$$

Let

$$A^*_{1n} = I(\mathbf{X}_1 \in \text{ Smallest } C_{F_n}(.) \text{ with volume } \geq V_F(p)) = I(\mathbf{X}_1 \in C_{F_n}(p_n))$$

and

$$A^{**}_1 = I(\mathbf{X}_1 \in C_F(p))$$

Now the moment generating function of $e^n(A^{**}_1 - A^*_{1n})$ is

$$\begin{aligned} & E(exp(it(A^{**}_1 - A^*_{1n})e^n)) \\ = \; & 1.P(A^{**}_1 = A^*_{1n}) + e^{ite^n}.P(A^{**}_1 - A^*_{1n} = 1) + e^{-ite^n}.P(A^{**}_1 - A^*_{1n} = -1) \\ \to \; & 1 \text{ if both } P(A^{**}_1 - A^*_{1n} = 1) \to 0, P(A^{**}_1 - A^*_{1n} = -1) \to 0 \end{aligned} \tag{3.27}$$

which would imply $e^n(A^{**}_1 - A^*_{1n}) \xrightarrow{p} 0$ and this would imply $A^{**}_1 - A^*_{1n} = o_p(e^{-n})$
It remains to show that $P(A^{**}_1 - A^*_{1n} = 1) \to 0$ and $P(A^{**}_1 - A^*_{1n} = -1) \to 0$ as $n \to \infty$.
Now

$$\begin{aligned} & P(A^{**}_1 - A^*_{1n} = 1) + P(A^{**}_1 - A^*_{1n} = -1) \\ = \; & P(C^*_{F_n}(p)\Delta C_F(p)) \\ = \; & P\left(\{||R_{F_n}(\mathbf{X}_1) \leq r_{F_n}(p_n)\} \,\Delta\, \{||R_{F_n}(\mathbf{X}_1)|| \leq r_F(p)\}\right). \end{aligned} \tag{3.28}$$

Hence, if we can show that $r_{F_n}(p_n) \xrightarrow{a.s.} r_F(p)$, then (3.27) will follow by Lemma 3.1.5.
We have already seen from Lemma 3.1.4, $V_{F_n}(p) \xrightarrow{a.s.} V_F(p)$.
Hence for a given $\epsilon > 0$, there exists N such that for all $n \geq N$,

$$V_{F_n}(p + \epsilon) \geq V_F(p) > V_{F_n}(p - \epsilon). \tag{3.29}$$

Further, as the volume of $C_{F_n}(p_n) = V_F(p)$, hence by the nesting property of $C_{F_n}(p)$ we have

$$C_{F_n}(p - \epsilon) \subseteq C_{F_n}(p_n) \subseteq C_{F_n}(p + \epsilon). \tag{3.30}$$

Thus,

$$p + \epsilon \geq p_n > p - \epsilon, \text{ a.s.}$$

So,

$$\begin{aligned} r_{F_n}(p+\epsilon) &\geq r_{F_n}(p_n) > r_{F_n}(p) \text{ for } n \geq N, \text{a.s.} \\ \Rightarrow r_{F_n}(p_n) &- r_{F_n}(p) \xrightarrow{a.s.} 0 \end{aligned}$$

As from Lemma 3.1.3, $r_{F_n}(p) - r_F(p) \xrightarrow{a.s.} 0$, we can say

$$r_{F_n}(p_n) \xrightarrow{a.s.} r_F(p)$$

and hence,

$$P(C^*_{F_n}(p) \Delta C_F(p)) \to 0$$

which proves (3.27). So,

$$\begin{aligned} I(\mathbf{X}_1 \in C_{F_n}(p_n)) &- I(\mathbf{X}_1 \in C_F(p)) = o(e^{-n}) \text{ w.p. } 1 \\ \frac{1}{n}\sum_{j=1}^{n} I(\mathbf{X}_j \in C_{F_n}(p_n)) &- \frac{1}{n}\sum_{j=1}^{n} I(\mathbf{X}_j \in C_F(p)) = o(e^{-n}) \text{ w.p. } 1 \end{aligned}$$

where $V_{F_n}^{-1}(V_F(p)) = \frac{1}{n}\sum_{j=1}^{n} I(\mathbf{X}_j \in C_{F_n}(p_n))$. Thus the lemma is proved. □

We now aim to prove a result which establishes a relation between $V_{F_n}^{-1}(x) - V_{F_n}^{-1}(V_F(p))$ and $V_F^{-1}(x) - V_F^{-1}(V_F(p))$ for x's close to $V_F(p)$. For this we will use the following result by Bernstein.

Fact 3.1.1 *For any n and any z, $0 \leq z \leq 1$, let $B(n,z)$ denote a binomial random variable. Then*

$$P(|B(n,z) - nz| \geq t) \leq 2e^{-h}, \tag{3.31}$$

for all $t > 0$*, where*

$$h = h(n, z, t) = \frac{t^2}{2[nz(1-z) + \frac{t}{3}\max\{z, 1-z\}]}. \tag{3.32}$$

The proof is omitted and can be found in Uspensky (1937).

Consider the following setup. Let us define

$$G_n(x) = \left[V_{F_n}^{-1}(x) - V_{F_n}^{-1}(V_F(p))\right] - \left[V_F^{-1}(x) - V_F^{-1}(V_F(p))\right] \tag{3.33}$$

Now let $\{a_n : n = 1, 2, \cdots\}$ be a sequence of positive constants such that

$$a_n \sim \frac{\log n}{n^{\frac{1}{2}}}, \text{ as } n \to \infty. \tag{3.34}$$

Let

$$I_n = (V_F(p) - a_n, V_F(p) + a_n) \tag{3.35}$$

and define $H_n = \sup_{x \in I_n} |G_n(x)|$.

The following lemma is similar to a lemma proved by Bahadur (1966).

Lemma 3.1.7 *Under A1-A5,*

$$H_n = O_P(n^{-\frac{3}{4}} \log n)$$

Proof : Let

$$b_n \sim n^{\frac{1}{4}}, \tag{3.36}$$

as $n = 1, 2, \cdots$. For any integer r, let

$$\eta_{r,n} = V_F(p) + a_n b_n^{-1} r.$$

Further, let

$$J_{r,n} = [\eta_{r,n}, \eta_{r+1,n}]$$

and define

$$\alpha_{r,n} = V_F^{-1}(\eta_{r+1,n}) - V_F^{-1}(\eta_{r,n}).$$

$V_{F_n}^{-1}$ and V_F^{-1} are clearly non-decreasing for $x \in J_{r,n}$. So,

$$\begin{aligned} G_n(x) &\leq V_{F_n}^{-1}(\eta_{r+1,n}) - V_{F_n}^{-1}(V_F(p)) - V_F^{-1}(\eta_{r,n}) + p \\ &= G_n(\eta_{r+1,n}) + \alpha_{r,n}. \end{aligned}$$

Also, for $x \in J_{r,n}$, $G_n(x) \geq G_n(\eta_{r,n}) - \alpha_{r,n}$. Hence,

$$\begin{aligned} H_n &\leq \max\{|G_n(\eta_{r,n})| : -b_n \leq r \leq b_n\} + \max\{\alpha_{r,n} : -b_n \leq r \leq b_n - 1\} \\ &= K_n + \beta_n \text{ (say)}. \end{aligned} \tag{3.37}$$

Since $\eta_{r+1,n} - \eta_{r,n} = a_n b_n^{-1} r$ for each r, since $|\eta_{r,n} - V_F(p)| \leq a_n$ for $|r| \leq b_n$ and since V_F^{-1} is sufficiently smooth in a bounded neighbourhood of $V_F(p)$, hence from (3.34) and (3.36), $\beta_n = O(n^{-\frac{3}{4}} \log n)$.

Let us now choose a suitable $c_1 > 0$ (whose choice is to be clarified later) and define $\nu_n = c_1 n^{-\frac{3}{4}} \log n$, for $n = 1, 2, \cdots$. We now show that $\sum_n P(K_n \geq \nu_n) < \infty$, which implies by Borel-Cantelli lemma that $P(\limsup_{n\to\infty}(K_n \geq \nu_n)) = 0$, which would in turn imply that $K_n = O_P(\nu_n)$ and this in turn proves $H_n = O_P(n^{-\frac{3}{4}} \log n)$. Hence, it remains to show that

$$\sum_n P(K_n \geq \nu_n) < \infty. \tag{3.38}$$

Now,

$$\begin{aligned} K_n &= \max\{|G_n(\eta_{r,n})| : -b_n \leq r \leq b_n\}, \\ \eta_{r,n} &= V_F(p) + a_n b_n^{-1} r, \\ \nu_n \sim n^{-\frac{3}{4}} \log n, \quad b_n &\sim n^{\frac{1}{4}}, \quad a_n \sim n^{-\frac{1}{2}} \log n. \end{aligned}$$

Choose and fix c_2 such that

$$c_2 > \frac{d}{dy} V_F^{-1}(y)\big|_{y=V_F(p)} = \frac{1}{v_F(p)}. \tag{3.39}$$

Let N be an integer so large such that

$$V_F^{-1}(V_F(p) + a_n) - V_F^{-1}(V_F(p)) < c_2 a_n$$

and

$$V_F^{-1}(V_F(p)) - V_F^{-1}(V_F(p) - a_n) < c_2 a_n$$

for all $n > N$. Now the distribution of $|G_n(\eta_{r,n})|$ is that of $n^{-1}|B(n,z) - nz|$, where $z = |V_F^{-1}(\eta_{r,n} - p)| = z_{r,n}$ (say).

$$\begin{aligned}
|G_n(\eta_{r,n})| &= |V_{F_n}^{-1}(\eta_{r,n}) - V_{F_n}^{-1}(V_F(p)) - V_F^{-1}(\eta_{r,n}) + p| \\
&= \left| \frac{1}{n} \sum_{j=1}^{n} I(X_j \in C_F(p) \Delta C_F(p_n^*)) - |p_n^* - p| \right| + o(1) \text{ (by Fact 3.1.1)} \\
&= \left| \frac{1}{n} Bin(n, |p_n^* - p|) - |p_n^* - p| \right| \\
&= \left| \frac{1}{n} Bin(n, z) - z \right|
\end{aligned}$$

where $p_n^* = V_F^{-1}(\eta_{r,n})$.

Consequently by (3.31),

$$P(|G_n(\eta_{r,n})| \geq \nu_n) \leq 2e^{-h_n(r)} \tag{3.40}$$

where $h_n(r) = h(n, z_n, n\nu_n)$ is given by (3.32). Since

$$h(n, z, t) \geq \frac{t^2}{2(nz + t)},$$

and since $n > N$ and $|r| \leq b_n$ imply $z_n \leq c_2 a_n$, it follows that

$$P(|G_n(\eta_{r,n})| \geq \nu_n) \leq 2e^{-\delta_n} \tag{3.41}$$

for $n > N$ and $|r| < b_n$, where $\delta_n = \frac{n^2\nu_n^2}{2(c_2 n a_n + n\nu_n)}$. Since δ_n does not depend on r, it follows from (3.37) and (3.41), using an argument similar to Bahadur (1966), that

$$P(K_n \geq \nu_n) \leq 4b_n e^{-\delta_n} = \lambda_n(say),$$

for $n > N$. It follows from (3.34) and (3.36) by definitions of ν_n, δ_n and λ_n that

$$\frac{\log \lambda_n}{\log n} \to \frac{1}{4} - \frac{c_1^2}{2c_2} \tag{3.42}$$

as $n \to \infty$. The limit in (3.42) is less than -1 if, given c_2, c_1 is chosen sufficiently large. Then $\sum_n \lambda_n < \infty$ and (3.38) holds. Hence the proof is complete. □

We are now ready to prove Theorem 3.1.1.

Proof of Theorem 3.1.1: We will prove the following Bahadur type representation

$$\sqrt{n}\left\{v_F(p)\right\}^{-1}\left(V_{F_n}(p) - V_F(p)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(p - I(\mathbf{X}_i \in C_F(p))\right) + R_n \tag{3.43}$$

as $n \to \infty$, where $R_n = o_p(1)$ and hence the result will follow as $\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(p - I(\mathbf{X}_i \in C_F(p))\right), 0 < p < 1\}$ converges to a Brownian bridge with covariance kernel $\Gamma(p_1, p_2) = p_1(1-p_2)$ where $0 < p_1 \leq p_2 < 1$.

Now let us fix a p, $0 < p < 1$. From Lemma 3.1.4 we have

$$V_{F_n}(p) - V_F(p) = o_P\left(\frac{1}{n^{\frac{1}{4}}}\right) \tag{3.44}$$

Considering the Taylor series expansion of V_F^{-1}, we get,

$$\begin{aligned} V_F^{-1}(V_{F_n}(p)) &= V_F^{-1}(V_F(p)) + (V_{F_n}(p) - V_F(p)) \{v_F(p)\}^{-1} + E_n \\ \Rightarrow \{v_F(p)\}^{-1} (V_{F_n}(p) - V_F(p)) &= V_F^{-1}(V_{F_n}(p)) - V_F^{-1}(V_F(p)) + E_n \end{aligned} \tag{3.45}$$

where E_n is the residual of the Taylor series expansion. E_n is of the form

$$E_n = \frac{d^2}{dy^2} \left((V_F^{-1}(y)) \right|_{y=c_n} \left(\frac{(V_{F_n}(p) - V_F(p))^2}{2!} \right), \tag{3.46}$$

where c_n is a number between $V_{F_n}(p)$ and $V_F(p)$. From (3.44) we can say,

$$\sqrt{n} \left(\frac{(V_{F_n}(p) - V_F(p))^2}{2!} \right) = o_P(1). \tag{3.47}$$

We now need to see whether

$$\frac{d^2}{dy^2} \left((V_F^{-1}(y)) \right|_{y=c_n}$$

is bounded. Let $y = V_F(p)$. Then $\frac{dy}{dp} = v_F(p)$. We assume that v_F' also exists in a neighbourhood of p. Note that,

$$\begin{aligned} \frac{d}{dy} V_F^{-1}(y) &= \frac{1}{v_F(V_F^{-1}(y))}. \\ \frac{d^2}{dy^2} V_F^{-1}(y) &= -\frac{1}{v_F^2(V_F^{-1}(y))} \frac{d}{dy} v_F(V_F^{-1}(y)) \\ &= -\frac{v_F'(V_F^{-1}(y))}{v_F^3(V_F^{-1}(y))} \end{aligned} \tag{3.48}$$

Thus,

$$\sqrt{n} E_n = -\frac{v_F'(V_F^{-1}(c_n))}{v_F^3(V_F^{-1}(c_n))} \sqrt{n} \left(\frac{(V_{F_n}(p) - V_F(p))^2}{2!} \right) \tag{3.49}$$

We now show that $\sqrt{n} E_n \xrightarrow{a.s.} 0$. Since $V_{F_n}(p)$ converges to $V_F(p)$ in probability and c_n is a sequence of numbers between them, hence $c_n \to V_F(p)$ in probability. Thus $V_F^{-1}(c_n) \to p$ in probability and hence $v_F'(V_F^{-1}(c_n))$ is bounded in the neighbourhood of p because of $A5$. Hence (3.49) combined with (3.44) proves $\sqrt{n} E_n \xrightarrow{a.s.} 0$.

As $V_{F_n}(p) - V_F(p) = o_P\left(\frac{1}{\sqrt{n}}\right)$ by Lemma 3.1.4 and $a_n \sim \frac{\log n}{\sqrt{n}}$ as defined in (3.34), eventually $V_{F_n}(p)$ is in I_n, for all $n \geq N$ for large N with probability 1, where I_n is defined in (3.35). Hence by Lemma 3.1.7 we can say that $\sup_{x \in I_n} \sqrt{n} G_n(x) = o_P(1)$, which implies that

$$\sqrt{n}(V_F^{-1}(V_{F_n}(p) - V_F^{-1}(V_F(p))) - \sqrt{n}(p - V_{F_n}^{-1}(V_F(p))) = o_P(1). \tag{3.50}$$

Now,

$$V_{F_n}^{-1}(V_F(p)) = \frac{1}{n} \sum_{j=1}^{n} I\left\{X_j \in \text{Smallest } C_{F_n}(.) \text{ with volume } \geq V_F(p)\right\}. \tag{3.51}$$

So from (3.50) and (3.51) we get

$$p - V_{F_n}^{-1}(V_F(p)) = p - \frac{1}{n} \sum_{j=1}^{n} I\left\{X_j \in \text{Smallest } C_{F_n}(.) \text{ with volume } \geq V_F(p)\right\}. \tag{3.52}$$

Now we have to take care of the difference between $\frac{1}{n}\sum_{j=1}^{n} I\left\{X_j \in C_F(p)\right\}$ and $\frac{1}{n}\sum_{j=1}^{n} I\left\{X_j \in \text{Smallest } C_{F_n}(.) \text{ with volume } \geq V_F(p)\right\}$. From Lemma 3.1.6 we get $V_{F_n}^{-1}(V_F(p)) - \frac{1}{n}\sum_{j=1}^{n} I(\mathbf{X}_j \in C_F(p)) = o_p(e^{-n})$. Thus we can replace $V_{F_n}^{-1}(V_F(p))$ in (3.52) by $\frac{1}{\sqrt{n}} I(\mathbf{X}_i \in C_F(p))$. Hence we have our result. □

The construction of our rank functions $R_F(\mathbf{x})$ and $R_{F_n}(\mathbf{x})$ are based on the transformation retransformation matrix $\{\mathbf{X}(\alpha)\}$ as discussed in Definition 2.2.1 of Chapter 2. $\{\mathbf{X}(\alpha)\}\{\mathbf{X}(\alpha)\}^T$ is chosen in such a way that it becomes a consistent estimate of the covariance matrix Σ. Without loss of generality, we may take Σ as the identity matrix I; then we arrive at the forms of the rank functions as we have taken them in this section.

3.2 The Scaling Factor $v_F(p)$

In the previous section we have seen that $\{V_{F_n}, 0 < p < 1\}$ converges to a Brownian bridge, where $v_F(p)$ appears as a scaling factor. Clearly to find a confidence interval for $V_F(p)$, $v_F(p)$ needs to be computed or estimated. We now state some results that give us some idea about the form of $v_F(p)$ in some special cases. In general $v_F(p)$ may be quite hard to obtain.

Theorem 3.2.1 *Suppose* $\mathbf{X}$ *is a d-dimensional random vector with a distribution function* F*, which is elliptically symmetric about* $\boldsymbol{\theta}$ *with the scale matrix* Σ*. Then,*

$$v_F(p) = \frac{d\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}+1) f(r_F(p))} \tag{3.53}$$

where $r_F(p)$ *is the p-th quantile of the distribution of* $\sqrt{(\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta})}$*, and* f *is the probability density function of* $\sqrt{(\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta})}$*.*

Proof : For a d-dimensional elliptically symmetric distribution with distribution function F,

$$V_F(p) = \frac{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} r_F^{d}(p)}{\Gamma(\frac{d}{2}+1)}. \tag{3.54}$$

Then

$$\begin{aligned} v_F(p) &= \frac{d}{dp}V_F(p) \\ &= \frac{d\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}+1)} \left(\frac{d}{dp} r_F(p)\right) \\ &= \frac{d\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}+1) f(r_F(p))} \\ &= \frac{2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}) f(r_F(p))} \end{aligned}$$

where $f(.)$ is the density function of $\sqrt{(\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta})}$. □

In the spherically symmetric case, $\Sigma = I_d$ and we have

$$\begin{aligned} v_F(p) &= \frac{d\pi^{\frac{d}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}+1) f(r_F(p))} \\ &= \frac{2\pi^{\frac{d}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}) f(r_F(p))} \end{aligned}$$

Now let us consider a few special spherically and elliptically symmetric cases, the multivariate normal distribution, the multivariate Laplace distribution and the multivariate t distribution.

Examples :

1. For a d-dimensional multivariate normal distribution $N_d(\boldsymbol{\mu}, \Sigma)$ with distribution function F,

$$v_F(p) = \frac{d\pi^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} r_F^{d-1}(p)}{\Gamma(\frac{d}{2}+1) f(r_F(p))} \tag{3.55}$$

   where $r_F^2(p)$ is the p-th quantile of a χ_d^2 random variable and $f(.)$ is the density function of $\sqrt{\chi_d^2}$ where

$$f(x) = \frac{1}{2^{\frac{d}{2}-1}\Gamma(\frac{d}{2})} x^{d-1} exp\left(-\frac{x^2}{2}\right).$$

   Hence

$$v_F(p) = (2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} exp\left(\frac{r_F^2(p)}{2}\right). \tag{3.56}$$

2. For a d-dimensional multivariate Laplace distribution with distribution function F, $v_F(p)$ is given by (3.55) where $r_F(p)$ is the p-th quantile of $||\mathbf{X}||$ and $f(.)$ is the density function of $\Gamma(d, 1)$ where

$$f(x) = \frac{x^{d-1} exp(-x)}{\Gamma(d)}.$$

   Hence

$$v_F(p) = \frac{2\pi^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \Gamma(d)}{\Gamma\left(\frac{d}{2}\right)} exp(r_F(p)).$$

3. For a d-dimensional multivariate t distribution with ν degrees of freedom and distribution function F, $v_F(p)$ is given by (3.55) where $r_F^2(p)$ is the p-th quantile of $dF_{d,\nu}$ random variable and $f(.)$ is the density function of $\sqrt{dF_{d,\nu}}$ where

$$f(x) = \frac{2\Gamma\left(\frac{d+\nu}{2}\right)}{d\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} x^{d-1}\left(1+\frac{1}{\nu}x^2\right)^{-\frac{(d+\nu)}{2}} . \left(\frac{d}{\nu}\right)^{\frac{d}{2}}.$$

Hence

$$v_F(p) = \frac{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)\nu^{\frac{d}{2}}\left(1+\frac{1}{\nu}r_F^2(p)\right)^{\frac{d+\nu}{2}}}{\Gamma\left(\frac{d+\nu}{2}\right)d^{\frac{d}{2}-1}}.$$

We note here that in all of the above examples $v_F(p) > 0$ for all $0 < p < 1$ and is bounded away from 0 for p in any compact subset of $[0, 1]$. Moreover in the above three examples $r_F(p)$ and hence $v_F(p)$ is a differentiable function. This clearly indicates that assumptions A4 and A5 hold for all of the above examples.

In the above examples we know the form of $v_F(p)$ exactly and we can just use a consistent estimator of Σ to get a consistent estimator of $v_F(p)$. However, if the form of $f(.)$ is unknown, we need to estimate $f(.)$ by some density estimation method, for example, kernel density estimation method or spline fitting, which may affect the accuracy of our estimation procedure. It may be noted here that the accuracy in estimation of $v_F(p)$ affects the accuracy in estimating $V_F(p)$ by $V_{F_n}(p)$. Especially when a nonparametric density estimator is employed in estimating $v_F(p)$, the $\sqrt{n}$ rate of convergence that was established in Theorem 3.1.1 will not be achieved.

CHAPTER 4

VISUAL TESTS BASED ON THE SCALE-SCALE PLOTS

One of the widely used graphical method for comparing univariate distributions is the quantile-quantile (Q-Q) plot that matches the quantiles of one distribution with the same quantiles of the other. The Q-Q plots, which were proposed by Wilk and Gnanadesikan (1968), are quite useful in revealing location and scale differences as well as identifying outliers. Though there is an extensive literature on univariate Q-Q plots, see Barnett (1976), Cook and Weisberg (1982), Cleveland (1993), Marden (2004), for some detailed discussions and examples, there are very few proposed generalisations to the multivariate distributions. Most of the multivariate procedures are based on dimension reduction techniques and are used to compare some specific reference distributions, such as the multivariate normal distribution and depend on their properties.

Healy (1968) used squared Mahalanobis distances of the observations from the sample mean vector to assess multivariate normality. These squared distances are approximately distributed as chi-squared random variables and a Q-Q plot can be constructed to assess that. Andrews, Gnanadesikan, and Warner (1973) proposed Q-Q plots based on the directions and the magnitude of the observations. The magnitudes or equivalently the squared distances are approximately distributed as chi-squared random variables as before and the angles obtained from the direction vectors are uniformly distributed.

There are some other multivariate Q-Q plotting techniques available in the literature,

which are based on assessing the commonality of the shape of the marginal distributions for certain commonly used multivariate distributions. For a detailed discussion on graphical methods for assessing multivariate distributional shape, see chapters 5 and 6 of Gnanadesikan (1977).

In a completely different approach, Easton and McCulloch (1990) proposed a generalisation of multivariate Q-Q plots based on matching the data with simulated observations from the reference distribution. For a d-dimensional data set $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the procedure is to find a permutation σ^* of $\{1, 2, \cdots, n\}$ a $d \times d$ matrix $\mathbf{A}$ and a $d \times 1$ vector $\mathbf{b}$ that solves

$$\min_{\mathbf{A},\mathbf{b},\sigma} \sum_{i=1}^{n} ||\mathbf{Y}_i - \mathbf{A}\mathbf{X}_{\sigma(i)} - \mathbf{b}||^2$$

where $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ is a random sample from the reference distribution. Suppose the $\mathbf{X}_i^*$'s are the best matching of the data set to the reference sample where $\mathbf{X}_i^* = \mathbf{A}^*\mathbf{X}_{\sigma^*(i)} + \mathbf{b}^*$. They suggested to display the matched pairs $(\mathbf{X}_i^*, \mathbf{Y}_i)$ using either coordinatewise Q-Q plots or some distance based Q-Q plots. One of the main problem in this approach is that it cannot be used to compare two multivariate samples. Visualising co-ordinatewise Q-Q plots can be difficult if the dimension d is large.

Singh, Tyler, Zhang and Mukherjee (2009) defined quantile scale curves to perform visual tests. Again consider the d-dimensional random sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ and denote by $\triangle_i$ the volume of simplex formed by $(d+1)$ points $\mathbf{X}_1, \cdots, \mathbf{X}_{d+1}$ for $i = 1, \cdots, N$ and $N = \binom{n}{d+1}$. Then their quantile scale curve is defined as $qsc(t) = \triangle_{([Nt])}$ for $0 \leq t \leq 1$.

In an approach based on the spatial quantiles, Marden (1998) proposed a bivariate Q-Q plot by drawing arrows from (bivariate) normal quantiles corresponding to bivariate ranks of the observations to the actual values of the corresponding observations as a check for normality. Chakraborty (2001) pointed out that Marden's plots are not affine invariant and hence might lead to erroneous inference for highly correlated data. Chakraborty (2001) suggested a modification based on a transformation retransformation (TR) ap-

proach to obtain a spherical transformation of the data based on which an affine invariant Q-Q plot can be constructed. He also proposed a version of quantile contours $\{\hat{\mathbf{Q}}_n^{(\alpha,p)}(\mathbf{u}) : ||\mathbf{u}||_q = r\}$ for $0 < r < 1$, where $\hat{\mathbf{Q}}_n^{(\alpha,p)}(\mathbf{u})$ is the $\mathbf{u}$-th multivariate TR l_p quantile based on a data-driven co-ordinate system $\{\mathbf{X}_j : j \in \alpha\}$, which were affine equivariant. In a recent work Dhar, Chakraborty and Chaudhuri (2011) proposed a method to construct a Q-Q plot for a d-dimensional multivariate dataset as a collection of d two dimensional plots, each plot corresponding to a component of the multivariate empirical spatial quantile of the data. They also proposed a test based on the norms of the spatial quantiles for comparing multivariate distributions. Their proposed test statistic is related to the arrow lengths of the arrow plots proposed by Marden (1998) described earlier.

In the rest of this chapter we explore visual tests based on our proposed scale-scale curves. In Section 4.1 we suggest a test for the one sample case and in Section 4.2 we suggest a test for the two sample case.

4.1 One Sample Problem

The proposed scale-scale plot can be used to check the distributional assumptions for the multivariate data visually as an extension of the univariate quantile-quantile plot. If F_0 is the hypothesised distribution function up to a location parameter and F_n denotes the empirical distribution function of the d dimensional multivariate data $\mathbf{X}_1, \ldots, \mathbf{X}_n$, we can construct a scale-scale plot by plotting $V_{F_n}(p)$ against $V_{F_0}(p)$. If the plot is close to a straight line, then we can say that there is not enough evidence that F_0 can not be considered as the true distribution up to location and scale parameters.

For illustration, we use the Iris data in Example 1 and check for multivariate normality of the underlying distribution of the three separate species Iris Setosa, Iris Versicolor and Iris Virginica. This data can be found at http://archive.ics.uci.edu/ml.

Example 1. The Iris data consists of three species, namely Setosa, Versicolor and Vir-

ginica. The dataset consists of measurements in centimetres of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of Iris. We considered the sepal length and sepal width for the Setosa species, computed the volume of the central rank region and plotted it against the volume obtained from a standard bivariate normal distribution. From Figure 4.1(a) we can see that the Scale-Scale plot is nearly linear, which suggests normality for the underlying distribution may be a valid assumption. Similarly we have taken Versicolor and Virginica species and performed the same calculations to get Figure 4.1(b) and (c) respectively from which we can see that these two plot are also nearly linear. Hence we can make the same conclusion for Versicolor and Virginica species as the case to Setosa species.

We can also construct formal tests based on the slope functional $s_n(p) = \frac{V_{F_n}(p)}{V_{F_0}(p)}$. Under the null hypothesis, $H_0 : F = F_0$, $s_n(p) = \frac{V_F(p)}{V_{F_0}(p)}$ is identically equal to 1 and we can use the following result for any test statistic based on s_n.

Theorem 4.1.1 *Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be a random sample from d-dimensional distribution F. Consider some d-dimensional distribution F_0 and the slope functional $s_n(p) = \frac{V_{F_n}(p)}{V_{F_0}(p)}$. Then*

$$\left\{ \sqrt{n} \left(\frac{v_F(p)}{V_{F_0}(p)} \right)^{-1} \left(s_n(p) - \frac{V_F(p)}{V_{F_0}(p)} \right), 0 < p < 1 \right\}$$

converges to a Brownian Bridge with covariance kernel

$$\Gamma(p_1, p_2) = \begin{cases} p(1-p), & \text{if } p_1 = p_2 = p \\ \min\{p_1, p_2\} - p_1 p_2, & \text{if } p_1 \neq p_2 \end{cases}$$

as $n \to \infty$.

Hence under $H_0 : F = F_0$,

$$\left\{ \sqrt{n} \left(\frac{v_{F_0}(p)}{V_{F_0}(p)} \right)^{-1} (s_n(p) - 1), 0 < p < 1 \right\}$$

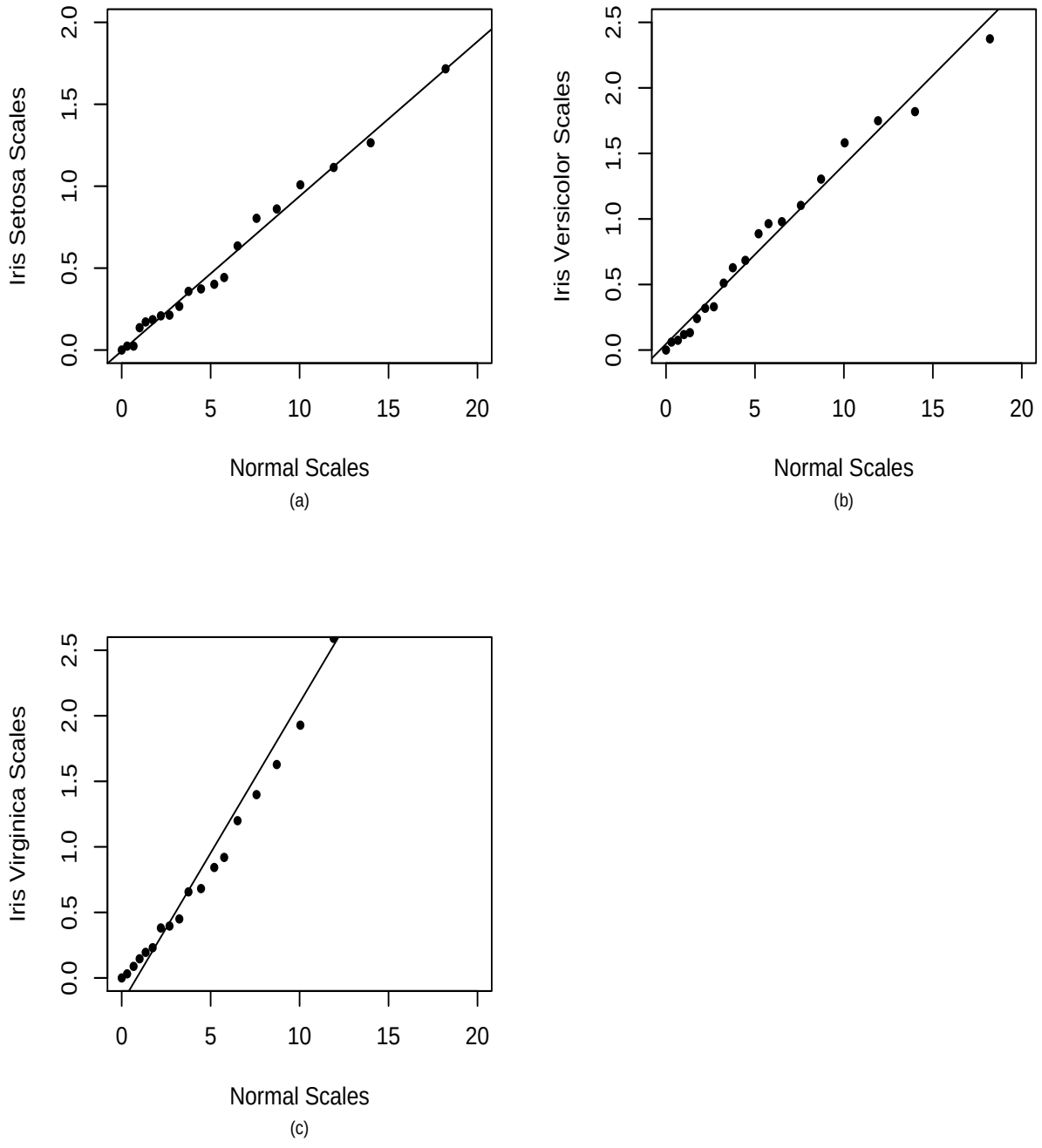

Figure 4.1: Scale-Scale plot for(a) Setosa Species of Iris data, (b) Versicolor Species of Iris data (c) Virginica Species of Iris data

converges to a Brownian Bridge with covariance kernel

$$\Gamma(p_1, p_2) = \begin{cases} p(1-p), & \text{if } p_1 = p_2 = p \\ \min\{p_1, p_2\} - p_1 p_2, & \text{if } p_1 \neq p_2 \end{cases}$$

as $n \to \infty$.

Proof : From Theorem 3.1.1 we have $\left\{\sqrt{n}\left\{v_F(p)\right\}^{-1}\left(V_{F_n}(p) - V_F(p)\right), 0 < p < 1\right\}$ converges to a Brownian bridge. Thus from this result and from the definition of the slope functional $s_n(p)$, we have our result by dividing the numerator and denominator by $V_{F_0}(p)$.

Also as under $H_0 : F = F_0$, we have $V_F(p) = V_{F_0}(p)$. Hence we have our result. □

When F_0 is standard multivariate normal distribution, one can simplify the above result to the following:

Corollary 4.1.1 *Suppose* F_0 *is a standard multivariate normal distribution function, then under* H_0,

$$\left\{\sqrt{n}\left(\frac{r_{F_0}(p) f(r_{F_0}(p))}{d}\right)\left(s_n(p) - 1\right), 0 < p < 1\right\}$$

converges to a Brownian Bridge as $n \to \infty$ *with covariance kernel*

$$\Gamma(p_1, p_2) = \begin{cases} p(1-p), & \text{if } p_1 = p_2 = p \\ \min\{p_1, p_2\} - p_1 p_2, & \text{if } p_1 \neq p_2 \end{cases}$$

and $f(.)$ *is the density function of* $\sqrt{\chi^2_d}$ *and* $r^2_{F_0}(p)$ *is the* p*-th quantile of* χ^2_d *distribution.*

Proof : From Theorem 4.1.1, under $H_0 : F = F_0$ we have the following result: $\left\{\sqrt{n}\left(\frac{v_{F_0}(p)}{V_{F_0}(p)}\right)^{-1}\left(s_n(p) - 1\right), 0 < p < 1\right\}$ converges to a Brownian Bridge. We know $V_F(p) = \frac{\pi^{\frac{d}{2}} r_F^d(p)}{\Gamma\left(\frac{d}{2}+1\right)}$. When F_0 is a standard multivariate normal distribution, then from (3.55) we have an expression for $v_F(p)$. Thus we have, $\left(\frac{v_{F_0}(p)}{V_{F_0}(p)}\right)^{-1} = \frac{r_{F_0}(p) f(r_{F_0}(p))}{d}$ and hence we have our result. □

One can define several statistics based on $s_n(p)$ to formally test for the distributions. Some suggested tests are discussed in detail in the next chapter.

4.2 Two Sample Problem

We can also use the scale-scale plot to compare two multivariate samples. Suppose F_n and G_n are the empirical distribution functions of the two independent samples, respectively. We plot $V_{G_n}(p)$ against $V_{F_n}(p)$ for $0 < p < 1$ to construct the scale-scale plot of the two samples.

To illustrate, we give the scale-scale plot for Setosa species versus Virginica species, Setosa species versus Versicolor species and for Versicolor species versus Virginica species of Iris data in Example 2. In Example 3 we take the open and closed book data set from Mardia, Kent and Bibby (1979) and plot scale-scale plot for open and closed book data set. In Example 4 we consider data on Alaskan and Canadian salmons (See Table 11.2, Johnson and Wichern (2002)).

Example 2. Here we have taken the data of sepal length and sepal width for Virginica species of Iris data. We have computed the volume of the central rank region for the data set and then have plotted it against the volumes of Setosa species in Figure 4.2(a) and Versicolor species in Figure 4.2(b) respectively. In Figure 4.2(c) we have plotted the scale-scale plot for Setosa species and Versicolor species of Iris dataset. All the three plots of Figure 4.2 appear to be almost linear. So we can say that there is not enough evidence against the null hypothesis that the distribution of the data set for Virginica species does not come from different distribution family as that of the Setosa species and the Versicolor species.

Example 3. In this example we consider a data set from Mardia, Kent and Bibby (1979). This data set has examination marks of 88 students in an open book test and a closed book test. The open book tests were taken on algebra and analysis and closed book tests

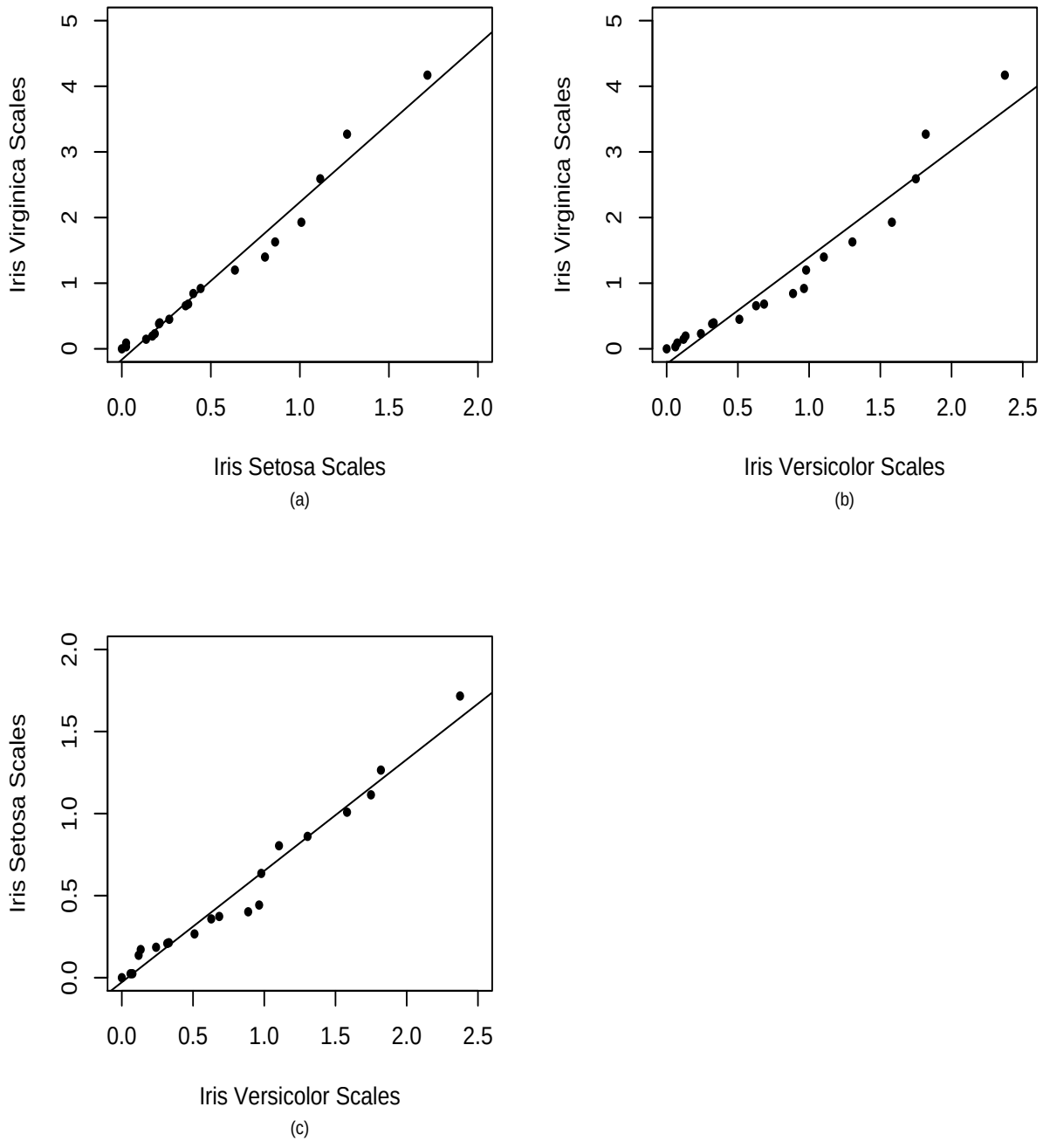

Figure 4.2: (a)Scale-scale plot for the Setosa species and the Virginica species of Iris data. (b)Scale-scale plot for the Versicolor species and the Virginica species of Iris data. (c) Scale-Scale plot for the Setosa species and the Versicolor species of Iris data.

were taken on mechanics and vectors. The total marks were out of 100.

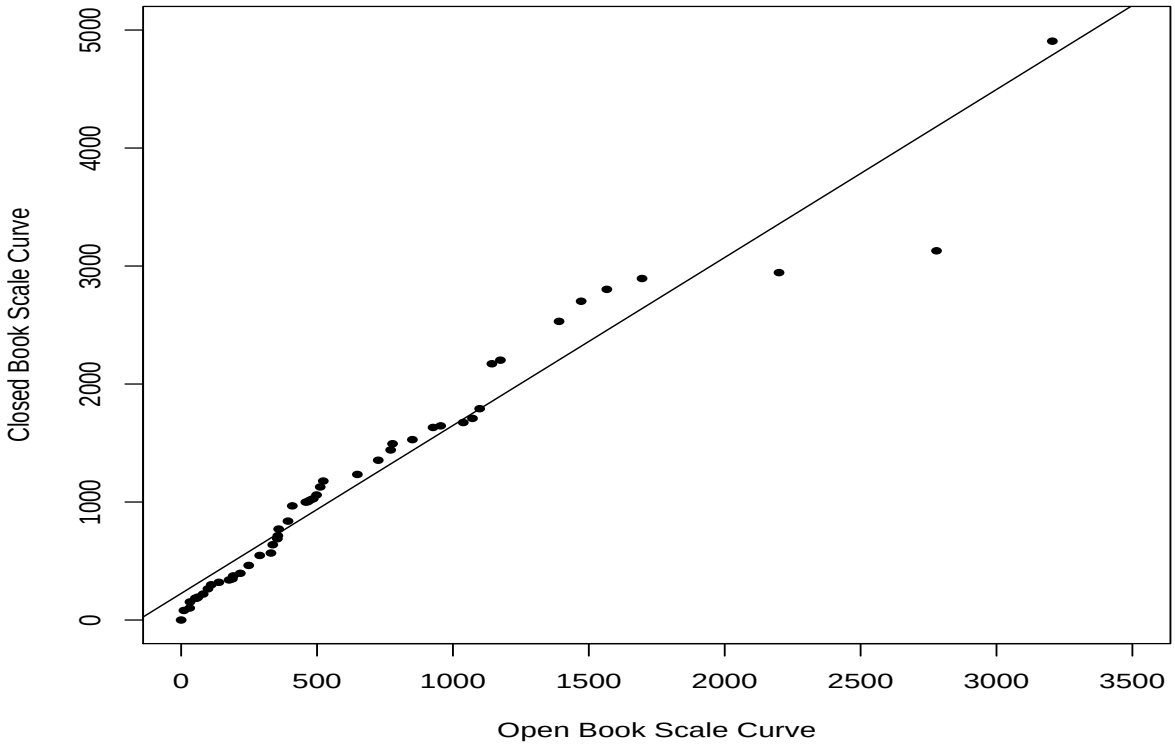

Figure 4.3: Scale-Scale plot for open and closed book examination marks

Here the number of observations is 88 and we have constructed two bivariate data sets, one consisting of open book marks of a student and the other consisting of closed book marks of a student. We have computed the volume of the central rank region for the two bivariate data set and then have plotted them against each other in Figure 4.3. We can see that the plot is nearly linear. So we can say that there is not enough evidence that the open book marks and the closed book marks are from different distribution family.

Example 4. The measurements were taken on diameter of rings for the first-year fresh-water growth and that for the first-year marine growth for Alaskan and Canadian salmons (See Table 11.2, Johnson and Wichern (2002)). Sample sizes are 50 for both Alaskan-born and Canadian born salmons. This is a nice example on classification techniques used in

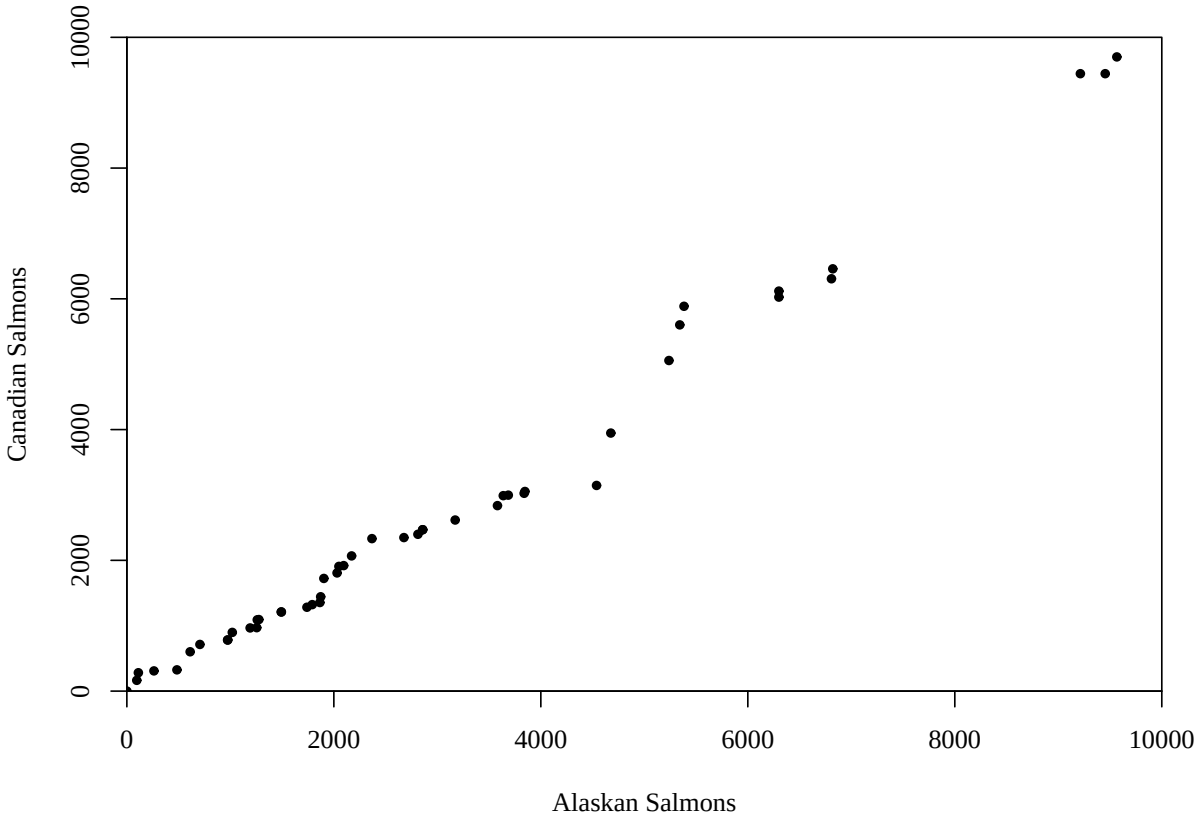

Figure 4.4: Scale-Scale plot comparing the distributions of Alaskan and Canadian Salmons.

the literature under the assumption of multivariate normality with the same covariance matrix. To justify, we may use our proposed scale-scale plot in Figure 4.4, which is almost on a 45° line. That suggests that the data distribution for both groups of salmons is the same elliptically symmetric distribution with the same scale matrix and possibly with a different location vector.

We now state a theorem which can be used to construct formal tests to compare the underlying distributions of two multivariate samples.

Theorem 4.2.1 *Let us suppose that* $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ *and* $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_m$ *are two d-dimensional random samples with distribution functions* F *and* G *respectively, symmetric around zero.*

Under $H_0 : F = G$, *assuming* $\frac{n}{m+n} \to \lambda$, *for some* $0 < \lambda < 1$, *as* $n, m \to \infty$

$$\left\{ \{v_F(p)\}^{-1} \left(\sqrt{\frac{nm}{n+m}} (V_{F_n}(p) - V_{G_m}(p)) \right), 0 < p < 1 \right\}$$

converges to a Brownian Bridge with covariance kernel

$$\Gamma(p_1, p_2) = \begin{cases} p(1-p), & \text{if } p_1 = p_2 = p \\ \min\{p_1, p_2\} - p_1 p_2, & \text{if } p_1 \neq p_2 \end{cases}$$

where F_n *and* G_m *are corresponding empirical distribution functions.*

Proof : From Theorem 3.1.1 we have,

$$\sqrt{n} \{v_F(p)\}^{-1} (V_{F_n}(p) - V_F(p)) \Rightarrow N(0, p(1-p))$$

and

$$\sqrt{m} \{v_G(p)\}^{-1} (V_{G_m}(p) - V_G(p)) \Rightarrow N(0, p(1-p)).$$

Let us take some constant λ, $(0 < \lambda < 1)$, such that $\frac{n}{(n+m)} \to \lambda$. Then $\frac{m}{(n+m)} \to 1 - \lambda$, and hence

$$\sqrt{n}\sqrt{\frac{m}{(n+m)}} \{v_F(p)\}^{-1} (V_{F_n}(p) - V_F(p)) \quad \Rightarrow \quad N(0, (1-\lambda)p(1-p)); \tag{4.1}$$

$$\sqrt{m}\sqrt{\frac{n}{(n+m)}} \{v_G(p)\}^{-1} (V_{G_m}(p) - V_G(p)) \quad \Rightarrow \quad N(0, \lambda p(1-p)). \tag{4.2}$$

As (4.1) and (4.2) are independent and under $H_0 : F = G$, thus we have,

$$\sqrt{\frac{nm}{(n+m)}} \{v_F(p)\}^{-1} ((V_{F_n}(p) - V_F(p)) - (V_{G_m}(p) - V_F(p))) \quad \Rightarrow \quad N(0, (1-\lambda+\lambda)p(1-p))$$

$$\Rightarrow \sqrt{\frac{nm}{(n+m)}} \{v_F(p)\}^{-1} ((V_{F_n}(p) - V_{G_m}(p)) \quad \Rightarrow \quad N(0, p(1-p)).$$

Similarly we can show the convergence of the higher finite order sub processes to the

appropriate multivariate normal distribution. Thus we have our result. □

Formal tests can be considered based on the above theorem, which we propose to carry out in some future work.

CHAPTER 5

TEST OF MULTIVARIATE NORMALITY

In statistical data analysis investigating whether the data points are from a specific distribution is a very important problem. For one dimensional data, several tests exist for that purpose, eg. chi-square goodness of fit test, Kolmogorov-Smirnov test, Shapiro-Wilks test, Cramer von-Mises test etc. A similar problem may also be defined in the multidimensional case. For $\mathbf{X}_1, \cdots, \mathbf{X}_n$ n d-dimensional data points with a distribution function F, we may wish to test whether F has a specific distribution, eg. a multivariate normal distribution, $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for some unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Many of the tests for multivariate normality are extensions of the univariate normality tests (see Royston (1983), Justel, Pefia, Zamar (1997), Doornik and Hansen (2008)). Quite a few of them depend on graphical representation. For example, the line test by Hald (1952) is based on deviation of points in a probability plot from a straight line, the ring test by Rietz (1943) and Cramer (1946) is based on sample points distributed within elliptical rings. For more details on these tests one can see Kowalski (1970). More recently Ghosh (1996) has proposed a graphical test for checking normality based on the third derivative of the logarithm of the empirical moment-generating function of observed data, where significant departure of this function from zero is an indication of departure from normality.

For the multivariate case, Wilk and Gnanadesikan (1968) proposed to study quantile-quantile plots of the observations considering the individual components separately while using a common distribution for comparison. Cox and Small (1978) considered both

coordinate dependent procedures and invariant procedures. In a coordinate dependent procedure, linearity of regression relationships are tested. For example, in a bivariate case if (x_{i1}, x_{i2}), $i = 1, \cdots, n$, are n independent observations, then the null hypothesis is that they are independent and identically distributed random variables (X_1, X_2) with a bivariate normal distribution. The conditional distribution of one variable, given the other, is checked to see whether the conditional distribution is linearly dependent on the other variables by taking linear and quadratic combinations of the other variable and checking whether the coefficients of the non linear terms are significant. For higher-dimensional data they suggested a visual test based on the quantile-quantile plot of the t-statistics of the estimates of the coefficients of the non linear terms against that of the standard normal distribution. Koziol (1983) proposed multivariate normality tests based on properties of radii and angles of normally distributed random vectors, and to combine different procedures to achieve omnibus tests. Subsequently, Koziol (1993) looked at quantile-quantile plots of the individual components of "smooth" skewness and kurtosis statistics of data, plotting them against their theoretical distributions under the hypothesis of multivariate normality. However, these theoretical distributions may be challenging to obtain. In a somewhat similar approach, Liang, Pan and Yang (2004) also considered the quantile-quantile plots of certain transformations of the multivariate data against their theoretical distributions under the hypothesis of multivariate normality; however in this case their proposed statistics have easily tractable standard distributions. Recently Liang and Ng (2009) proposed a visual test to detect non-normality by performing a principal-component analysis and then applying Royston (1993)'s test for univariate normality to test for normality of each principal component, rejecting the hypothesis of multivariate normality if a departure from normality is detected in any principal component direction. They also performed a detailed literature review of presently existing graphical techniques for detection of departure from multivariate normality, and formal tests based on such techniques.

In the univariate case, the Kolmogorov-Smirnov test and the Cramer von-Mises test

are perhaps the two most popular tests to check goodness of fit of a specific distribution in general and normality in particular. We will discuss these tests in more details in Section 5.2 where we perform a power study based on these tests.

Among other notable univariate tests of normality, Shapiro and Wilk (1965) defined a test statistic to test for normality, where the test statistic is the ratio of the square of an appropriate linear combination of the sample order statistics to the usual symmetric estimate of variance. Also this ratio is scale and location invariant. Malkovich and Afifi (1973) extended the univariate test statistic of Shapiro and Wilk to test a hypothesis of multivariate normality. This new test statistic is affine invariant. For $d = 1$, it reduces to the Shapiro and Wilk's test statistic.

Malkovich and Afifi (1973) also suggested a test method based on multivariate skewness and kurtosis, which are generalised notions of univariate skewness and kurtosis. Their proposed test statistic is affine invariant and rejects the null hypothesis of multivariate normality for large values of measures of skewness and kurtosis. Mardia (1970) suggested some measures of multivariate skewness $\beta_{1,d}$ and kurtosis $\beta_{2,d}$ for d-dimensional distributions and proposed a test for multivariate normality based on the sample skewness $b_{1,d}$ and the sample kurtosis $b_{2,d}$ of a multivariate distribution. According to his proposed method, the test of normality is performed by testing $\beta_{1,d} = 0$ and $\beta_{2,d} = d(d+2)$ separately. Under $H_0 : \beta_{1,d} = 0$, $\frac{1}{6}nb_{1,d}$ has a χ^2 distribution with $d(d+1)(d+2)/6$ degrees of freedom and we reject the null hypothesis if the value of $\frac{1}{6}nb_{1,d}$ is large, which would imply the presence of skewness present in the distribution. Also, under the assumption of normality $(b_{2,d} - \beta_{2,d})/\{8d(d+2)/n\}^{\frac{1}{2}}$ is distributed as $N(0,1)$ and the null hypothesis is rejected for large values of $|(b_{2,d} - \beta_{2,d})/\{8d(d+2)/n\}^{\frac{1}{2}}|$. Further, Mardia (1974) gave alternative forms of these measures, which were convenient for computer programming and checking the invariance property of these measures. Kankainen, Taskinen and Oja (2007) proposed some tests of multinormality depending on location vectors and scatter matrices. They defined a d-vector valued test statistic T as a location vector if it is affine equivariant and

a $d \times d$-matrix valued test statistic C as a scatter matrix if it was affine equivariant. In the univariate case the Mahalanobis difference between two location vectors is used to detect skewness. Also measures of univariate kurtosis are usually ratios of two scale measures. A similar idea is used here to detect skewness and kurtosis using two separate location vectors and two separate scatter matrices respectively. They defined two test statistics for multinormality, one to detect skewness and the other to detect kurtosis. For two different location vectors T_1 and T_2 and a scatter matrix C, their statistic to detect skewness, U, is defined as $U = (T_1 - T_2)^T C^{-1}(T_1 - T_2)$. If C_1 and C_2 are two different scatter matrices, then the statistic to detect kurtosis, W, is defined as $W = ||C_1^{-1}C_2 - I_d||^2$. These two statistic are affine invariant and it can be shown that the test construction method is a generalised version of Mardia's measures of skewness and kurtosis.

Ghosh and Ruymgaart (1992) proposed a test of normality based on the studentised empirical characteristic function. The computation of their test statistic is, however, complicated in practice and does not have a tractable asymptotic distribution beyond $d = 2$.

Romeu and Ozturk (1993) divided the existing methods for testing multivariate normality into six classes and compared among themselves. Szekely and Rizzo (2005) proposed a class of V-statistics to test multivariate normality where the test statistic was affine invariant and consistent against all fixed alternatives.

In the remainder of this chapter we develop a test of multivariate normality based on volume functionals of central rank regions, which can readily be generalised to other distributions. We then compare our proposed test with extensions of the Kolmogorov-Smirnov test and the Cramer-von Mises test proposed by Malkovich and Afifi (1973). In Section 5.1, we define our test statistic as a function of the scale curve and develop the asymptotic properties of the test statistic. In Section 5.2, we study the power of test for different sample sizes under various alternatives and compare the power with the power of the Kolmogorov-Smirnov test and the Cramer-von Mises test as proposed by Malkovich

and Afifi (1973).

5.1 Test of Multivariate Normality Based on Scale Curves

For the remainder of this chapter we consider $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n \in \mathbb{R}^d$ independent with common distribution function F. Suppose that we want to test the null hypothesis $H_0 : F \in \mathcal{F}$ where $\mathcal{F} = \{N_d(\boldsymbol{\theta}, \Sigma) : \boldsymbol{\theta} \in \mathbb{R}^d$ and Σ is a $d \times d$ positive definite matrix $\}$. Let us assume F_0 is the d-variate standard normal distribution. Then all the 5 conditions $A1, A2, A3, A4$ and $A5$ will hold under H_0. It is quite obvious that $A1$ is satisfied by the multivariate normal distributions. The conditions $A2$ and $A3$ follows from Theorem 2.2.1. The last two conditions $A4$ and $A5$ will clearly hold due to (3.56).

Let $V_{F_n}(p)$ denote the scale curve based on the data and let $V_{F_0}(p)$ be the scale curve based on F_0. Define the slope functional

$$s_n(p) = \frac{V_{F_n}(p)}{V_{F_0}(p)}. \tag{5.1}$$

We now state some properties of the slope functional $s_n(p)$.

Theorem 5.1.1 *Let us assume that H_0 is true, then the following will hold:*

1. $s_n(p) \xrightarrow{p} c_0$, $0 < p < 1$, *as* $n \to \infty$ *for a constant* $c_0 > 0$ *independent of* p.
2. $\frac{\sqrt{n} V_{F_0}(p)(s_n(p) - c_0)}{v_F(p)}$ *converges weakly to a Brownian bridge* $\{B(p), 0 < p < 1\}$ *as* $n \to \infty$.

Proof :

1. From Theorem 3.1.1, we have $V_{F_n}(p) \xrightarrow{p} V_F(p)$ where F is the true distribution of $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Now, by Theorem 2.3.1, $V_F(p) = c_0 V_{F_0}(p)$ whenever $F \in \mathcal{F}$, where c_0 is a constant dependent only on Σ but not on p. Hence by the definition of the slope functional $s_n(p)$, we can say that $s_n(p) \xrightarrow{p} c_0$.

2. By Theorem 3.1.1, we have $\left\{\sqrt{n}\left\{v_F(p)\right\}^{-1}\left(V_{F_n}(p)-V_F(p)\right), 0<p<1\right\}$ converges to a Brownian bridge. Hence under H_0, $\left\{\sqrt{n}\left\{v_F(p)\right\}^{-1}\left(V_{F_n}(p)-V_F(p)\right), 0<p<1\right\}$ converges to a Brownian bridge. Therefore from the definition of the slope functional $s_n(p)$, and the fact that $V_F(p)=c_0V_{F_0}(p)$, we have the result. □

As a consequence, for any finite integer k independent of n, we can write

$$\sqrt{n}\Sigma_k^{-\frac{1}{2}}\left(\left(s_n\left(\frac{1}{k+1}\right),s_n\left(\frac{2}{k+1}\right),\cdots,s_n\left(\frac{k}{k+1}\right)\right)^T-c_0\left(1,1,\cdots,1\right)^T\right) \tag{5.2}$$

converges in distribution to $N_k(\mathbf{0},\mathbf{I}_{k\times k})$, where $\Sigma_k=(\sigma(i,j))_{k\times k}$ has elements

$$\sigma(i,j)=\begin{cases} v_{F_0}\left(\frac{i}{k+1}\right)\times v_{F_0}\left(\frac{j}{k+1}\right).\left(\frac{\min(i,j)}{(k+1)}-\frac{ij}{(k+1)^2}\right)/\left(V_{F_0}(\frac{i}{k+1}).V_{F_0}(\frac{j}{k+1})\right), \text{ if } i\neq j,\\ \frac{i}{(k+1)}\left(1-\frac{i}{(k+1)}\right)\times v_{F_0}^2\left(\frac{i}{k+1}\right)/V_{F_0}^2(\frac{i}{k+1}), \text{ if } i=j. \end{cases} \tag{5.3}$$

By Theorem 3.1.1, any k-dimensional sub process of $\sqrt{n}\left\{V_{F_n}(p)-V_F(p), 0<p<1\right\}$, $\sqrt{n}((V_{F_n}(p_1),\cdots,V_{F_n}(p_k))-(V_F(p_1),\cdots,V_F(p_k))$ will converge in distribution to $N(0,\Sigma_k')$, where $\Sigma_k'=((\sigma'(i,j)))_{k\times k}$ with

$$\sigma'(i,j)=\begin{cases} v_{F_0}(p_i)\times v_{F_0}(p_j)\times\left(\min\left\{p_i,p_j\right\}-p_ip_j\right), \text{ if } i\neq j,\\ p_i\left(1-p_i\right)\times v_{F_0}^2\left(p_i\right), \text{ if } i=j. \end{cases} \tag{5.4}$$

Under H_0, we expect to have $s_n(p)$ close to a constant independent of p for all $p\in(0,1)$. Hence to test H_0 we consider a test statistic, which we will refer to as the *scale test* statistic, T_n defined as follows:

$$T_n=\log U_n-V_n, \tag{5.5}$$

where

$$U_n=\frac{1}{k}\sum_{i=1}^{k}s_n\left(\frac{i}{(k+1)}\right) \tag{5.6}$$

and

$$V_n=\frac{1}{k}\sum_{i=1}^{k}\log\left[s_n\left(\frac{i}{(k+1)}\right)\right]. \tag{5.7}$$

From the definition of U_n and V_n, we can say that $U_n \approx \int s_n d\lambda$ and $V_n \approx \int log(s_n) d\lambda$, for large k, where λ is the Lebesgue measure on $(0,1)$.

We now establish an invariance property of T_n.

Lemma 5.1.1 *T_n is invariant under affine transformations of the sample $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$.*

Proof : Let $\mathbf{A}$ be a non-singular $d \times d$ matrix and $\mathbf{b} \in \mathbb{R}^d$. Define $\mathbf{Y}_i = \mathbf{A}\mathbf{X}_i + \mathbf{b}$ for $i = 1, \cdots, n$. Suppose the empirical distribution of $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ is F_n and that of $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ is G_n. Let $T_n^{(\mathbf{X})}$ and $T_n^{(\mathbf{Y})}$ be the versions of T_n as defined in (5.5) for the $\mathbf{X}_i$s and $\mathbf{Y}_i$s respectively. We need to show $T_n^{(\mathbf{X})} = T_n^{(\mathbf{Y})}$.

We have seen earlier that $V_{G_n}(p) = |\mathbf{A}| V_{F_n}(p)$ and hence $s_n^{(\mathbf{Y})}(p) = |\mathbf{A}| s_n^{(\mathbf{X})}(p)$ where $s_n^{(\mathbf{Y})}(p)$ and $s_n^{(\mathbf{X})}(p)$ are versions of the slope functional as defined in (5.1) for the $\mathbf{X}_i$s and $\mathbf{Y}_i$s respectively. Then

$$\begin{aligned} T_n^{(\mathbf{Y})} &= \log \left(\frac{\frac{1}{k} \sum_{i=1}^{k} s_n^{(\mathbf{Y})} \left(\frac{i}{(k+1)} \right)}{\left(\prod_{i=1}^{k} s_n^{(\mathbf{Y})} \left(\frac{i}{(k+1)} \right) \right)^{\frac{1}{k}}} \right) \\ &= \log \left(\frac{|\mathbf{A}|^{\frac{1}{k}} \sum_{i=1}^{k} s_n^{(\mathbf{X})} \left(\frac{i}{(k+1)} \right)}{|\mathbf{A}| \left(\prod_{i=1}^{k} s_n^{(\mathbf{X})} \left(\frac{i}{(k+1)} \right) \right)^{\frac{1}{k}}} \right) \\ &= T_n^{(\mathbf{X})}. \end{aligned} \tag{5.8}$$

Thus the lemma is proved. □

By Theorem 5.1.1, $\left(s_n\left(\frac{1}{k+1}\right), s_n\left(\frac{2}{k+1}\right), \cdots, s_n\left(\frac{k}{k+1}\right) \right)$ is approximately multivariate normal under H_0 for large n. Hence for the proposed test we reject H_0 at level α if $T_n > c_n(\alpha)$, for some suitable $c_n(\alpha) > 0$. In the following theorem we state some properties of T_n.

Theorem 5.1.2 *Let us assume that H_0 is true. Then the following will hold:*

1. $T_n \geq 0$.

2. *Under* H_0, $T_n \xrightarrow{p} 0$.

3. nT_n *converges in distribution to a weighted sum of* k χ^2_1 *random variables as* $n \to \infty$.

Proof : Note that the distribution of T_n remains invariant for any $F \in \mathcal{F}$, so without loss of generality let us assume that $F = F_0$.

1. The term U_n is the arithmetic mean (A. M.) of $s_n\left(\frac{i}{(k+1)}\right)$ and V_n is the logarithm of the geometric mean (G. M.) of $s_n\left(\frac{i}{(k+1)}\right)$ for $i = 1, \cdots, k$. Thus as A.M.$\geq$ G.M. we have $T_n \geq 0$.

2. By (5.2) we have

$$\left(s_n\left(\frac{1}{k+1}\right), s_n\left(\frac{2}{k+1}\right), \cdots, s_n\left(\frac{k}{k+1}\right)\right)^T \xrightarrow{p} (1, 1, \cdots, 1)^T.$$

   As T_n is a continuous function of $\left(s_n\left(\frac{1}{k+1}\right), s_n\left(\frac{2}{k+1}\right), \cdots, s_n\left(\frac{k}{k+1}\right)\right)^T$, hence the result follows by the continuous mapping theorem.

3. From (5.5) we can see that

$$T_n = \log\left(\frac{1}{k}\sum_{i=1}^{k} s_n\left(\frac{i}{(k+1)}\right)\right) - \frac{1}{k}\sum_{i=1}^{k} \log s_n\left(\frac{i}{(k+1)}\right). \tag{5.9}$$

   As before we define

$$s_{ni} = s_n\left(\frac{i}{(k+1)}\right) \tag{5.10}$$

   and $\mathbf{S}_n = (s_{n1}, s_{n2}, \cdots, s_{nk})^T$. Also note that from Theorem 3.1.1,

$$E(s_{ni}) = 1 + o\left(\frac{1}{\sqrt{n}}\right). \tag{5.11}$$

Let

$$f(x_1, x_2, \cdots, x_k) := \log\left(\frac{1}{k}\sum_{i=1}^{k} x_i\right) - \frac{1}{k}\left(\sum_{i=1}^{k} \log x_i\right).$$

Therefore,

$$\frac{\partial f}{\partial x_i} = \frac{1}{\sum_{j=1}^{k} x_j} - \frac{1}{kx_i}, 1 \leq i \leq k;$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x_i \partial x_j} &= \begin{cases} -\frac{1}{(\sum_{l=1}^{k} x_l)^2}, \text{ if } i \neq j. \\ -\frac{1}{(\sum_{l=1}^{k} x_l)^2} + \frac{1}{kx_i^2}, \text{ if } i = j \end{cases} \\ &= -\frac{1}{(\sum_{l=1}^{k} x_l)^2} + \frac{1}{kx_i^2} I(i=j) \end{aligned} \tag{5.12}$$

Now, by using the Taylor series expansion up to the second term,

$$\begin{aligned} &(f(\mathbf{S}_n) - f(E(\mathbf{S}_n)) \\ &\approx \bigtriangledown f(E(\mathbf{S}_n)).(\mathbf{S}_n - E(\mathbf{S}_n)) + \frac{1}{2}(\mathbf{S}_n - E(\mathbf{S}_n))^T \bigtriangledown^2 f(E(\mathbf{S}_n)).(\mathbf{S}_n - E(\mathbf{S}_n)) \end{aligned} \tag{5.13}$$

where

$$\bigtriangledown f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_k}\right)^T$$

and

$$\bigtriangledown^2 f(\mathbf{x}) = \left(\left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right)\right).$$

Now

$$f(E(\mathbf{S}_n)) \approx \log\left(\frac{1}{k}\sum_{i=1}^{k} 1\right) - \frac{1}{k}\sum_{i=1}^{k} \log 1 = 0$$

up to order $\sqrt{n}$ by (5.11). Again

$$\bigtriangledown f(E(\mathbf{S}_n)) \approx \frac{1}{k} - \frac{1}{k} + o_P\left(\frac{1}{\sqrt{n}}\right) = o_P\left(\frac{1}{\sqrt{n}}\right) \text{ by (5.11).}$$

Now $\mathbf{S}_n - E(\mathbf{S}_n) = O_P\left(\frac{1}{\sqrt{n}}\right)$, which implies $\bigtriangledown f(E(\mathbf{S}_n))(\mathbf{S}_n - E(\mathbf{S}_n)) = o_P\left(\frac{1}{n}\right)$. Hence $\sqrt{n}T_n \xrightarrow{d} 0$ and hence $\sqrt{n}T_n \xrightarrow{p} 0$ as $\sqrt{n}(\mathbf{S}_n - E(\mathbf{S}_n)) \xrightarrow{d} N_k(0, \Sigma_k)$ by Theorem 3.1.1 where $\Sigma_k = (\sigma(i,j))_{k\times k}$ has elements

$$\sigma(i,j) = \begin{cases} v_{F_0}\left(\frac{i}{k+1}\right) \times v_{F_0}\left(\frac{j}{k+1}\right) \cdot \left(\frac{\min(i,j)}{(k+1)} - \frac{ij}{(k+1)^2}\right) / \left(V_{F_0}(\frac{i}{k+1}).V_{F_0}(\frac{j}{k+1})\right), \text{ if } i \neq j, \\ \frac{i}{(k+1)}\left(1 - \frac{i}{(k+1)}\right) \times v^2_{F_0}\left(\frac{i}{k+1}\right) / V^2_{F_0}(\frac{i}{k+1}), \text{ if } i = j. \end{cases} \tag{5.14}$$

Hence,

$$f(\mathbf{S}_n) = \frac{1}{2}(\mathbf{S}_n - E(\mathbf{S}_n))^T \bigtriangledown^2 f(E(\mathbf{S}_n)).(\mathbf{S}_n - E(\mathbf{S}_n)) + o_P\left(\frac{1}{n}\right)$$

where

$$\begin{aligned} \left(\left(\bigtriangledown^2 f(E(\mathbf{S}_n))\right)\right)_{ij} &= -\frac{1}{k^2} + \frac{1}{k} I(i=j) \\ \bigtriangledown^2 f(E(\mathbf{S}_n)) &= \frac{1}{k} I - \frac{1}{k^2}\mathbf{1}\mathbf{1}^T. \end{aligned}$$

Thus

$$nT_n = nf(\mathbf{S}_n) = (\sqrt{n}(\mathbf{S}_n - \mathbf{1}))^T(\frac{1}{2} \bigtriangledown^2 f(\mathbf{1}))(\sqrt{n}(\mathbf{S}_n - \mathbf{1})) + o_P(1).$$

Therefore $nT_n \xrightarrow{d} \sum_{i=1}^k c_i \chi^2_1$, where c_i's are the eigen values of

$$\frac{1}{2}\Sigma_k^{\frac{1}{2}} \bigtriangledown^2 f(\mathbf{1})\Sigma_k^{\frac{1}{2}} \equiv \frac{1}{2k}\Sigma_k - \frac{1}{2k^2}(\Sigma^{\frac{1}{2}}\mathbf{1}) \cdot (\Sigma^{\frac{1}{2}}\mathbf{1})^T.$$

□

5.2 Finite Sample Power: A Numerical Study

We now construct a test of multinormality based on the test statistic T_n. Instead of using the asymptotic distribution derived in the previous section, we simulate the finite sample null distribution of the statistic T_n and tabulate the 95 percentile in the Table 5.1. We describe the procedure in detail in the following:

To find out the size and power of the test for testing of $H_0 : F \in$ the set of all d-variate normal distributions, we need to find out the cut off values for the test statistic T_n. In finite samples we estimate the level α cut off $c_n(\alpha)$ by taking it to be the upper α-th quantile of a simulated distribution of T_n.

We now provide an algorithm for obtaining cut off values for T_n to perform a goodness of fit test for multivariate normal distribution for dimensions 2 and 3 and for sample sizes $n = 30, 50, 100$ and level of significance, $\alpha = 0.05$.

1. Suppose F_0 is a standard multivariate normal distribution. Compute the value of the volume functional V_{F_0} for F_0 using Example 1 and (3.54) of Chapter 3.

2. Generate n random samples $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ from a multivariate standard normal distribution and compute the corresponding volume functional $V_{F_n}(p_i)$ where $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$.

3. Compute the slope functional $s_n(p_i)$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$ using (5.1) and then compute the test statistic T_n for each i using (5.5).

4. Repeat steps 2-3 1000 times and get 1000 values of T_n and order them.

5. Take the 95-th percentile value of the ordered T_n as our cutoff value $c_n(0.05)$.

In Table 5.1 we give the cut off values for bivariate normal distribution and trivariate normal distributions as null distributions. The computer program we have used for simulation was written in C language. We took $k = 10$, in all of our simulations.

Table 5.1: The values of $c_n(\alpha)$ for proposed test for bivariate and trivariate standard normal distributions as the null distributions where $\alpha = 0.05$ and $n = 30, 50, 100$ after 1000 iterations.

Null Distribution	$n = 30$	$n = 50$	$n = 100$
Bivariate Normal Distribution	0.099202	0.059806	0.029459
Trivariate Normal Distribution	0.112553	0.067311	0.041632

5.2.1 Power Under Mixture Normal Alternatives

We compute the power of the scale test under mixture normal alternatives for multivariate normal distribution. Here we have considered two different sets of alternatives. At first we use the following mixture model as our alternative,

$$F = (1 - \delta)F_0 + \delta F_1,$$

where F_0 is a standard multivariate normal distribution and F_1 is another normal distribution with a different mean vector. Both F and F_0 have identity matrix as the variance matrix.

We now provide the steps required for estimating the power of the test against our pre-specified alternative distribution, F_A with location vector $\boldsymbol{\mu}$, for a sample size n:

1. Generate a random sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ from F_A.

2. Compute the corresponding volume functional $V_{F_n}(p_i)$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$.

3. Compute the slope functional $s_n(p_i) = \frac{V_{F_n}(p_i)}{V_{F_0}(p_i)}$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$, using (5.1) and then compute the test statistic T_n for each i using (5.5). Here F_0 is the standard multivariate normal distribution.

4. Repeat steps 1-4 1000 times and get 1000 values of T_n.

5. Estimate the power of the test for F_A for sample size n at 5% level by the proportion of T_ns greater than the cut off value $c_n(0.05)$.

First we vary the value of δ to explore the various degrees of departure from normality. We take $\delta = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ and calculate the powers under the respective alternative mixture models, which are presented in Table 5.2. $\delta = 0.0$ is the null case and it gives us the size of the test.

Table 5.2: Power of the proposed test under mixture alternatives bivariate normal with $\mu = (5, 5)^T$, and trivariate normal with $\mu = (5, 5, 5)^T$.

		δ						
Distribution		0.0	0.05	0.1	0.2	0.3	0.4	0.5
Bivariate Normal	$n = 30$	0.044	0.140	0.182	0.119	0.213	0.507	0.636
	$n = 50$	0.045	0.277	0.407	0.223	0.252	0.690	0.894
	$n = 100$	0.048	0.613	0.834	0.530	0.326	0.890	0.996
Trivariate Normal	$n = 30$	0.036	0.325	0.584	0.651	0.544	0.351	0.289
	$n = 50$	0.053	0.428	0.762	0.813	0.598	0.373	0.359
	$n = 100$	0.051	0.589	0.933	0.907	0.695	0.468	0.401

Next we consider another departure from normality by keeping δ fixed at 0.1 and taking $\boldsymbol{\mu} = r\mathbf{1}_{d\times 1}$ where $r = 1, 2, 3, 4, 5, 10$, where we look at dimension $d = 2$ and 3. Table 5.3 gives the power of the proposed test for this situation for bivariate and trivariate normal distribution. From Table 5.3 we can see that our proposed test in bivariate normal case starts detecting a shift when $r = 4$ for $n = 30$. For $n = 50$ and 100, the test can detect a shift for $r = 3$ or larger. In the trivariate normal distribution case our proposed test can detect a shift when $r = 2$ or more.

Table 5.3: Power of the proposed test under mixture alternatives bivariate normal with $\boldsymbol{\mu} = (r, r)^T$, and trivariate normal with $\boldsymbol{\mu} = (r, r, r)^T$ and $\delta = 0.1$.

		r						
Distribution		0.0	1.0	2.0	3.0	4.0	5.0	10.0
Bivariate Normal	$n = 30$	0.045	0.046	0.044	0.053	0.102	0.187	0.683
	$n = 50$	0.050	0.051	0.055	0.085	0.231	0.434	0.936
	$n = 100$	0.048	0.056	0.067	0.210	0.566	0.841	0.997
Trivariate Normal	$n = 30$	0.041	0.034	0.107	0.228	0.435	0.584	0.816
	$n = 50$	0.049	0.060	0.158	0.383	0.618	0.762	0.929
	$n = 100$	0.060	0.055	0.208	0.544	0.816	0.933	0.992

The Kolmogorov-Smirnov test and the Cramer von-Mises test are two very common tests to check the goodness of fit for univariate distributions (see Malkovich and Afifi (1973)). For the univariate Kolmogorov-Smirnov test, the test statistic for the null hypothesis $H_0 : F = F_0$ is $\Delta_n = \sup_x |F_n(x) - F_0(x)|$, where $F_n(x)$ is the sample distribution function at x. The distribution of Δ_n is independent of F if F is continuous. We reject H_0 for large values of Δ_n.

The Cramer-von Mises test of goodness of fit of H_0 is based on the test statistic $\omega_n = \int_{-\infty}^{\infty}[F_n(x) - F_0(x)]^2 dF_0(x)$. The distribution of ω_n does not depend on F when F is continuous. Again we reject H_0 for large values of ω_n.

Table 5.4: Power of the Kolmogorov-Smirnov test statistic under mixture normal alternative $(1-\delta)N(\mathbf{0}, I) + \delta N(\boldsymbol{\mu}, I)$ for the bivariate and trivariate case respectively with $\boldsymbol{\mu} = (5,5)^T$, and $\boldsymbol{\mu} = (5,5,5)^T$.

		δ						
Distribution		0.0	0.05	0.1	0.2	0.3	0.4	0.5
Bivariate Normal	$n = 30$	0.022	0.389	0.362	0.149	0.039	0.057	0.078
	$n = 50$	0.051	0.806	0.631	0.139	0.172	0.574	0.799
	$n = 100$	0.028	0.902	0.926	0.323	0.033	0.188	0.425
Trivariate Normal	$n = 30$	0.013	0.306	0.232	0.090	0.018	0.005	0.018
	$n = 50$	0.014	0.081	0.038	0.018	0.090	0.250	0.342
	$n = 100$	0.043	0.977	0.912	0.280	0.041	0.070	0.126

When the Kolmogorov-Smirnov test is extended to a d-variate case, where $d > 1$, one possible generalised test statistic would be $\sup_{x_1,\cdots,x_d} |F_n(x_1, \cdots, x_d) - F(x_1, \cdots, x_d)|$. However Simpson (1951) showed that the distribution of this statistic is not the same for all continuous F when $d > 1$. Malkovich and Afifi (1973) suggested another way of extending the Kolmogorov-Smirnov test to a multidimensional case when F is normal. They observed that when $(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ are independent samples from $F_0 = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the d-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$, then $(\mathbf{X}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu})$ are iid χ^2_d. Based on that observation, they used a transformation $V_i = (\mathbf{X}_i - \overline{\mathbf{X}})^T \mathbf{S}^{-1} (\mathbf{X}_i - \overline{\mathbf{X}})$. They developed tests based on the idea that for a large n, the distribution of V_i's would be close to χ^2_d. Hence they used the

Kolmogorov-Smirnov test statistic given by $\Delta_n^V = \sup_x |F_n^V(x) - F_d(x)|$ where $F_n^V(x)$ is the sample distribution function based on $(V_1, \cdots, V_n)$ and $F_d(x)$ is the distribution function of χ_d^2 at x. In the same paper Malkovich and Afifi also extended the Cramer-von Mises test statistic similarly to a test of multivariate normality by using the test statistic $\omega_n^V = \int_{-\infty}^{\infty}[F_n(x) - F(x)]^2 dF(x)$.

Table 5.5: Power of the Cramer von-Mises test statistic under mixture normal alternative $(1-\delta)N(\mathbf{0}, I)+\delta N(\boldsymbol{\mu}, I)$ for the bivariate and trivariate case respectively with $\boldsymbol{\mu} = (5, 5)^T$, and $\boldsymbol{\mu} = (5, 5, 5)^T$.

		δ						
Distribution		0.0	0.05	0.1	0.2	0.3	0.4	0.5
Bivariate Normal	$n = 30$	0.061	0.659	0.651	0.280	0.104	0.198	0.265
	$n = 50$	0.050	0.874	0.759	0.219	0.274	0.728	0.918
	$n = 100$	0.048	0.968	0.976	0.424	0.080	0.413	0.720
Trivariate Normal	$n = 30$	0.050	0.576	0.507	0.223	0.045	0.026	0.040
	$n = 50$	0.049	0.084	0.049	0.042	0.245	0.515	0.637
	$n = 100$	0.070	0.989	0.954	0.333	0.030	0.079	0.168

Table 5.6: Power of the Kolmogorov-Smirnov test statistic under mixture normal alternative $0.9N(\mathbf{0}, I) + 0.1N(\boldsymbol{\mu}, I)$ for the bivariate and trivariate case respectively with $\boldsymbol{\mu} = (r, r)^T$, and $\boldsymbol{\mu} = (r, r, r)^T$

		r						
Distribution		0.0	1.0	2.0	3.0	4.0	5.0	10.0
Bivariate Normal	$n = 30$	0.034	0.042	0.060	0.155	0.262	0.418	0.687
	$n = 50$	0.012	0.017	0.053	0.178	0.411	0.594	0.841
	$n = 100$	0.042	0.043	0.035	0.147	0.423	0.701	0.930
Trivariate Normal	$n = 30$	0.047	0.056	0.099	0.202	0.328	0.373	0.550
	$n = 50$	0.025	0.031	0.028	0.082	0.155	0.231	0.435
	$n = 100$	0.022	0.037	0.165	0.472	0.711	0.808	0.919

We present the power of our proposed test, Kolmogorov-Smirnov test and Cramer von-Mises test in Tables 5.2, 5.4 and 5.5 respectively when the departure from normality in the mixture model is achieved by varying the values of δ for $d = 2$ and $d = 3$. The size of the proposed test is slightly better than the Kolmogorov-Smirnov test but not as good as the Cramer von-Mises test. Both Kolmogorov-Smirnov and Cramer von-Mises tests

Table 5.7: Power of the Cramer von-Mises test statistic under mixture normal alternative $0.9N(\mathbf{0}, I) + 0.1N(\boldsymbol{\mu}, I)$ for the bivariate and trivariate case respectively with $\boldsymbol{\mu} = (r, r)^T$, and $\boldsymbol{\mu} = (r, r, r)^T$

		r						
Distribution		0.0	1.0	2.0	3.0	4.0	5.0	10.0
Bivariate Normal	$n = 30$	0.051	0.057	0.087	0.201	0.332	0.528	0.785
	$n = 50$	0.035	0.046	0.118	0.340	0.654	0.801	0.934
	$n = 100$	0.033	0.027	0.036	0.205	0.621	0.860	0.970
Trivariate Normal	$n = 30$	0.045	0.070	0.119	0.256	0.418	0.475	0.680
	$n = 50$	0.036	0.047	0.059	0.149	0.299	0.407	0.627
	$n = 100$	0.047	0.071	0.298	0.687	0.873	0.935	0.971

can detect the shift from normality better than the proposed test in the bivariate case. In the trivariate case, the proposed test performs better than both the tests. It should be noted that power is not an increasing function of δ. As a variant we can fix $\delta = 0.1$ and vary $\boldsymbol{\mu}$. As an illustration, we fix $\delta = 0.1$ and compute the power for $\boldsymbol{\mu} = r\mathbf{1}_{d\times 1}$ where $r = 1.0, 2.0, 3.0, 4.0, 5.0, 10.0$, for dimensions $d = 2$ and 3. The powers and the levels of the proposed test, Kolmogorov-Smirnov test and Cramer von-Mises test are given in Tables 5.3, 5.6 and 5.7 respectively. The proposed test performs slightly better than Kolmogorov-Smirnov test but not as good as Cramer- von-Mises test.

In this Section 5.2 we have presented the empirical power of our proposed test in tables 5.2 and 5.3 and compared these powers with powers obtained from Kolmogorov-Smirnov test and Cramer von-Mises test as proposed by Malkovich and Afifi (1973) in tables 5.4, 5.5, 5.6 and 5.7. To compute the powers of the tests in tables 5.3, 5.6 and 5.7 we have taken $\delta = 0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ keeping $\boldsymbol{\mu}$ fixed to obtain the mixture alternatives. In all those tables we can see a pattern for the power variation. One would think that as δ increases shift from normality would increase. But actually that is not the case. We have seen in tables 5.3, 5.6 and 5.7 that the power does not increase monotonically as δ increases. In the power tables for a fixed $\boldsymbol{\mu}$ for $d = 2$ the power first increases then decreases around $\delta = 0.2, 0.3$ and then again increases. As an illustration we have presented the density plots for the test statistic T_n in figures 5.1, 5.2 and 5.3 for $d = 2$ and

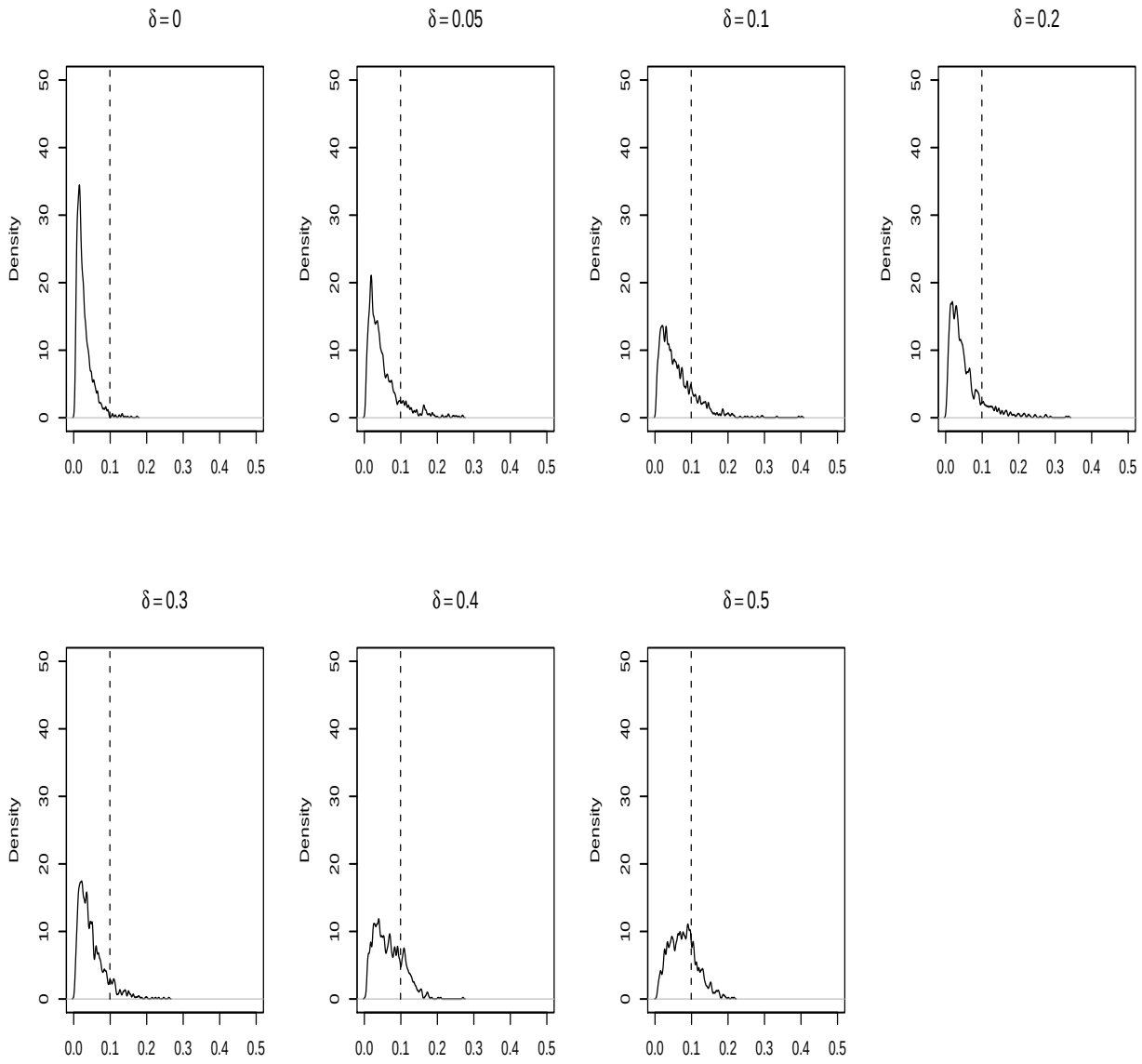

Figure 5.1: Density Plot for the Test Statistic T_n for $n = 30$ and $d = 2$. The dotted line gives the 95-th percentile of the T_n values based on the 1000 simulated values of T_n for $\delta = 0.0$

$n = 30, 50, 100$ respectively. In all these plots the dotted line gives the 95-th percentile of the T_n values based on the 1000 simulated values of T_n for $\delta = 0.0$. In Figure 5.1 we can see that though the dotted line shifts to the left side for $\delta = 0.05$ and $\delta = 0.1$, it returns at a similar place with respect to the density plot of T_n for $\delta = 0.05$ at $\delta = 0.2$. Then it again shifts towards the left side of the density plot for $\delta = 0.3, 0.4, 0.5$. In Figure 5.2 we can see that the dotted line first shifts towards the left for $\delta = 0.05$ and $\delta = 0.1$ but it shifts towards the right for $\delta = 0.2$ and $\delta = 0.3$ and again shifts towards the left for $\delta = 0.4, 0.5$. For $n = 100$ we present the density plot for T_n in Figure 5.3. Here we can also see that the dotted line first shifts towards the left for $\delta = 0.05, 0.1, 0.2$ but at $\delta = 0.3$ the dotted line shifts towards the right and again shifts towards the left for $\delta = 0.4$ and $\delta = 0.5$. From the Figures 5.1, 5.2 and 5.3, it seems like the power function obtained due to our proposed test statistic T_n is not monotonically increasing. We see a similar phenomenon for the powers of hte Kolmogorov-Smirnov and Cramer von-Mises tests as well. For small sample sizes, the simulation results are a bit unstable and for that reason, we do not see any monotonicity of the power with the sample size n for different δ. However, for a fixed δ, this monotonicity is evident from Table 5.3. We would also like to mention that increasing the value of k in computing the test statistic T_n would lead to a greater accuracy in the power calculations, however, due to computational time complexity, we had to restrict k to 10 only for simulation purposes.

5.2.2 Multivariate t Alternatives

Now we compute the power of the proposed test based on T_n under alternatives of multivariate t-distributions with degrees of freedom $3, 5, 10$ and 20 for $d = 2$ and $d = 3$. The power of the proposed test is computed in the same way as described in Section 5.2.1 and are presented in Table 5.8.

We know as the degree of freedom increases in a t distribution, it converges to a normal distribution. From the Table 5.8 we can see that the power decreases as the degrees of freedom of the t distribution increases. In Tables 5.9 and 5.10 we present the power of the

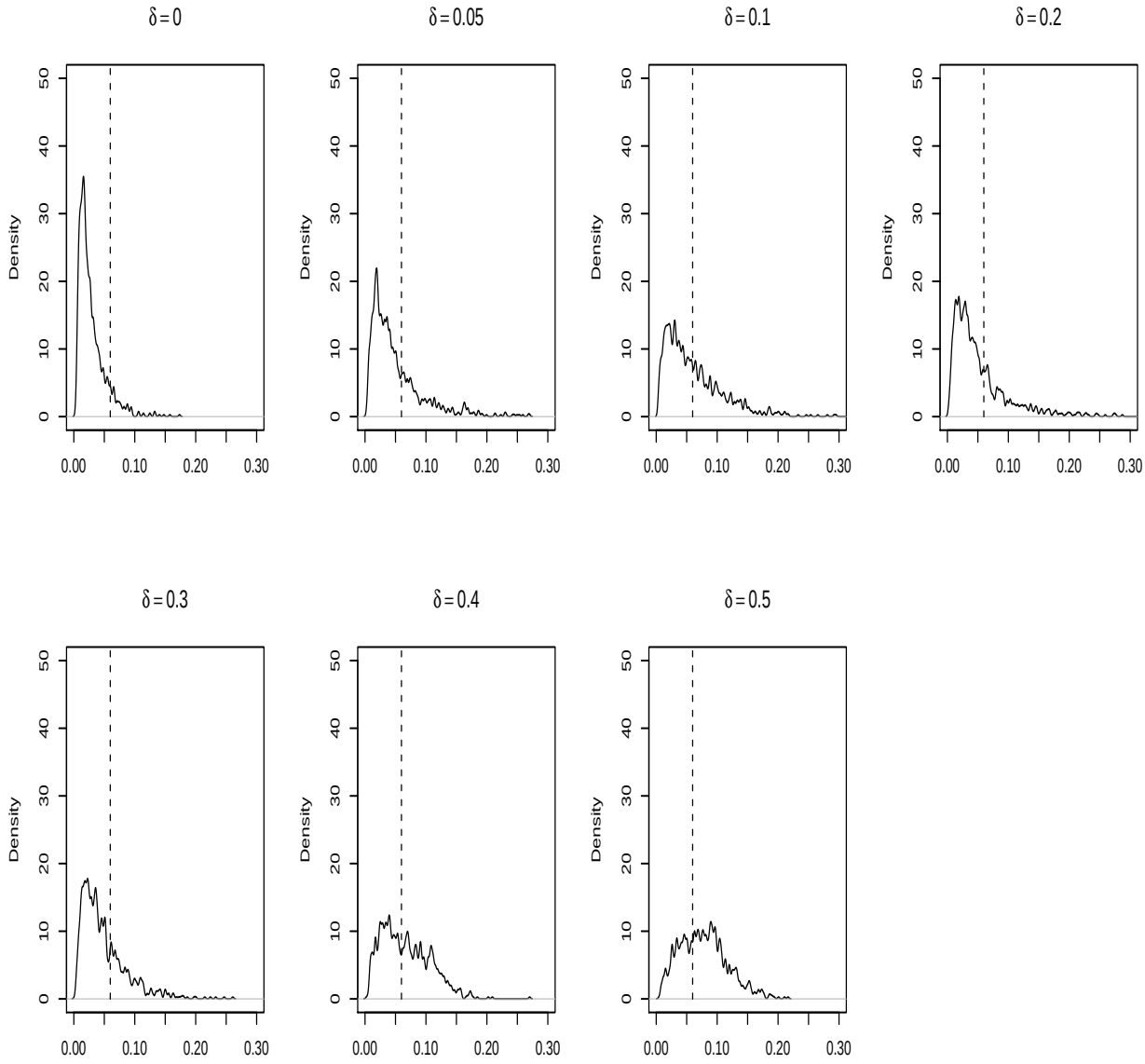

Figure 5.2: Density Plot for the Test Statistic T_n for $n = 50$ and $d = 2$. The dotted line gives the 95-th percentile of the T_n values based on the 1000 simulated values of T_n for $\delta = 0.0$

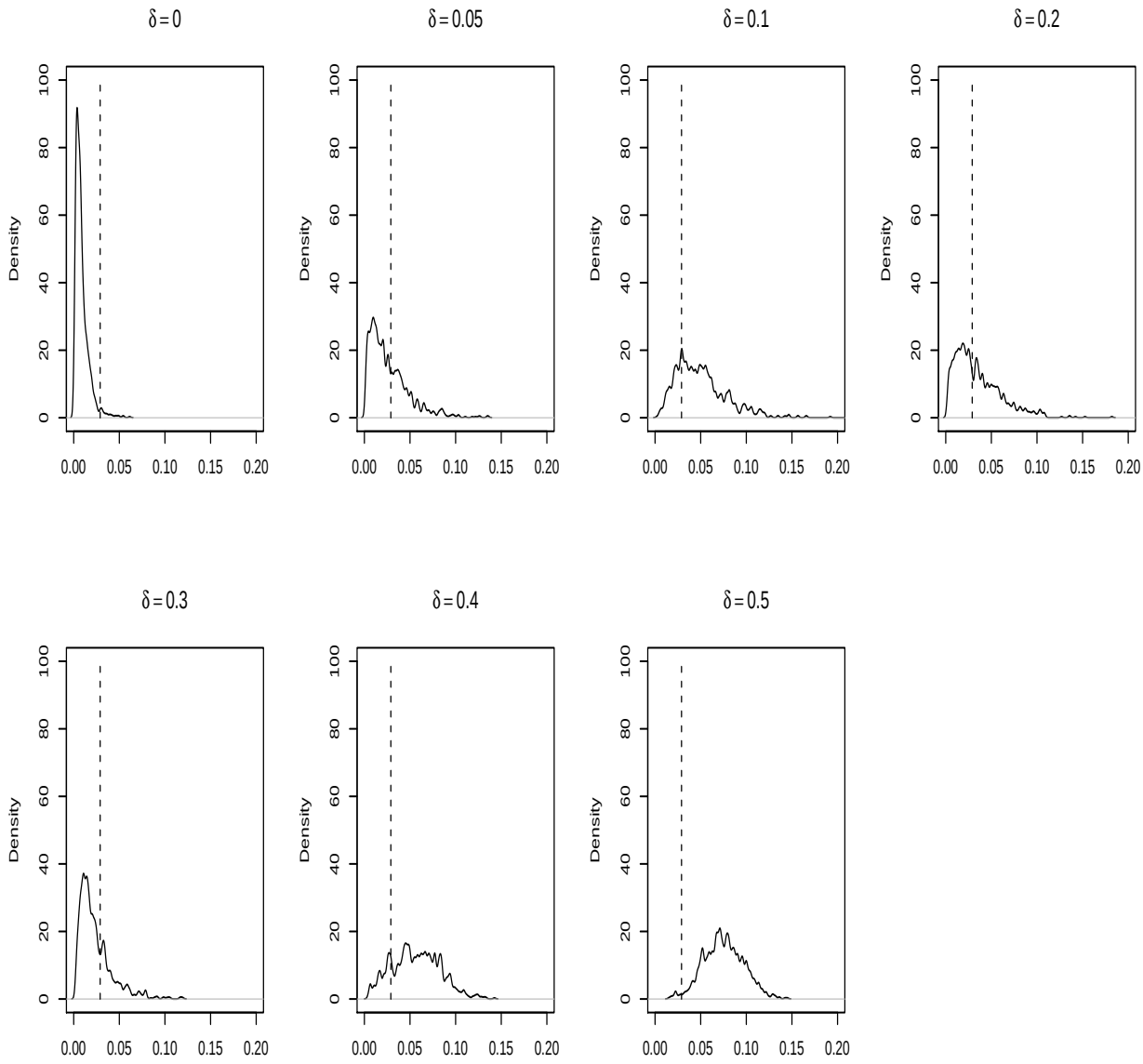

Figure 5.3: Density Plot for the Test Statistic T_n for $n = 100$ and $d = 2$. The dotted line gives the 95-th percentile of the T_n values based on the 1000 simulated values of T_n for $\delta = 0.0$

Table 5.8: Power of the proposed test, under alternatives of bivariate and trivariate t with 3, 5, 10 and 20 degrees of freedom, based on 1000 iterations.

Alternative Distribution		3 df	5 df	10 df	20 df
Bivariate t	$n = 30$	0.364	0.130	0.030	0.036
	$n = 50$	0.610	0.225	0.052	0.040
	$n = 100$	0.881	0.437	0.106	0.052
Trivariate t	$n = 30$	0.613	0.314	0.106	0.061
	$n = 50$	0.856	0.503	0.214	0.104
	$n = 100$	0.978	0.736	0.275	0.132

test with t distribution as alternative for the Kolmogorov-Smirnov test and the Cramer von-Mises test respectively, in the sense of Malkovich and Afifi (1973).

Table 5.9: Power of the Kolmogorov-Smirnov test under alternatives bivariate and trivariate t with 3, 5, 10 and 20 degrees of freedom, based on 1000 iterations.

Alternative Distribution		3 df	5 df	10 df	20 df
Bivariate t	$n = 30$	0.443	0.181	0.060	0.039
	$n = 50$	0.594	0.215	0.051	0.022
	$n = 100$	0.934	0.597	0.204	0.084
Trivariate t	$n = 30$	0.150	0.008	0.005	0.009
	$n = 50$	0.718	0.286	0.078	0.042
	$n = 100$	0.985	0.707	0.212	0.080

Comparing Tables 5.8, 5.9 and 5.10 we can see that for bivariate case both of the Kolmogorov-Smirnov and the Cramer von-Mises tests perform slightly better than the proposed test. In the trivariate case, the proposed test performs slightly better than the Kolmogorov-Smirnov test but not as good as the Cramer von-Mises test.

5.3 Tests for Other Multivariate Distributions

We can extend the proposed methodology for testing multivariate normality to test for other elliptically symmetric distributions as well. To illustrate, we consider $H_0 : F \in$ the

Table 5.10: Power of the Cramer von-Mises test under alternatives bivariate and trivariate t with 3, 5, 10 and 20 degrees of freedom, based on 1000 iterations.

Alternative Distribution		3 df	5 df	10 df	20 df
Bivariate t	$n = 30$	0.616	0.329	0.129	0.088
	$n = 50$	0.809	0.426	0.154	0.078
	$n = 100$	0.987	0.789	0.335	0.116
Trivariate t	$n = 30$	0.305	0.054	0.026	0.041
	$n = 50$	0.864	0.484	0.146	0.081
	$n = 100$	0.995	0.841	0.329	0.126

Table 5.11: The values of $c_n(\alpha)$ for the proposed test for bivariate and trivariate standard Laplace distributions as the null distributions where $\alpha = 0.05$ and $n = 30, 50, 100$ after 1000 iterations.

Null Distribution	$n = 30$	$n = 50$	$n = 100$
Bivariate Laplace Distribution	0.107296	0.066518	0.033765
Trivariate Laplace Distribtuion	0.147575	0.106263	0.064838

set of all d-variate Laplace distributions. As before we need to compute the percentile points of the test statistic T_n. Instead of using the asymptotic distribution we simulate the 95-th percentile $c_n(0.05)$ for finite samples of a simulated distribution of T_n. This is a similar method to the multivariate normal distribution case.

We use a similar algorithm for obtaining the 95-th percentile values for the test statistic for the scale test T_n as described in Section 5.1. The 95-th percentile values for T_n are calculated for multivariate Laplace distribution for dimensions $d = 2, 3$ and for sample sizes $n = 30, 50, 100$ and level of significance, $\alpha = 0.05$. In this case F_0 is a standard multivariate Laplace distribution, and we compute the value of the volume functional V_{F_0} for F_0 using (3.54) of Chapter 3.

In Table 5.11 we give the 95-th percentile values for bivariate and trivariate Laplace distributions as null distributions.

Table 5.12: Power of the proposed test under mixture alternatives bivariate Laplace with $\mu = (5,5)^T$, and trivariate Laplace with $\mu = (5,5,5)^T$.

		δ						
Distribution		0.0	0.05	0.1	0.2	0.3	0.4	0.5
Bivariate Laplace	$n = 30$	0.049	0.072	0.083	0.095	0.267	0.552	0.666
	$n = 50$	0.055	0.072	0.090	0.093	0.380	0.799	0.932
	$n = 100$	0.054	0.095	0.139	0.125	0.581	0.967	0.999
Trivariate Laplace	$n = 30$	0.051	0.090	0.111	0.082	0.081	0.100	0.132
	$n = 50$	0.044	0.093	0.120	0.097	0.066	0.138	0.196
	$n = 100$	0.055	0.144	0.170	0.091	0.074	0.225	0.390

Power Under Mixture Laplace Alternatives

We now compute the power of the proposed test under mixture alternatives for the multivariate Laplace distribution using a similar algorithm as described in Section 5.1. We use the mixture model as our alternative, where the alternative distribution is

$$F = (1 - \delta)F_0 + \delta F_1,$$

where F_0 is a standard multivariate Laplace distribution and F_1 is another Laplace distribution with a different mean vector. Both F and F_0 have identity scale matrix. As before we first vary the value of δ to explore the various degrees of departure from standard Laplace distribution. We take $\delta = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ and calculate the powers under the respective alternative mixture models which are represented in Table 5.12. $\delta = 0.0$ is the null case and it gives us the level of the test.

From Table 5.12 we can see that the power in the bivariate case gradually increases as the value of δ increases, that is, the degree of departure from standard Laplacian increases. In the trivariate case we can see that the power first increases, then it decreases around $\delta = 0.2, 0.3$ and then it again increases.

Finally, as in the previous section, we fix $\delta = 0.1$ and vary $\boldsymbol{\mu} = (r, r, r)^T$ where $r = 1.0, 2.0, 3.0, 4.0, 5.0, 10.0$ to obtain the power of the proposed test. $\boldsymbol{\mu} = (0, 0, 0)^T$ gives

Table 5.13: Power of the proposed test under mixture alternatives bivariate Laplace with $\boldsymbol{\mu} = (r, r)^T$, and trivariate normal with $\boldsymbol{\mu} = (r, r, r)^T$ and $\delta = 0.1$.

		r						
Distribution		0.0	1.0	2.0	3.0	4.0	5.0	10.0
Bivariate Laplace	$n = 30$	0.049	0.036	0.048	0.050	0.064	0.083	0.237
	$n = 50$	0.055	0.046	0.042	0.052	0.065	0.075	0.352
	$n = 100$	0.054	0.047	0.057	0.056	0.066	0.142	0.720
Trivariate Laplace	$n = 30$	0.056	0.048	0.061	0.060	0.070	0.111	0.392
	$n = 50$	0.044	0.036	0.049	0.047	0.065	0.120	0.485
	$n = 100$	0.055	0.044	0.033	0.062	0.089	0.170	0.721

us the size of our proposed test. From Table 5.13 we can see that in bivariate case our proposed test can detect shift when $\boldsymbol{\mu} = (4.0, 4.0, 4.0)^T$ and in the trivariate case it can detect shift from $\boldsymbol{\mu} = (3.0, 3.0, 3.0)^T$.

CHAPTER 6

A TEST OF ELLIPTICAL SYMMETRY

Elliptical symmetry is one of the best known notions of multivariate symmetry and is well studied in the literature. There are various approaches to perform a test of elliptical symmetry. We now briefly overview some of the more well-known works in this field.

In one of the earliest works on tests for elliptical symmetry, Beran (1979) developed a rank-based test. Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are independent random samples from $\mathbb{R}^d$ where each $\mathbf{X}_i$ has density of the form $|\Sigma|^{-\frac{1}{2}} h(\Sigma^{-1}(\mathbf{x} - \boldsymbol{\theta}))$. Beran's test statistic B_n for testing $H_0(\boldsymbol{\theta}, \Sigma)$ is based on scaled residuals. For simplicity, we consider $\mathbf{Z}_i = S_n^{-\frac{1}{2}}(\mathbf{X}_i - \overline{\mathbf{X}})$. Let R_i be the univariate ranks, divided by $(n+1)$, of the distances $\{||\mathbf{Z}_i||\}$ and

$$\mathbf{u}_i = \frac{\mathbf{Z}_i}{||\mathbf{Z}_i||}. \tag{6.1}$$

Consider $\{a_k : k \geq 1\}$ to be a family of functions which are orthonormal with respect to the Lebesgue measure on $[0, 1]$ and orthogonal to the constant function on $[0, 1]$. Also consider $\{b_m : m \geq 1\}$ to be a family of functions orthonormal with respect to the uniform distribution on the unit sphere in d-dimension and orthogonal to the constant function on the unit sphere in d-dimension. Then Beran's test statistic B_n is

$$B_n = \sum_{k=1}^{K_n} \sum_{m=1}^{M_n} \left[n^{-\frac{1}{2}} \sum_{i=1}^{n} a_k(R_i) b_m(\mathbf{u}_i) \right]^2,$$

for suitable choices of K_n, M_n, which are functions of n, increasing with n. Large values

of B_n would lead to the rejection of H_0. Properly scaled, Beran's test statistic has an asymptotically normal distribution depending on the rate of convergence of K_n and M_n.

Beran (1979) proposed another test statistic U_n for this test, given by

$$U_n = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} w_{k,m}^2 \left[n^{-\frac{1}{2}} \sum_{i=1}^{n} a_k(R_i) b_m(\mathbf{u}_i) \right]^2,$$

where $\{w_{k,m}\}$ are the weights such that $\sum_{k,m} w_{k,m}^2 < \infty$ and they are independent of n. This test is a generalisation of the Cramer-Von Mises test, but it has a complex asymptotic distribution.

In his review of elliptically symmetric distributions Chmielewski (1981) reviewed, among other properties of the elliptically symmetric distributions, some tests of elliptical symmetry until his period.

Beran's test statistics, as defined above, are not typically affine invariant, which is a much desired property in a test statistic for elliptical symmetry. More recently various tests for elliptically symmetric distributions have been developed by, among others, Li, Fang and Zhu (1997), Koltchinskii and Sakhanenko (2000), Manzottia, Perez and Quiroz (2002), Schott (2002) and Huffer and Park (2007), all of which satisfy this property. Also see Zhu and Neuhaus (2003) for a test based on the empirical characteristic function.

Li, Fang and Zhu (1997) proposed some robust statistics and related plots to visually test for spherical and elliptical symmetry. They also suggested some numerical measures of detecting deviation from spherical or elliptical symmetry that could be used along with their proposed plots. They defined $t(\mathbf{X}_i) = \frac{\sqrt{d}\,\overline{X}_{i\cdot}}{s_{i\cdot}}$, where $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{id})$, $\overline{X}_{i\cdot} = \frac{1}{d}\sum_{j=1}^{d} X_{ij}$ and $s_{i\cdot}^2 = \frac{1}{d-1}\sum_{j=1}^{d}(X_{ij} - \overline{X}_{i\cdot})^2$, which are distributed independently as t with (d-1) degrees of freedom when $\mathbf{X}_i$ are independent standard d-variate normals. They argued that $t(\cdot)$ was a robust transformation for the class of spherically symmetric distributions and hence suggested a plot of ordered $t(\mathbf{X}_i)$'s against appropriate quantiles of the t_{d-1} distribution, which would be close to the 45^0 line through the origin when the

distribution is spherically symmetric. For elliptically symmetric situations, they suggested looking at standardised observations $\mathbf{Z}_i$'s as defined before and consider $t(\mathbf{Z}_i)$ as their test statistic which they claim would be close to t distribution with (d-1) degrees of freedom for sufficiently large n.

Koltchinski and Sakhanenko (2000) proposed a general class of tests: for a function f on $\mathbb{R}^d$, they defined $m_f(\rho) = E(fU)$ where U is uniformly distributed on the sphere of radius ρ around the origin; they took $\mathcal{F}$, a class of functions closed under orthogonal transformations and considered test statistics of the form

$$sup_{f\in\mathcal{F}} n^{-\frac{1}{2}} \sum_{i=1}^{n} (f(Z_i) - m_f(|Z_i|)).$$

The tests reject the null hypothesis of elliptical symmetry for large values of the test statistic. However, they do not have an asymptotic distribution for their test statistic even under the simplest situation of multivariate standard normal distribution and rely on bootstrap-based procedures to obtain the critical values.

The test statistic Schott (2002) proposed for testing elliptical symmetry is based on fourth moments of the scaled data. The test statistic is given by

$$T_1 = n\left[\hat{\beta}_1 \ \mathrm{tr}(\hat{M}_{4*}^2) + \hat{\beta}_2 \ \mathrm{vec}(I_d)^T \hat{M}_{4*}^2 \ \mathrm{vec}(I_d) - \left\{3\hat{\beta}_1 + (m+2)\hat{\beta}_2\right\} d(d+2)(1+\hat{\kappa})^2\right],$$

where $\hat{\beta}_1 = \frac{1+\hat{\xi}}{24}$, $\hat{\beta}_2 = -3a\left\{24(1+\hat{\xi})^2 + 12(d+4)a(1+\hat{\xi})\right\}^{-1}$, $a = (1+\hat{\xi}) + (1+\hat{\kappa})^3 - 2(1+\hat{\kappa})(1+\hat{\eta})$, $\eta = 8\psi^{(3)}(0) - 1$, $\xi = 16\psi^{(4)}(0) - 1$, $\kappa = 4\psi^{(2)}(0) - 1$, ψ being the characteristic function. $\hat{M}_{4*}$ is the centred and normalised fourth sample moment and $\hat{\kappa}$, $\hat{\xi}$, $\hat{\eta}$ are sample estimates of κ, ξ, η. See Schott (2002) for the exact expressions of these estimates.

The basic idea of Schott's test is to use a Wald-type test to compare the sample fourth moments of the scaled data to the fourth moments expected for an elliptically symmetric distribution. This procedure requires the underlying distribution to have finite moments

up to order 8. Under this condition the test statistic T_1 is asymptotically $\chi^2_{\nu_d}$ with

$$\nu_d = d^2 + \frac{d(d-1)(d^2+7d-6)}{24} - 1.$$

Manzottia, Perez, Quiroz (2002) proposed a test statistic based on the averages of spherical harmonics over the projections of the scaled residual of the d-dimensional data on the unit d-dimensional sphere. They developed a test statistic based on the $\mathbf{u}_i$s, $i = 1, \cdots, n$, as defined earlier in (6.1), which would be asymptotically uniformly distributed on the d-dimensional unit sphere. If h is a spherical harmonic function, "the deviation from non-uniformity in the direction of h" is measured by

$$Q_n(h) = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{u}_i) I\left(|\mathbf{Z}_i|^2 > \rho_{n\epsilon}\right),$$

where $\rho_{n\epsilon}$ is the ϵ sample quantile of $|\mathbf{Z}_1|^2, \cdots, |\mathbf{Z}_n|^2$. The indicator function deletes the fraction ϵ of the data having the smallest radii to avoid residuals falling near the origin "for avoiding extra assumptions" (Manzottia, Perez, Quiroz (2002)). They combined these individual deviations to form the test statistic given by

$$Z_n^2 = n \sum_{h \in \mathcal{I}_{jl}} Q_n^2(h),$$

where $j \geq 3$, $\mathcal{I}_{jl} = \cup_{j \leq i \leq l} H_i$, H_i iss the set of spherical harmonics in an orthonormal basis of the d-dimensional unit sphere as obtained by Muller (1966). The asymptotic distribution of this test statistic was shown to be $(1-\epsilon)\chi^2_\nu$ where ν is the number of functions in $\mathcal{I}_{ji}$.

Huffer and Park (2007) developed a test statistic X^2 based on Pearsons chi-square statistic which is calculated after transforming the data into a spherical form. To compute the test statistic X^2, the space of $\mathbf{Z}_i$'s is divided into c spherical shells centred at the origin with each shell containing an equal number of $\mathbf{Z}_i$'s. Then $\mathbb{R}^d$ is divided into g sectors

coming out from the origin, congruent to an orthogonal transformation mapping: hence $\mathbb{R}^d$ is divided into gc congruent sectors asymptotically expected to contain equal number of $\mathbf{Z}_i$'s. The test statistic X^2 is the chi-squared test statistic comparing the observed counts in each shell with the expected frequencies. They showed that the distribution X^2 converges asymptotically to a weighted linear combination of independent χ^2 variables. For the exact form of the asymptotic distribution see Huffer and Park (2007).

In this chapter, we generalise the test for multivariate normality proposed in Chapter 5 to a test of elliptical symmetry. The proposed procedure provides a graphical tool to check for elliptical symmetry.

6.1 Non-elliptical Distributions

In the previous chapters, we have restricted our discussions of the scale-scale plots to the elliptically symmetric distributions. We now consider the situations when the distributions are not necessarily elliptical and develop a test of elliptical symmetry based on the scale-scale curve. As illustrative examples we look into the scale-scale plots for the bivariate gamma distribution and Tukey's g and h distribution for different parametric values. For more discussion on non-elliptical distributions one can see Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999).

In Figure 6.1, we compare samples from simulated bivariate gamma distributions with the standard bivariate normal distribution. A sample of size $n = 1000$ is simulated from the bivariate gamma density

$$f(x_1, x_2) = \frac{1}{\{\lambda^\alpha \Gamma(\alpha)\}^2} e^{-(x_1+x_2)/\lambda} (x_1 x_2)^{\alpha-1}, x_1, x_2 \geq 0$$

for different values of α and λ, so that the mean vectors of the distributions remain the same. For small values of α, bivariate gamma is a highly skewed distribution and their scale-scale plot compared to the bivariate normal is a nonlinear curve with scales, $V_G(p)$,

increasing sharply with higher values of p. However, as α increases, the scale-scale plots become more linear and we observe that the scale-scale plot for the sample from bivariate gamma with $\alpha = 10$ and $\lambda = 0.1$ is very close to the 45° line, which is in line with the distributional convergence of the gamma distribution to the normal distribution as $\alpha \to \infty$.

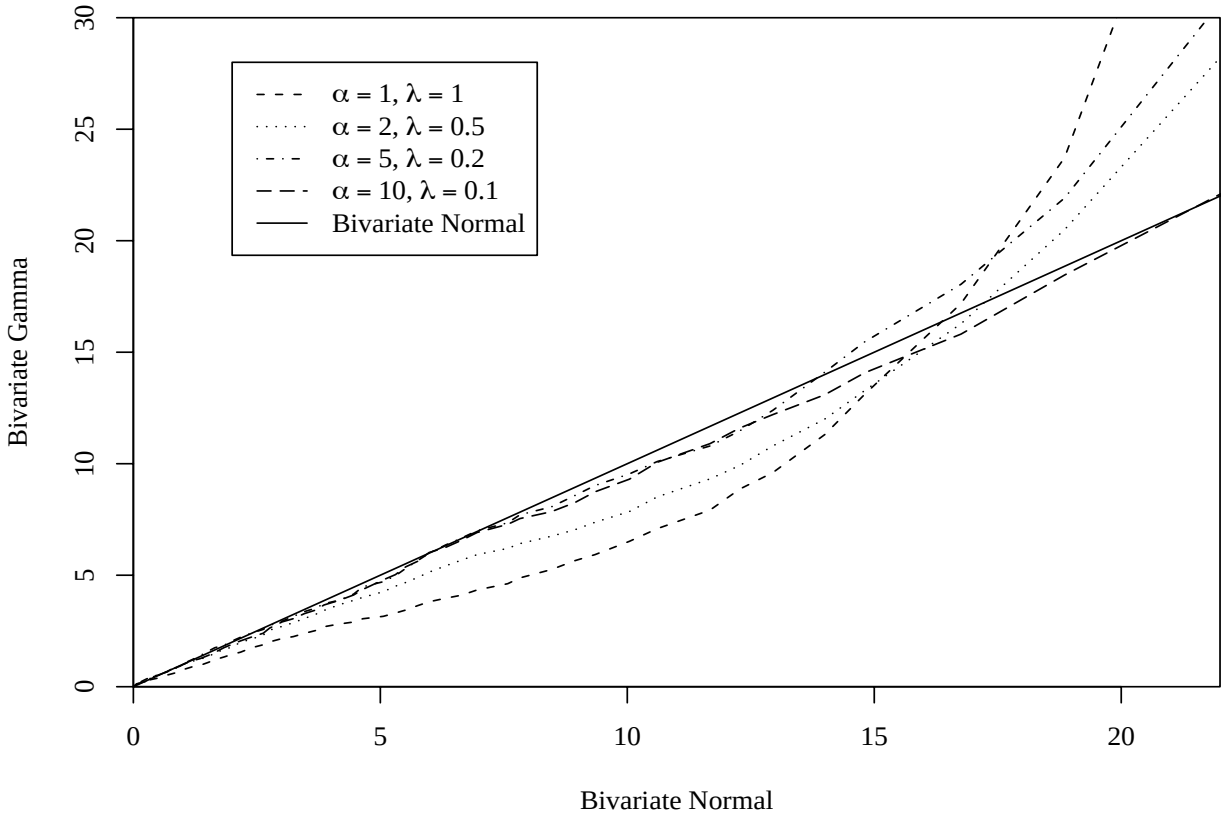

Figure 6.1: Scale-Scale plots comparing bivariate gamma distribution with the standard bivariate normal distribution.

Another example of non-elliptical distributions is the g and h distribution, introduced by Tukey in 1977 (see Field, Genton (2006)). For more discussion on the univariate g and h distribution one can see Hoaglin (1983). It is generated by a single transformation of the standard normal, which allows for a symmetry and heavier tails. When a standard

normal random variable Z is transformed to $X = \psi_{g,h}(Z)$, where

$$\psi_{g,h}(Z) = \frac{(e^{gZ} - 1)e^{\frac{hZ^2}{2}}}{g} \tag{6.2}$$

with $h \geq 0$, then the resulting random variable X is said to have a g and h distribution. In the multivariate setup a random vector $\mathbf{X} \in \mathbb{R}^d$ would have a standard multivariate g and h distribution, where $\mathbf{g} = (g_1, g_2, \cdots, g_d)^T \in \mathbb{R}^d$ and $\mathbf{h} = (h_1, h_2, \cdots, h_d)^T \in \mathbb{R}^d_+$, if

$$\mathbf{X} = (\psi_{g_1,h_1}(Z_1), \psi_{g_2,h_2}(Z_2), \cdots, \psi_{g_d,h_d}(Z_d))^T = \boldsymbol{\psi}_{\mathbf{g},\mathbf{h}}(\mathbf{Z}) \tag{6.3}$$

where $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_d)^T \sim \mathbf{N}_d(\mathbf{0}, \mathbf{I}_d)$. The parameter $\mathbf{g}$ is responsible for the skewness and the parameter $\mathbf{h}$ controls kurtosis of the distribution. A discussion about the multivariate g and h distribution can be found in Field and Genton (2006).

In Figure 6.2 we are comparing the scale-scale plots of standard bivariate normal distribution and bivariate g and h distributions. For our simulation we have taken $g_1 = g_2 = g$ and $h_1 = h_2 = h$. Samples of size $n = 1000$ are simulated from the bivariate g and h distribution for different values of g and h. As g and h get closer to 0, the scale-scale plots become more linear. We observe that the scale-scale plot for the sample from bivariate g and h distribution with $g = 0.1$ and $h = 0.1$ is very close to the 45° line.

6.2 Test of Multivariate Elliptical Symmetry Based on the Proposed Test Statistic

In the previous chapter we mainly discussed spherically symmetric distributions and tests to detect multivariate normality. We now suggest a test to detect multivariate elliptically symmetric distributions, which is based on the following idea:

Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be a random sample from a distribution F with location vector $\boldsymbol{\mu}$ and

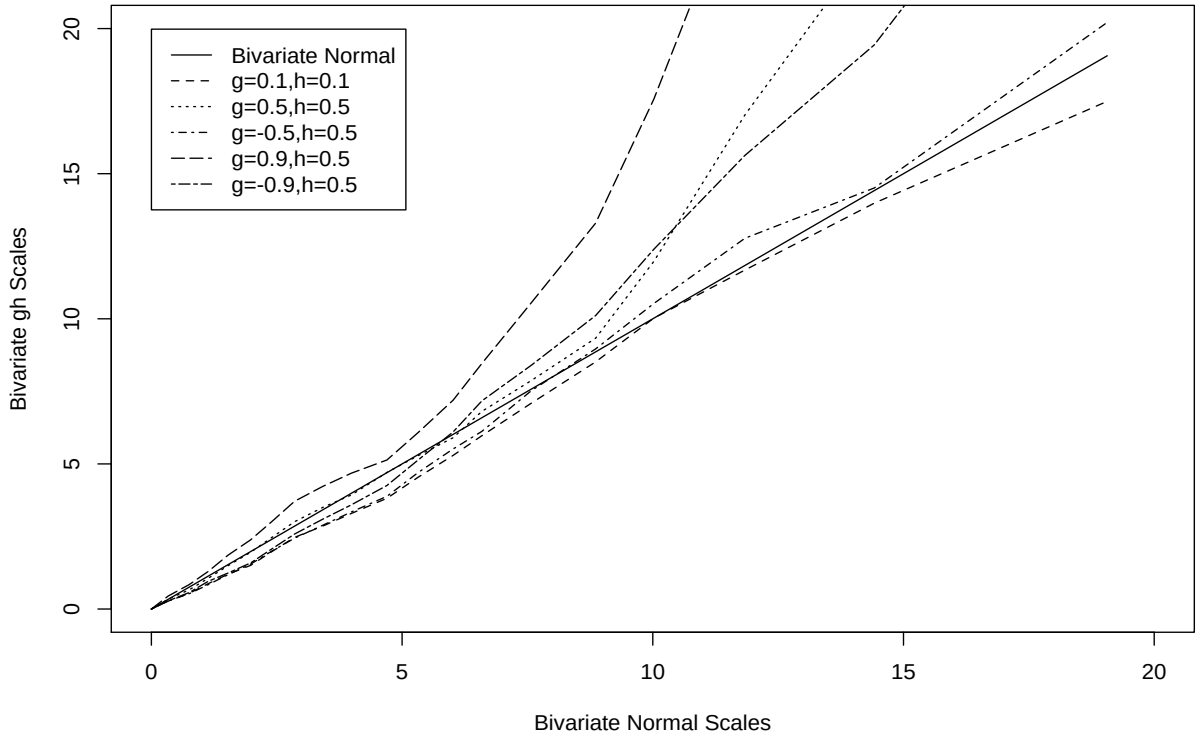

Figure 6.2: Scale-Scale plots comparing bivariate gh distribution with the standard bivariate normal distribution.

scale matrix Σ. Under the null hypothesis of elliptical symmetry of F,

$$\mathbf{U}_i = \frac{\mathbf{X}_i - \boldsymbol{\mu}}{||\Sigma^{-\frac{1}{2}}(\mathbf{X}_i - \boldsymbol{\mu})||}$$

is uniformly distributed over the ellipsoid determined by Σ and independent of the length of $||\Sigma^{-\frac{1}{2}}(\mathbf{X}_i - \boldsymbol{\mu})||$ and hence if we obtain a sample $r_1, \cdots, r_n$ from the χ^2 distribution with d degrees of freedom, independent of the $\mathbf{U}_i$s, then $\mathbf{Y}_i = \sqrt{r_i}\mathbf{U}_i$s are independent multivariate normal random variables.

Therefore, a test of multivariate normality based on the $\mathbf{Y}_i$s should work as a test of elliptical symmetry. One may, for example, employ the test of multivariate normality

based on the volume functional-based test statistic T_n as defined in (5.5).

In the remainder of this section, we present two power studies for our proposed test against the alternatives of bivariate gamma and bivariate g and h distribution for sample sizes $n = 30, 50$ and 100 for various choices of parameters at a 5% level of significance.

However, before proceeding to the power study, we note that in practice one usually does not know $\boldsymbol{\mu}$ or Σ, and hence instead of the $\mathbf{U}_i$s it is often required to use

$$\mathbf{u}_i = \frac{\mathbf{X}_i - \hat{\boldsymbol{\mu}}}{||\hat{\Sigma}^{-\frac{1}{2}}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})||}$$

based on some estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ of $\boldsymbol{\mu}$ and Σ respectively. As a result the tests are performed using

$$\mathbf{Y}_i = \sqrt{r_i}\mathbf{u}_i,$$

for which we expect some departure of normality, especially when the sample size is small. Hence instead of using cut off from Table 5.1 which were derived for normal Y_is we need re-estimating the cut off values by a re-sampling technique, as described below, for better accuracy.

We now provide an algorithm for obtaining cut off values for T_n as defined in (5.5) to perform a goodness of fit test for multivariate elliptically symmetric distributions at a pre-specified level of significance.

1. Generate a random sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ from the standard multivariate normal density.

2. Estimate the location vector $\boldsymbol{\mu}$ and scale matrix Σ by some method of estimation. In our simulation we use the usual sample mean and covariance. Compute the $\mathbf{u}_i$s, $i = 1, \cdots, n$.

3. Simulate an auxiliary independent sample $r_1, \cdots, r_n$ from the χ^2 distribution with d degrees of freedom and obtain $\mathbf{Y}_i = \sqrt{r_i}\mathbf{u}_i$ for $i = 1, \cdots, n$. Compute the corre-

Table 6.1: The values of cut off at 5% level, $c_n(0.05)$, for scale test for bivariate standard normal distributions as the null distributions where $n = 30, 50, 100$ based on 1000 iterations.

	$n = 30$	$n = 50$	$n = 100$
$c_n(0.05)$	0.071834	0.043736	0.021732
Size of the test	0.049	0.051	0.053

sponding volume functional $V_{F_n}(p_i)$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$.

4. Compute the slope functional $s_n(p_i) = \frac{V_{F_n}(p_i)}{V_{F_0}(p_i)}$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$, using (5.1) and then compute the test statistic T_n for each i using (5.5). Here F_0 is the standard multivariate normal distribution and $V_{F_0}(p_i)$ can be computed using (3.54).

5. Repeat steps 1-4 1000 times and get 1000 values of T_n and order them.

6. Take the 95-th percentile value of the ordered T_n as our cut off value $c_n(0.05)$.

In Table 6.1 we give a set of cut-off values obtained using the above algorithm. The computer program we have used for simulation was written in C language.

When we estimate the size based on 1000 simulation each and for the 5% level cut offs in Table 6.1, for $n = 30, 50, 100$ we observe the estimated sizes of $0.049, 0.051$ and 0.053 respectively, which seem reasonably accurate.

We are now ready to provide the steps required for estimating the power of the test against some pre-specified alternative distribution, say F_A with location vector $\boldsymbol{\mu}$ and scale matrix Σ, for a sample size n:

1. Generate a random sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ from F_A.

2. Estimate the location vector $\boldsymbol{\mu}$ and scale matrix Σ take

$$\mathbf{u}_i = \frac{\mathbf{X}_i - \hat{\boldsymbol{\mu}}}{||\hat{\Sigma}^{-\frac{1}{2}}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})||}.$$

3. Simulate an auxiliary independent sample $r_1, \cdots, r_n$ from the χ^2 distribution with d degrees of freedom and obtain $\mathbf{Y}_i$s for $i = 1, \cdots, n$. Compute the corresponding volume functional $V_{F_n}(p_i)$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$.

4. Compute the slope functional $s_n(p_i) = \frac{V_{F_n}(p_i)}{V_{F_0}(p_i)}$ for $p_i = \frac{i}{(k+1)}, i = 1, \cdots, k$, using (5.1) and then compute the test statistic T_n for each i using (5.5). Here F_0 is the standard multivariate normal distribution.

5. Repeat steps 1-4 1000 times and get 1000 values of T_n.

6. Estimate the power of the test for F_1 for sample size n at 5% level by the proportion of T_ns greater than the cut off value $c_n(0.05)$.

Table 6.2: Power of test of elliptical symmetry under bivariate gamma alternatives based on 1000 iterations with new cut off values.

		$\alpha = 1$	$\alpha = 2$	$\alpha = 5$	$\alpha = 10$
Distribution		$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0.2$	$\lambda = 0.1$
Bivariate Gamma	$n = 30$	0.080	0.049	0.054	0.056
	$n = 50$	0.095	0.063	0.066	0.060
	$n = 100$	0.137	0.080	0.076	0.062

In Table 6.2 we present the power of our proposed test at a 5% level of significance against bivariate gamma alternatives for sample sizes $n = 30, 50, 100$, $\alpha = 1, 2, 5, 10$ and $\lambda = 1/\alpha$. We see from the table that for $n = 30$ and 50, significant shift is detected only at $\alpha = 1$. For $n = 100$, we observe that as the value of α increases, the power of the test decreases. This is consistent with the change in the bivariate gamma distribution with the increasing value of α as bivariate gamma distribution converges to bivariate normal distribution as $\alpha \to \infty$. We can also see that at a specific value of α, the power increases as the sample size n increases from 30 to 100. The power of our proposed test (again, at a 5% level of significance), against bivariate gh alternatives for sample sizes $n = 30, 50, 100$, $(g, h) = (0.1, 0.1), (0.5, 0.5), (-0.5, 0.5), (0.9, 0.5)$ and $(-0.9, 0.5)$, is presented in Table 6.3. We see from the table that the power of the test is negligible for $(g, h) = (0.1, 0.1)$,

Table 6.3: Power of the test of elliptical symmetry under bivariate gh alternatives based on 1000 iterations with new cut off values.

Distribution		g = 0.1 h = 0.1	g = 0.5 h = 0.5	g = -0.5 h = 0.5	g = 0.9 h = 0.5	g = -0.9 h = 0.5
gh	$n = 30$	0.045	0.162	0.046	0.340	0.057
	$n = 50$	0.054	0.238	0.045	0.536	0.067
	$n = 100$	0.059	0.476	0.056	0.846	0.066

$(-0.5, 0.5)$ and $(-0.9, 0.5)$ uniformly across all choices of the sample size. In the remaining cases the power increases as the sample size n increases from 30 to 100. It may be seen that the findings based on our proposed test statistic for relatively smaller sample sizes are consistent with the scale-scale plots in Figure 6.2 for $n = 1000$.

CHAPTER 7

OTHER APPLICATIONS

7.1 A Measure of Tail Weight

Kurtosis is a location and scale free measure of peakedness and tail weight which describes the distribution of the probability mass "from the shoulders of a distribution into its centre and tails" (Balanda and MacGillivary (1990)) where by "shoulders" of a distribution one would mean the points $\mu \pm \sigma$ (Serfling (2004)). It is sometimes also described as "lack of shoulders"(Finucan (1964)). Some authors prefer to distinguish the concepts of peakedness and tail weight and prefer a good measure of kurtosis to be a mixture of measures for both, see Serfling (2004). For the univariate case, most commonly kurtosis is defined as the ratio of the fourth central moment to the square of the second central moment and is denoted by β_2. The value of β_2 for the univariate normal distribution is 3. Sometimes $\gamma_2 = \beta_2 - 3$ is also taken as a measure of kurtosis. Depending on the value of β_2, the probability distribution curves are named platykurtic (if $\beta_2 < 3$), mesokurtic (if $\beta_2 = 3$) and leptokurtic (if $\beta_2 > 3$) depending upon the flatness of the curves with respect to the normal distribution curve. These names were used first by Pearson (1905) and Dyson (1943).

Kurtosis can also be defined in terms of the quantiles. Groeneveld and Meeden (1984) defined a quantile based kurtosis for symmetric univariate distributions. The kurtosis defined in terms of the moment measures the dispersion of the probability mass in the

region of the shoulders of the density function but it does not give any idea about the shape of the distribution. The quantile based kurtosis gives information about the shape of the distribution. Using this definition we would get a high value of kurtosis if the probability mass gets reduced near the shoulders of the distribution.

In perhaps one of the earliest attempts to generalise the concept of kurtosis to a multivariate situation, Mardia (1970) defined the kurtosis for a multivariate distribution in $\mathbb{R}^d$ with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ as

$$\kappa = E\left\{\left[(\mathbf{X}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\right]^2\right\}. \tag{7.1}$$

Malkovitch and Affifi (1973) defined a measure of multivariate kurtosis as

$$\beta_2^{*2} = \max_{\mathbf{c}}\left(\frac{E(\mathbf{c}^T(\mathbf{Y}-E(\mathbf{Y}))^4}{\mathrm{Var}(\mathbf{c}^T\mathbf{Y})^2}-3\right)^2 \tag{7.2}$$

based on which they defined a test of multivariate normality for some vector $\mathbf{c}$. Oja (1983) developed a similar measure of kurtosis in d dimensions based on functions of the volumes of d-dimensional sample simplices. Liu, Parelius and Singh (1999) discussed a visual measure of kurtosis through data-depth-based "fan-plots", which actually gave an idea of heavy tailedness. Serfling (2004) discussed a spatial kurtosis functional based on volumetric considerations, invariant of shift and orthogonal and homogeneous scale transofrmations, that generalised a quantile-based kurtosis functional proposed by Groenveld and Meeden (1984). For some measures of multivariate measures of peakedness, one can also see Olkin and Tong (1988) and Zuo and Serfling (2000).

Balanda and MacGillivary (1990) proposed an ordering based on quantile-based kurtosis, in the sense of van Zwet (1964), which was scale and location free. This ordering involved the idea of spread function $S_F(u)$ of a distribution function F defined by:

$$S_F(x) = F^{-1}(0.5+x) - F^{-1}(0.5-x), \text{ for } 0 \le x < \frac{1}{2}.$$

This measure of spread is location invariant but not scale invariant. Hence, to develop a measure of kurtosis, which would be scale and location free, they proposed using

$$t_{x,y}(F) = \frac{S_F(x)}{S_F(y)}$$

for $0 < x < y < \frac{1}{2}$. This is primarily a measure of tail weight: for a fixed y, $S_F(x)$ and hence $t_{x,y}(F)$ increase more quickly in x for more heavy-tailed distributions, and hence a relatively steeper slope of $t_{x,y}(F)$ for a fixed y is indicative of more tail weight.

Using an idea somewhat similar to Balanda and MacGillivary (1990), we can define a measure of peakedness using the volume determined by the spatial rank regions.

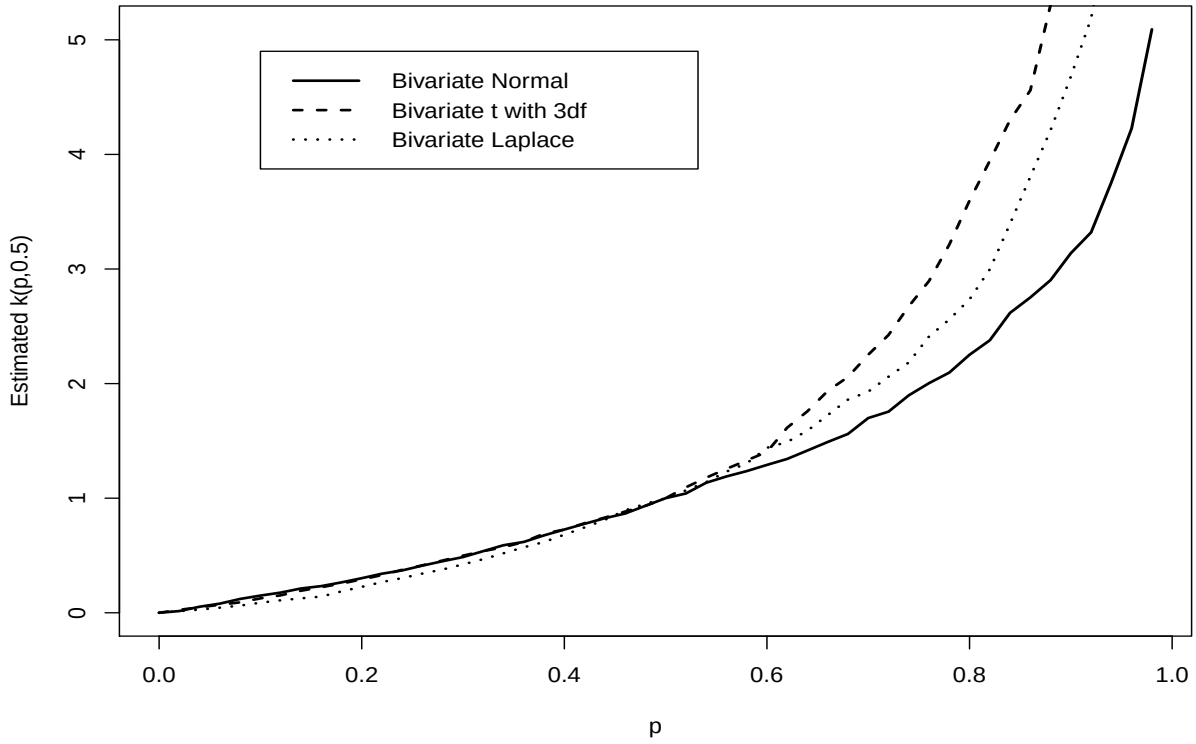

Figure 7.1: Sample Kurtosis plot for Bivariate Laplace, Bivariate t with 3 df and Bivariate Normal Distribution with $q = 0.5$

Definition 7.1.1 *Suppose F is a d-dimensional distribution function. Let $V_F(p)$ denote the volume of the central rank region $C_F(p)$, with respect to the p-th quantile of the distribution of $||R_F(\mathbf{X})||$, where $0 < p < 1$. Similarly define $V_F(q)$ to be the volume of the central rank region $C_F(q)$, with respect to the q-th quantile of the distribution of $||R_F(\mathbf{X})||$, where $0 < q < 1$. Then define,*

$$k(p,q) = \frac{V_F(p)}{V_F(q)}, 0 < p < 1, 0 < q < 1. \tag{7.3}$$

For a fixed q, $V_F(p)$ and hence $k(p,q)$ increases faster in p for more heavy-tailed distributions, and hence should have a relatively steeper slope compared to distributions with less tail weight.

As an illustration, in Figure 7.1 we have plotted estimates of $k(p,q)$, based on 1000 samples, against p, for $0 < p < 1$ and $q = 0.5$, for the standard bivariate normal distribution, the standard bivariate t distribution with 3 degrees of freedom and the standard bivariate Laplace distribution. We choose $q = 0.5$ as the boundary region of $C_F(0.5)$ can be considered as the "shoulder" separating a central region from an outlying tail region, see Serfling (2004) for a discussion. As p moves from 0.0 to 1.0, we get an idea about how the probability mass shifts from the shoulders to the tails: the t distribution with 3 degrees of freedom distribution has the fattest tail compared to the normal and the Laplace distribution as is clearly demonstrated by its steepest slope beyond $q = 0.5$.

7.2 A Visual Test of Location

In this section, we propose a test of location as an application of the proposed scale-scale plots following the ideas of Singh, Tyler, Zhang, and Mukherjee (2009). Let $\mathbf{X}_1, \cdots, \mathbf{X}_n$ be a random sample from a d-dimensional distribution F which is symmetric about $\boldsymbol{\theta} \in \mathbb{R}^d$ in the sense that $\mathbf{X}_i - \boldsymbol{\theta} \stackrel{d}{=} \boldsymbol{\theta} - \mathbf{X}_i$. We are interested to test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Note that under H_0, the distributions of $\mathbf{X}_i$ and

its reflection $2\boldsymbol{\theta}_0 - \mathbf{X}_i$ are identical and the scale-scale plot of the combined sample $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\} \cup \{2\boldsymbol{\theta}_0 - \mathbf{X}_1, \cdots, 2\boldsymbol{\theta}_0 - \mathbf{X}_i\}$ against the original data $\{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$ will be nearly a 45° line. However, if the null hypothesis does not hold, the scale of the combined data will be larger and the scale-scale plot will move away from the 45° line. Using this principle, we construct a test procedure as follows:

1. $$\text{Define } \mathbf{Y}_i = \begin{cases} \mathbf{X}_i & \text{with probability 0.5,} \\ 2\boldsymbol{\theta}_0 - \mathbf{X}_i & \text{with probability 0.5,} \end{cases}$$

   for $i = 1, \cdots, n$.

2. Construct a scale-scale plot of $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ against $\mathbf{X}_1, \cdots, \mathbf{X}_n$.

3. There are 2^n possible samples $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_n\}$. However, it is not practical to construct scale-scale plots for all of them for large n. One can repeat Steps 1 and 2 for a large number of random subsets $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_n\}$ and construct a band of scale-scale plots.

4. If the 45° line is in the bottom 5% of the band of scale-scale plots or below the band altogether, the null hypothesis is rejected.

For illustration, we present plots of the above test procedure in Figure 7.2 where the data are simulated from a bivariate normal distribution of sample size $n = 500$ with $\Sigma = \mathbf{I}$ and different values of the mean $\boldsymbol{\theta}$. Figure 7.2(a) is the plot under null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ and (b), (c) and (d) present the plots for the alternatives $\boldsymbol{\theta} = (0.2, 0.2)^T$, $(0.5, 0.5)^T$ and $(1.0, 1.0)^T$ respectively. We see that for a small shift in location, the scale-scale plot is not that effective in detecting the shift. However, for moderate to large shifts of the location, this visual tool detects the shift quite effectively.

One can use different criterion to formally compute the p-values of the proposed test. We have used the proportion of scale-scale plots in the band below the 45° line for some specific values of p, for example, at $p = 0.50$, which is the comparison of volumes of the

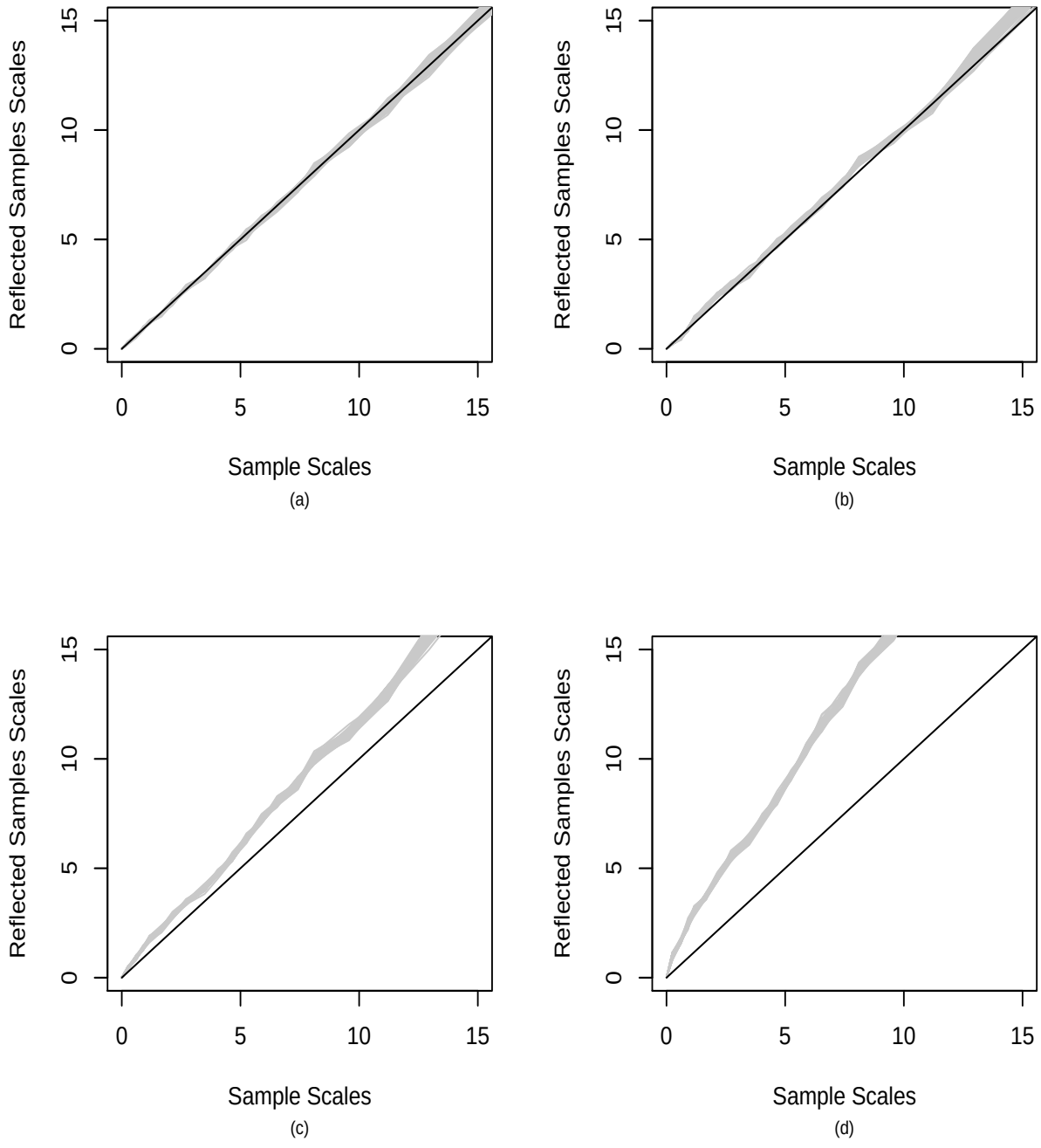

Figure 7.2: Scale-Scale plot for test of location for bivariate normal distribution for $n = 500$. (a) $\boldsymbol{\theta} = (0, 0)^T$, (b) $\boldsymbol{\theta} = (0.2, 0.2)^T$, (c) $\boldsymbol{\theta} = (0.5, 0.5)^T$, (d) $\boldsymbol{\theta} = (1.0, 1.0)^T$

Table 7.1: Finite sample power at the specified values of p for $n = 30$ and $d = 2$. Number of band is 100, simulation size is 1000 and level of significance = 5%. Powers are computed at $\boldsymbol{\theta} = (r, r)^T$ for different values of r. The underlying distribution is a bivariate normal distribution, with $\mathbf{I}_2$ as the scale matrix.

r	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.057	0.068	0.066
0.1	0.065	0.089	0.071
0.2	0.114	0.141	0.138
0.3	0.172	0.234	0.251
0.4	0.294	0.401	0.418
0.5	0.359	0.531	0.594
0.6	0.498	0.659	0.732
0.7	0.611	0.783	0.843
0.8	0.695	0.905	0.925
0.9	0.820	0.951	0.966
1.0	0.878	0.980	0.985

Table 7.2: Finite sample power at the specified values of p for $n = 30$ and $d = 2$. Number of band is 100, simulation size is 1000 and level of significance = 5%. Powers are computed at $\boldsymbol{\theta} = (r, r)^T$ for different values of r. The underlying distribution is a bivariate t distribution with 3df, with scale matrix $\Sigma = \mathbf{I}_2$.

r	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.071	0.071	0.072
0.1	0.075	0.077	0.055
0.2	0.113	0.131	0.100
0.3	0.106	0.206	0.162
0.4	0.280	0.297	0.208
0.5	0.384	0.448	0.318
0.6	0.507	0.578	0.448
0.7	0.619	0.665	0.486
0.8	0.708	0.809	0.577
0.9	0.822	0.881	0.719
1.0	0.867	0.916	0.740

central rank regions containing 50% of the data. Using that criteria, we present a small sample simulation study of the power of the test for different values of p.

In this study we compute the power for three elliptically symmetric distributions, bivariate normal, bivariate Laplace and bivariate t with 3 degrees of freedom with the scale matrix $\Sigma = \mathbf{I}$ and sample size $n = 30$. The null hypothesis is $H_0 : \boldsymbol{\theta} = (0, 0)^T$ and the alternatives are $\boldsymbol{\theta} = (r, r)^T$, for $r = 0.1, 0.2, \cdots, 1.0$. The test is performed based on 100 bands with level of significance 0.05, i.e., if 5 of reflected scale curves at p are below the original sample scale curve at p, then we reject the null hypothesis. The size and power of this test computed based on 1000 simulations for $p = 0.25, 0.50, 0.75$. We observe that the estimated sizes are slightly higher but the powers are nevertheless quite encouraging as they increase with r for every p. Hence instead of looking at the entire plot we can perform the test based on one single p. Perhaps using $p = 0.5$ might be a good idea as the power increases most rapidly for $p = 0.5$ for all three distributions.

Table 7.3: Finite sample power at the specified values of p for $n = 30$ and $d = 2$. Number of band is 100, simulation size is 1000 and level of significance = 5%. Powers are computed at $\boldsymbol{\theta} = (r, r)^T$ for different values of r. The underlying distribution is a bivariate Laplace distribution, with $\mathbf{I}_2$ as the scale matrix.

r	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.053	0.069	0.057
0.1	0.070	0.069	0.070
0.2	0.090	0.104	0.071
0.3	0.167	0.136	0.111
0.4	0.248	0.216	0.165
0.5	0.281	0.280	0.229
0.6	0.418	0.408	0.316
0.7	0.506	0.521	0.379
0.8	0.584	0.595	0.438
0.9	0.689	0.690	0.559
1.0	0.772	0.772	0.618

7.3 A Visual Test of Scale

Suppose $\mathbf{X}_1, \cdots, \mathbf{X}_n$ and $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ are random samples from the same family of elliptically symmetric distributions with possibly different location vectors $\boldsymbol{\theta}$ and scale matrix Σ. Then we have $\mathbf{Y}_i \stackrel{d}{=} \mathbf{A}\mathbf{X}_i + \mathbf{b}$ for some $d \times d$ matrix $\mathbf{A}$ and $d \times 1$ vector $\mathbf{b}$ and as we noted earlier, the scale-scale plot of such competing samples will be close to a straight line. We would like to test $H_0 : |\Sigma_{\mathbf{X}}| = |\Sigma_{\mathbf{Y}}|$ or alternatively, $H_0 : |\mathbf{A}| = 1$ against $H_1 : |\mathbf{A}| \neq 1$. Under H_0, the scale-scale plot will be very close to a 45° line. To construct a graphical test, plot a band of scale-scale plots of G_n^* against F_n^*, where F_n^* and G_n^* are bootstrap distributions of $\mathbf{X}_1, \cdots, \mathbf{X}_n$ and $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$, respectively. Under H_0, this band will lie on both sides of the 45° line, whereas under H_1, nearly all of it will lie above or below the 45° line. For illustration, we present a few plots in Figure 7.3. The data are simulated from the bivariate Laplace distribution with mean $\boldsymbol{\theta} = (0, 0)^T$ and $\mathbf{X}_1, \cdots, \mathbf{X}_n$ have $\Sigma_X = \mathbf{I}_2$. In Figure 7.3(a), the scale matrix for $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ is $\Sigma_{\mathbf{Y}} = \mathbf{I}_2$ and it is $0.25\mathbf{I}_2$, $2\mathbf{I}_2$ and $4\mathbf{I}_2$ in (b), (c) and (d) respectively. We observe that these plots detect the change in scales quite effectively.

One may also construct formal tests based on the asymptotic distribution of the process $\sqrt{n}\{v_F(p)\}^{-1}(s_n(p) - 1)$, which is a standard Brownian bridge as derived in Chapter 3.

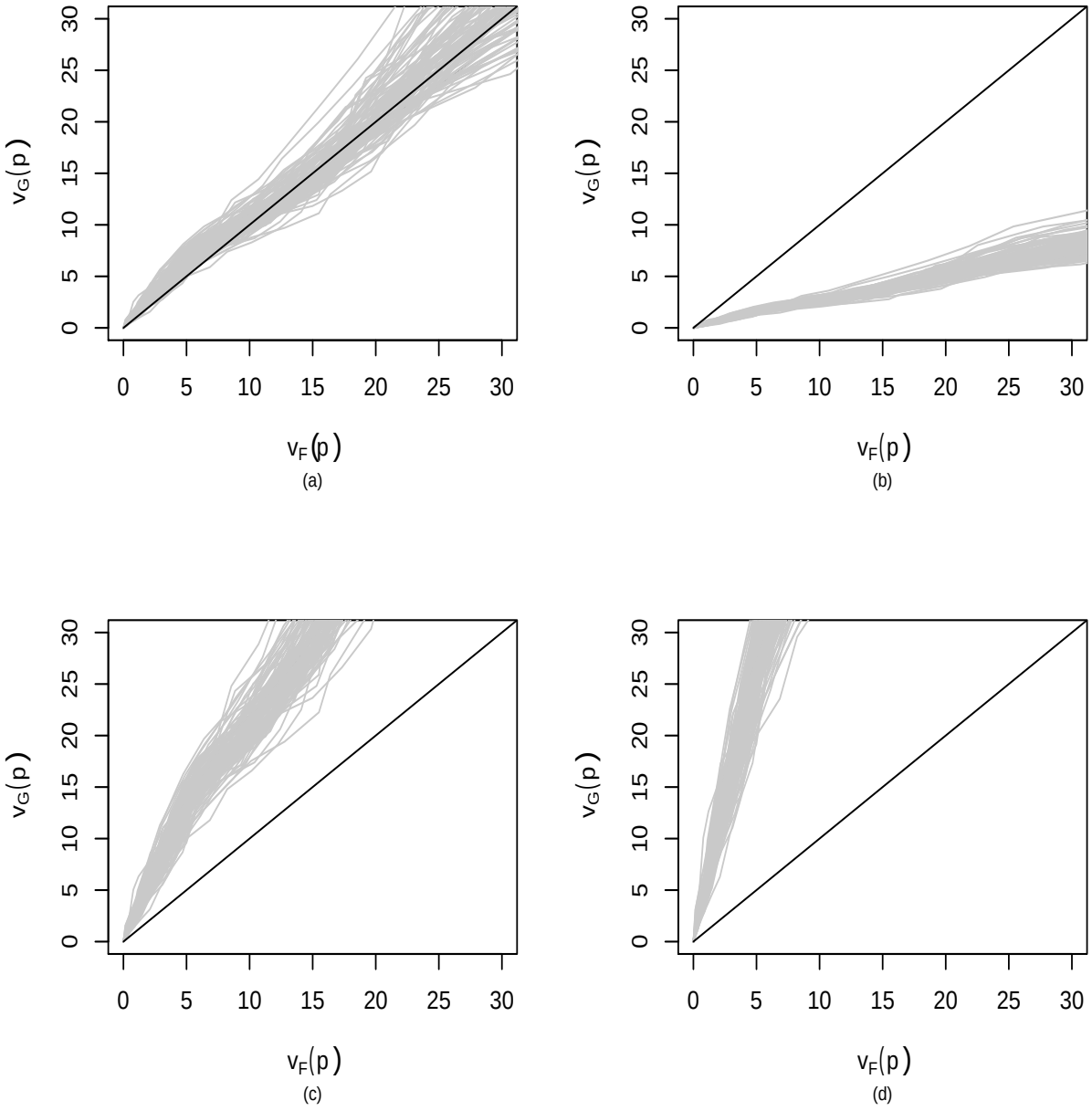

Figure 7.3: Scale-Scale plot for test of scale for bivariate Laplace distribution for $n = 500$. (a) $\Sigma_{\mathbf{Y}} = \mathbf{I}_2$. (b) $\Sigma_{\mathbf{Y}} = 0.25\mathbf{I}_2$. (c) $\Sigma_{\mathbf{Y}} = 2\mathbf{I}_2$. (d) $\Sigma_{\mathbf{Y}} = 4\mathbf{I}_2$.

CHAPTER 8

CONCLUDING REMARKS AND FUTURE WORK

Graphical representation is sometimes very useful for understanding descriptive properties of the data. One example of visual representation of multivariate data in two dimension is the star plot (see Fienberg (1979)) or the radar plot. In this method the each observation is represented as a star-shaped figure spacing out all variables at equal angles around a circle with one ray for each variable. For a given observation, the length of each ray is proportional to the size of that variable. For this type of plot it is difficult to visually compare lengths of different rays, because radial distances are not easy to compare. Visual tests are helpful for describing the underlying properties of a distribution. The most popular plots for visual test purposes, for example the Q-Q plots and the Bag plots, are restricted to univariate and bivariate case. Our proposed plot, on the other hand, is a more general one. The visual test based on scale-scale plots can be extended to more than two dimensions.

The scale-scale plot proposed here is based on a volume functional of central rank regions. For spherically and elliptically symmetric distributions, there are nice and closed form expressions for the theoretical volume functional. The sample version of the volume functional V_{F_n} does not have a proper closed form expression. So to compute it we are first computing the rank vector of the data points. Then we compute the lengths of the rank vectors, order them and take the p-th quantile value as $r_{F_n}(p)$ as the radius of the rank region $C_{F_n}(p)$. Then we generated vector $\mathbf{u}$s such that $||\mathbf{u}|| = r_{F_n}(p)$. After that,

we compute $Q(\mathbf{u})$s corresponding to $\mathbf{u}$s and using the qhull program (Barber, Dobkin and Huhdanpaa (1996)) we compute the volume V_{F_n} of the convex hull formed by $Q(\mathbf{u})$s. The qhull program uses a convex hull algorithm for the volume computation and this algorithm uses virtual memory of a computer rather than actual memory. This makes the program simpler compared with the other available convex hull algorithms and can be used for general dimension convex hulls. For more detail one can see Barber, Dobkin and Huhdanpaa (1996). Computing $V_{F_n}(p)$ is a major issue here; as the dimension d and the sample size n increases, it takes longer to compute the volume of the rank region $C_{F_n}(p)$.

From the asymptotic properties of the volume functional, we have observed that $\sqrt{n}(V_{F_n}(p) - V_F(p))$ scaled by $v_F(p) = \frac{d}{dp}V_F(p)$ obtains a limiting Brownian bridge. If we want to construct a test for $V_F(p)$ or construct a confidence interval for $V_F(p)$ based on $V_{F_n}(p)$, then we need an estimate for $v_F(p)$. From Chapter 3 we can see that $v_F(p)$ is a function of the density function of $(\mathbf{X}-\boldsymbol{\theta})^T\Sigma^{-1}(\mathbf{X}-\boldsymbol{\theta})$ for elliptically symmetric distributions. Thus we may need to estimate $v_F(p)$ by some density estimation method. Using an efficient estimate of $v_F(p)$ would lead to obtaining a better estimation of $V_F(p)$. However, in the most general case, when the distribution may not be elliptically symmetric, the exact form of $v_F(p)$ might not be obtainable easily and a numerical method to estimate it directly may be required, affecting the efficiency of our proposed testing procedure.

Visual tests may not be as powerful as some more theoretical tests but they often provide more insight to the situations. We propose in Chapter 4 that the scale-scale plots can be used to visualise the underlying properties of a multivariate distribution in a similar way. We give a few examples of such uses; in Example 1 by comparing a bivariate random sample with a standard bivariate normal distribution and in Examples 2, 3, 4 by comparing two data sets with each other. These examples show that the scale-scale plot can be considered as a multivariate analogue of the quantile-quantile plots which are a popular visual tool for comparing two distributions.

In Chapter 5 we have defined a slope functional $s_n(p)$ as the ratio of sample volume

functional $V_{F_n}(p)$ to theoretical volume functional $V_F(p)$. In that chapter we have constructed a test statistic T_n which is a function of the slope functional $s_n(p)$. We have also discussed some of the asymptotic properties of T_n. A test of fit of multivariate normality and multivariate Laplace is suggested. We have used our proposed test statistic T_n and have computed relevant cut off values for the test statistic at 5% level of significance. The power of the test for bivariate and trivariate variants of the normal, t and Laplace distributions has been computed. We have also compared the powers obtained from our proposed test with the powers obtained from Kolmogorov-Smirnov test and Cramer von-Mises test as described in Malkovich and Afifi (1973).

In Chapter 6, we have suggested a test for elliptical symmetry based on the test statistic T_n. We performed a power study for the bivariate gamma distribution and the bivariate g and h distribution. We have also provided asymptotic results for performing one sample and two sample tests. We also note here that in Chapter 6 to compute the value of the $\mathbf{u_i}$s from the data we estimated the location vector and the scale matrix respectively by the sample mean and the sample covariance matrix, which are quick to compute but usually not very robust. This was done to save on computational cost which, as mentioned earlier in this chapter, was an issue for most of our data analysis work which required significant computing time even with relatively advanced computing resources. In future, with improved computing resources, we hope to replace our present estimating procedures by something more robust, eg. estimates obtained by the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999). From the examples given in the Chapter 6 we can see that this test of elliptical symmetry can be useful for comparing multivariate distributions. It can also be used as a test of location and scale and for calculating the power of the tests of location. We have presented power tables for the test of location. The power study is done on samples of size 30. For this small sample size we obtained power values which were quite encouraging, but we also observed that the tests were slightly anti-conservative in some cases with the estimated size of a test at a 5% level of significance sometimes being close to 7%. The computer programs which

we used for our computation are quite time consuming: computing the power of the test for a sample of size 30 based on 1000 iterations taking a little over a day in a machine with a dual-core 2.27GHz CPU with 4GB RAM. Hence, although performing a test with a larger sample size with more simulations could have helped us to ascertain whether the observed over-estimation of sizes are a systematic problem of these tests or just a small sample error, at present we are unable to do so for the computing costs involved. However, even with this problem of sizes being slightly higher, the obtained powers are shown to increase rapidly with the location parameter for the small sample size of 30, which is very encouraging.

It may be worth mentioning here that most of our simulation work was restricted by the heavy computation cost and hence were somewhat limited in scope.

We have developed a test of location based on scale-scale plot and presented a detailed power study. We have also developed a visual test for scale as well. Although similar ideas were briefly introduced in Singh, Tyler, Zhang and Mukherjee (2009) for quantile scale curves, no detailed discussion was provided. We provide a power study using simulations for the visual test for location based on scale-scale plot.

8.1 Further Work and Possible Extensions

In Chapter 5 we took our test statistic $T_n = \log U_n - V_n$ where U_n and V_n are defined in (5.6) and (5.7). One may think of other distance functionals based on L_1 distance, $\int_0^1 |V_{F_n}(p) - V_F(p)|dp$ or L_2 distance, $\int_0^1 (V_{F_n}(p) - V_F(p))^2 dp$ (which is a Cramer von-Mises type test statistic) or the L_∞ distance, $\sup_{0<p<1} |V_{F_n}(p) - V_F(p)|$ (which is a Kolmogorov-Smirnov type test statistic). The asymptotics in these cases are much more difficult to work out. Also in order to compute integrals of this type accurately, one may need to estimate the volume function at more points which can be computationally expensive. We hope to take a look at those in the future.

Similarly for the two sample case also we hope to look at the Cramer von-Mises type test statistic, $\int_0^1 (V_{F_n}(p) - V_{G_n}(p))^2 dp$ and the Kolmogorov-Smirnov type test statistic, $\sup_{0<p<1} |V_{F_n}(p) - V_{G_n}(p)|$ in the future.

While comparing the performance of our proposed test statistic T_n with existing methodologies we choose a version of Kolmogorov-Smirnov and Cramer von-Mises test statistic proposed by Malkovich and Afifi (1973). Many more sophisticated recent tests for multivariate normality are available e.g. Ghosh and Ruymgaart (1992) based on the studentised empirical function, Szekely and Rizzo (2005) based on a class of V-statistics, Kankainen, Taskinen and Oja (2007) based on location vectors and scatter matrices to detect skewness and kurtosis. But all of these tests are computationally intensive. There also exist many visual tests, see Chapter 5 for some examples. However most of them are often difficult to compute and computationally costly. With the aid of better computers we plan to compare our test statistic with some of these tests in future.

In Chapter 6 we introduced a test of elliptical symmetry using a scale based test statistic and performed a power study against some skew elliptic and g-h distribution alternatives using simulation. We plan to study the behaviour of the scale curves for skew-elliptic and g-h distributions rigorously and obtain theoretical expressions for the power of the test we obtained using simulations.

Most of the existing tests for elliptical symmetry as described in Chapter 6 have complicated expressions and either complex or non-existent asymptotic distributions, see Huffer and Park (2007). Hence comparing our test with the existing tests is a computationally challenging job which we plan to do in the future with better computing resources.

LIST OF REFERENCES

[1] Aloupis, G., Langerman, S., Soss M., Toussaint G. (2003), Algorithms for bivariate medians and a Fermat-Torricelli problem for lines, *Computational Geometry: Theory and Applications*, **26**, pp 69–79.

[2] Aloupis, G., Cortes, C., Gomez, F., Soss M., Toussaint G. (2002), Lower bounds for computing statistical depth, *Computational Statistics and Data Analysis*, **40**, pp 223-229.

[3] Andrews, D.F., Gnanadesikan, R., and Warner, J.L. (1973), Methods for assessing multivariate normality. In *Multivariate Analysis - III*, (ed. P.R. Krishnaiah), Academic Press, New York, pp. 95–116.

[4] Arcones, M. A., Chen, Z., Gine, E., (1994) Estimators Related to U-Process with Applications to Multovariate Medians: Asymptotic Normality, *The Annals of Statistics*, **22**, pp. 1460–1477.

[5] Avérous, J., and Meste, M. (1994), Multivariate Kurtosis in L_1-sense, *Statistics and Probability Letters*, **19**, pp. 281–284.

[6] Avérous, J., and Meste, M. (1997), Median balls: an extension of the interquantile intervals to multivariate distributions, *Journal of Multivariate Analysis*, **63**, pp. 222–241.

[7] Azzalini, A., Dalla Valle, A. (1996), The Multivariate Skew-Normal Distribution, *Biometrika*, **83**, pp. 715–726.

[8] Azzalini, A., Capitanio, A. (1998) Statistical Applications of the Multivariate Skew-Normal Distribution, *Journal of Royal Statistical Society, Series B*, **61**, pp. 579–602.

[9] Babu, G. J., Rao, C. R. (1988), Joint Asymptotic Distribution of Marginal Quantiles and Quantile Functions in Samples from a Multivariate Population, *Journal of Multivariate Analysis*, **27**, pp. 15–23.

[10] Bahadur, R. (1966), A Note on Quantiles in Large Samples, *The Annals of Mathematical Statistics*, **37**, pp. 577–580

[11] Bai, Z., He, X. (1999), Asymptotic Distributions of the Maximal Depth Estimators for Regression and Multivariate Location, *The Annals of Statistics*, **27**, pp. 1616–1637.

[12] Balanda, K.P., and MacGillivray, H.L. (1988), Kurtosis : A Critical Review, *The American Statistician*, **42**, pp. 111–119.

[13] Balanda, K.P., and MacGillivray, H.L. (1990), Kurtosis and spread, *The Canadian Journal of Statistics*, **18**, pp. 17–30.

[14] Barber, C. B., Dobkin, D. P., Huhdanpaa, H. T. (1996), The Quickhull Algorithm For Convex Hulls, *ACM Transactions on Mathematical Software*, **22**, pp. 469–483.

[15] Bartlett, M. S. (1934), The Vector Representation of a Sample, *Proceedings of the Cambridge Philosophical Society*, **30**, pp. 327–340.

[16] Barnett, V. (1976), The Ordering of Multivariate Data, *Journal of Royal Statistical Society, Series A*, **139**, pp. 319–354.

[17] Bennett, B. M. (1962), On Multivariate Sign Tests, *Journal of the Royal Statistical Society, Series B*, 24, pp. 159–161.

[18] Bennett, B. M., Lindley, D. V. (1961), On a certain multivariate non-normal distribution, *Mathematical Proceedings of the Cambridge Philosophical Society*, **57**, pp. 434–436

[19] Bickel, P.J. (1964), On some alternative estimates for shift in the p-variate one sample problem. *The Annals of Mathematical Statistics*, **35**, pp. 1079–1090

[20] Bickel, P.J. (1965), On some asymptotically nonparametric competitors of Hotelling's T^2. *The Annals of Mathematical Statistics*, **36**, pp. 160–173.

[21] Bickel, P.J., Lehmann, E.L. (1975a), Descriptive statistics for nonparametric models. I. Introduction. *The Annals of Statistics*, **3**, pp. 1038–1044.

[22] Bickel, P.J., Lehmann, E.L. (1975b), Descriptive statistics for nonparametric models. II. Location. *The Annals of Statistics*, **3**, pp. 1045–069.

[23] Bickel, P.J., Lehmann, E.L. (1976), Descriptive statistics for nonparametric models. III. Dispersion. *The Annals of Statistics*, **4**, pp. 1139–1158.

[24] Bickel, P.J., and Lehmann, E.L. (1979), Descriptive statistics for nonparametric models. IV. Spread. In *Contributions to Statistics, Jaroslav Hájek Memorial Volume* (ed. J. Jureckova), Academia, Prague, pp. 33–40.

[25] Beran, R. (1979), Testing for ellipsoidal symmetry of a multivariate density, *The Annals of Statistics* **7**, pp. 150-162.

[26] Beran R. J., Millar P. W., (1997), Multivariate Symmetry Models, *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, D. Pollard, E. Torgensen and G. L. Yang eds. Springer, New York, pp. 13–42.

[27] Blough, D. K., (1989), Multivariate Symmetry Via Projection Pursuit, *Annals of the Institute of Statistical Mathematics*, **41**, pp. 461 –475.

[28] Blumen, I. (1958), A New Bivariate Sign Test for Location, *Journal of the American Statistical Association*, **53**, pp. 448–456.

[29] Brown, B. M. (1983), Statistical Uses of the Spatial Median. *Journal of Royal Statistical Society, Series B*, **45**, pp. 25–30.

[30] Brown, B. M., Hettmansperger, T. P. (1987), Affine Invariant Rank Methods in the Bivariate Location Model, *Journal of Royal Statistical Society, Series B*, **49**, pp. 301–310.

[31] Brown, B. M., Hettmansperger, T. P. (1989), The Affine Invariant Bivariate Version of the Sign Test, *Journal of the Royal Statistical Society, Series B,* **51**, pp. 117–125.

[32] Brown, B. M., Hettmansperger, T. P., Nyblom, J., Oja, H. (1992), On Certain Bivariate Sign Tests and Medians, *Journal of the American Sta- tistical Association*, **87**, pp. 127–135.

[33] Cambanis, S., Keener, R., Simons, G., (1983), On α-symmetric Multivariate Distributions, *Journal of Multivariate Analysis*, **13**, pp. 213–233.

[34] Chakraborty, B. (2001), On Affine Equivariant Multivariate Quantiles. *Annals of the Institute of Statistical Mathematics*, **53**, pp. 380–403.

[35] Chakraborty, B. (2003), On multivariate quantile regression. *Journal of Statistical Planning and Inference*, **110**, pp. 109–132.

[36] Chakraborty, B., Chaudhuri, P. (1996), On a Transformation and Re-Transformation Technique for Constructing an Affine Equivariant Multivariate Median. *Proceedings of the American Mathematical Society*, **124**, pp. 2539–2547.

[37] Chaudhuri, P., Sengupta, D. (1993), Sign Tests in Multidimension: Inference Based on the Geometry of the Data Cloud, *Journal of the American Statistical Association*, **88**, pp. 1363–1370.

[38] Chaudhuri, P. (1992), Multivariate Location Estimation Based on Using Extension of R-Estimates Through U-Statistics Type Approach, *The Annals of Statistics*, **20**, pp. 897–916.

[39] Chaudhuri, P. (1996), On a Geometric Notion of Quantiles for Multivariate Data, *Journal of the American Statistical Association*, **91**, pp. 862–872.

[40] Chen, Z. (1995), Robustness of the half-space median, *Journal of Statistical Planning and Inference* **46**, pp. 175–181.

[41] Cheng A, Ouyang, M. (2001) On algorithms for simplicial depth, *Proceedings of the 13th Canadian Conference on Computational Geometry*, pp. 53–56.

[42] Chmielewski, M. A. (1980), Invariant scale matrix hypothesis tests under elliptical symmetry *Journal of Multivariate Analysis*, **10**, pp. 343–350

[43] Chmielewski, M. A. (1981), Elliptically Symmetric Distributions: A Review and Bibliography, *International Statistical Review*, **49**, pp. 67–74.

[44] Chatterjee, S. K. (1966), A Bivariate Sign Test for Location, *The Annals of Mathematical Statistics*, **37**, pp. 1771-1782.

[45] Cleveland, W. (1993), *Visualizing Data*, Hobart Press.

[46] Cook, D.R., Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman & Hall, New York.

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

[48] Cramer, H. (1946), Mathematical Methods of Statistics. Princeton University Press, Princeton.

[49] Dempster, A. P. (1969), Elements of Continuous Multivariate Analysis. Addison-Wesley, New York.

[50] Dhar, S. S., Chakraborty, B., Chaudhuri, P. (2011) Comparison of Multivariate Distributions Using Quantile-Quantile Plots and Related Tests. *To be submitted.*

[51] Donoho, D. L. (1982), Breakdown properties of multivariate location estimators. Ph.D. Qualifying Paper, Department of Statistics, Harvard University.

[52] Donoho, D. L., Huber, P. J., (1983) The Notion of Breakdown Point, In *A Festschrift for Erich L. Lehmann in Honor of His Sixty-fifth Birthday* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) pp. 157–184. Wadsworth, Belmont, California.

[53] Donoho, D. L., Gasko, M. (1987), Multivariate Generalization of the Median and Trimmed Sum. I., Technical Report 133, Department of Statistics, University of California, Berkeley.

[54] Donoho, D. L., Gasko, M. (1992), Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness, *The Annals of Statistics*, **20**, pp. 1803–1827.

[55] Dümbgen, L. (1992), Limit Theorems for the Simplicial Depth, *Statistics and Probability Letters*, **14**, pp. 119–128.

[56] Doornik, J. A., Hansen, H., (2008), An Omnibus Test for Univariate and Multivariate Normality, *Oxford Bulletin of Economics and Statistics*, **70**, pp. 927-939.

[57] Dyson, F. J. (1943), A Note on Kurtosis, *Journal of the Royal Statistical Society, Series B*, **106**, pp. 360–361.

[58] Easton, G.S., and McCulloch, R.E. (1990), A multivariate generalization of quantile-quantile plots, *Journal of the American Statistical Association*, **85**, pp. 376–386.

[59] Einmahl, J. H. J., Mason, D. M. (1992), Generalized Quantile Processes, *The Annals of Statistics*, **20**, pp. 1062–1078.

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

[61] Fang, K.T., Li, R., and Liang, J. (1998), A multivariate version of Ghosh's T_3-plot to detect non-multinormality, *Computational Statistics and Data Analysis*, **28**, pp. 371–386.

[62] Field, C., Genton, M., G., (2006), The Multivariate g-and-h Distribution, *Technometrics*, **48**, pp. 104–111.

[63] Fienberg, S., (1979), Graphical Methods in Statistics, *The American Statistician*, **33**, pp. 165–178.

[64] Finucen, H. M., (1964), A Note on Kurtosis, *Journal of the Royal Statistical Society. Series B (Methodological)*, **26**, pp. 111–112.

[65] Fraiman, R., Meloche, J. (1996), Multivariate L-estimation. *Preprint.*

[66] Ghosh, S. (1996), A New Graphical Tool to Detect Non-Normality, *Journal of the Royal Statistical Society. Series B (Methodological)*, **59**, pp. 691-702.

[67] Ghosh, S., Ruymgaart, F. H. (1992), Applications of empirical characteristic functions in multivariate problems, *Canadian Journal of Statistics* **20**, pp. 429-440.

[68] Gini, C., Galvani, L. (1929), Di Talune Extensioni Dei Concetti di Media ai Caratteri Quatitativi, *Metron*, **8**. partial translation in *Journal of American Statistical Association*, **25**, pp. 448–450.

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

[70] Groeneveld, R. A. and Meeden, G. (1984), Measuring Skewness and Kurtosis, *The Statistician*, **33**, pp. 391–399.

[71] Gower, J. C., Groeneny, P. J. F., van de Velden, M. (2010), Area Biplots, *The Journal of Computational and Graphical Statistics*, **19**, pp. 46–61.

[72] Hald, A. (1952), Statistical Theory With Engineering Applications, New York: John Wiley

[73] Haldane, J. B. S. (1948), A Note on the Median of a Multivariate Distribution, *Biometrika*, **35**, pp. 414–415.

[74] Hartman, P., Wintner, A., (1940) On the Spherical Approach to the Normal Distribution Law, *American Journal of Mathematics*, **62** pp. 759–779.

[75] Hayford, J. L. (1902), What is the center of an area, or the center of a population?, *Journal of American Statistical Association*, **8**, pp. 47–58.

[76] He, X., Wang, G. (1997), Convergence of Depth Contours for Multivariate Datasets, *The Annals of Statistics*, **18**, pp. 405–414.

[77] Healy, M.J.R. (1968), Multivariate normal plotting. *Applied Statistics*, **17**, pp. 157–161.

[78] Hettmansperger, T. P., (1984), *Statistical Inference Based on Ranks*, Wiley, New York.

[79] Hoaglin, D.C., (1983), Summarizing Shape Numerically: The g-and-h Distributions, In *Exploring Data, Tables, Trend and Shapes*, Wiley, New York, pp. 461–513.

[80] Hodges, J. L. (1955), A Bivariate Sign Test, *The Annals of Mathematical Statistics*, **26**, pp. 523–527.

[81] Hodges, J. L., Lehmann, E. L. (1963), Estimates of Location Based on Rank Tests, *The Annals of Mathematical Statistics*, **34**, pp. 598–611.

[82] Huffer, F. A., Park, C. (2007), A test for elliptical symmetry, *Journal of Multivariate Analysis*, **98**, pp. 256–281.

[83] Hyndman, R. J., Shang, H. L. (2009), Rainbow Plots, Bagplots and Boxplots for Functional Data, *The Journal of COmputational and Graphical Statistics*, **19**, pp. 29–45.

[84] Isogai, T. (1985), Some extension of Haldanes multivariate median and its application, *Annals of the Institute of Statistical Mathematics* **37**, pp. 289-301.

[85] Johnson, D.S., Preparata, F.P. (1978), The densest hemisphere problem, *Theoretical Computer Science* **6** (1978), pp. 93-107.

[86] Johnson, R.A., Wichern, D.W. (2002), Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey.

[87] Justel, A., Pefia, D, Zamar, R. (1997), A Multivariate Kolmogorov-Smirnov Test of Goodness of Fit, *Statistics and Probability Letters*, **35**, pp.251–259.

[88] Kankainen, A., Taskinen, S., Oja, H. (2007), Tests of multinormality based on location vectors and scatter matrices, *Statistical Methods and Applications*, **16**, pp. 357–379.

[89] Koltchinskii, V. I. (1997), M-Estimation, Convexity and Quantiles, *The Annals of Statistics*, **25**, pp. 435–477.

[90] Koltchinskii, V. I., Sakhanenko, L. (2000), Testing for ellipsoidal symmetry of a multivariate distribution. In *High Dimensional Probability II, Progress in probability*, Birkhauser, Boston, pp. 493–510.

[91] Koshevoy, G., Mosler, K. (1997), Zonoid Trimming for Multivariate Distributions, *Annals of Statistics*, **25**, pp. 1998–2017.

[92] Kowalski, C.J. (1970), The Performance of Some Rough Tests for Bivariate Normality Before and After Coordinate Transformations to Normality, *Technometrics*, **12**, pp. 517–544.

[93] Koziol, J. A. (1983), On Assessing Multivariate Normality, *Journal of the Royal Statistical Society, Series B*, **45**. pp. 358–361.

[94] Koziol, J. A. (1993), Probability Plots for Assessing Multivariate Normality, *The Statistician*, **42**, pp. 161-173.

[95] Li, R., Fang, K., Zhu, L. X. (1997), Some Q-Q Probability Plots to Test Spherical and Elliptical Symmetry *Journal of Computational and Graphical Statistics*, **6**, pp. 435–450

[96] Liang, J., and Ng, K.W. (2009), A multivariate normal plot to detect nonnormality, *Journal of Computational and Graphical Statistics*, **18**, pp. 52–72.

[97] Liang, J., Pan, W., and Yang, Z.H. (2004), Characterization-based Q-Q plots for testing multinormality, *Statistics & Probability Letters*, **70**, pp. 183–190.

[98] Liu, R. Y. (1988), On a Notion of Simplicial Depth, *Proceedngs of the National Academy of Sciences*, **85**, pp. 1732–1734.

[99] Liu, R. Y. (1990), On a Notion of Data Depth Based on Random Simplices, *The Annals of Statistics*, **18**, pp. 405–414.

[100] Liu, R. Y. (1992), Data Depth and Multivariate Rank Tests, *Proceedings of 2nd International Conference on Statistical Data Analysis Based on the L-1 Norm and Related Methods*, Y. Dodge, Ed., pp. 279–294.

[101] Liu, R. Y. (1998) On a Notion of Simplical Depth, *Proceeding of the National Academy of Science, USA,* **85**, pp. 1732–1734.

[102] Liu, R. Y. (2003), Data depth: center-outward ordering of multivariate data and nonparametric multivariate statistics, *Recent Advances and Trends in Nonparametric Statistics*, eds. M. Akritas and D. Politis, Elsevier Science, pp. 155-168.

[103] Liu, R. Y., Parelius, J. M., Singh, K. (1999), Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference, *The Annals of Statistics*, **27**, pp. 783–858.

[104] Lord, R. D. (1954), The Use of the Hankel Transform in Statistics. I. General Theory and Examples, *Biometrika*, **41**, pp. 44–55.

[105] Mahalanobis, P. C. (1936), On the Generalized Distance in statistics, *Proceedings of the National Academy of India*, **12**, pp. 49–55.

[106] Malkovich J. F., Afifi, A. A. (1973), On Tests for Multivariate Normality, *Journal of the American Statistical Association*, **68**, pp. 176–179.

[107] Manzottia, A., Prezb, F. J., Quiroz, A. J., (2002) A Statistic for Testing the Null Hypothesis of Elliptical Symmetry, *The Journal of Multivariate Analysis*, **81**, pp. 274–285.

[108] Marden, J.I. (1998), Bivariate QQ-plots and spider web plots, *Statistica Sinica*, **8**, pp. 813–826.

[109] Marden, J.I. (2004), Positions and QQ Plots, *Statistical Science*, **19**, pp. 606–614.

[110] Mardia, K., V. (1970), Measures of Multivariate Skewness and Kurtosis with Applications, *Biometrika*, **57**, pp. 519–530.

[111] Mardia, K., V. (1974), Applications of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Study, *Sankhya Series B*, **36**, pp. 114–118.

[112] Mardia, K., V., Kent, J., T., Bibby, J., M. (1979), Multivariate Analysis (Probability and Mathematical Statistics) *Academic Press*.

[113] Maxwell, J. C., (1860), Illustration of the Dynamical Theory of Gases - Part I. On the Motions and Collisions of Perfectly Elastic Bodies, *Taylor's Philosophical Magazine*, **19** pp. 19–32.

[114] Muller, C. (1966), Lecture Notes in Mathematics, **17**, Springer-Verlag, New York.

[115] Niinimaa, A., Oja, H., Tableman, M. (1990), The Finite-Sample Breakdown Point of the Oja Bivariate Median and of the Corresponding Half-Sample Version, *Statistics and Probability Letters*, **10**, pp. 325–328.

[116] Oja, H. (1983), Descriptive Statistics for Multivariate Distributions, *Statistics and Probability Letters*, **1**, pp. 327–332.

[117] Oja, H., Nyblom, J (1989), On Bivariate Sign Tests, *Journal of the American Statistical Association*, **84**, pp. 249-259.

[118] Oja, H. (1999), Affine Invariant Multivariate Sign and Rank Tests and Corresponding Estimates: A Review, *Scandinavian Journal of Statistics*, **26**, pp. 319–343.

[119] Oja, H., Niinimaa, A. (1985), Asymptotic Properties of the Generalized Median in the Case of Multivariate Normality, *Journal of the Royal Statistical Society. Series B (Methodological)*, **47**, pp. 372–377.

[120] Oja, H., Nyblom, J. (1989), Bivariate Sign Tests, *Journal of the American Statistical Association*, **84**, pp. 249–259.

[121] Olkin, I., Tong, Y. L. (1988), Peakedness in multivariate distributions. In *Statistical decision theory and related topics, IV, Vol. 2.* Springer, New York, pp. 373–383.

[122] Pearson, K. R. (1905), Skew Variation, a Rejoinder, *Biometrika*, **4**, pp. 169–212

[123] Puri, M.L. and Sen, P.K. (1971), Nonparametric Methods in Multivariate Analysis. Wiley, New York.

[124] Rao, C. R. (1988), Methodology Based on the L_1 Norm in Statistical Inference, *Sankhya, Series A*, **50**, pp. 289–313.

[125] Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *Journal of the American Statistical Association*, **84**, pp. 1045–1050.

[126] Rietz, H. L. (1943). Mathematical Statistics. *Carus Math. Monographs, No. 3.*

[127] Romanazzi, M., (2004) Data Depth and Correlation, *Allgemeines Statisticshes Archive* **88**, pp. 191–214.

[128] Ronkainen, T., Oja, H., Orponen, P. (2001) Computation of the multivariate Oja median. In *Developments in Robust Statistics: Proceedings of the International Conference on Robust Statistics* pp. 344-359, Springer-Verlag, Berlin.

[129] Rousseeuw, P. J., Van Driessen K. (1999), A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, **41**, pp. 212–223.

[130] Rousseeuw, P. J., Ruts, I. (1996), Bivariate location depth, *Applied Statistics* **45**, pp. 516–526.

[131] Rousseeuw, P.J., Ruts, I. (1998), Constructing the bivariate Tukey median, *Statistica Sinica* **8**, pp. 827–839.

[132] Rousseeuw, P.J., Struyf, A. (1998), Computing location depth and regression depth in higher dimensions. *Statistics and Computing* **8**, pp. 193–203.

[133] Royston, J. P. (1983), Some Techniques for Assessing Multivarate Normality Based on the Shapiro-Wilk W, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **32**, pp. 121–133.

[134] Royston J.P. (1993), Graphical Detection of Non-Normality by Using Michaels Statistic, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **42**, pp. 153–158.

[135] Ruts, I., Rousseeuw, P. J. (1996), Computing Depth Contours of Bivariate Point Clouds, *Computational Statistics and Data Analysis*, **23**, pp. 153–168.

[136] Schott, J. R. (2002), Testing for Elliptical Symmetry in Covariance Matrix Based Analyses, *Statistics and Probability Letters*, **60**, pp. 395–404.

[137] Serfling, R. (1980), Approximation Theorems of Mathematical Statistics, Wiley, New York.

[138] Serfling, R. (2002a), Quantile Functions for Multivariate Analysis: Approaches and Applications, *Statistica Neerlandica*, **56**, pp. 214–232.

[139] Serfling, R. (2002b), Generalised Quantile Processes Based on Multivariate Depth Functions, with Applications in Nonparametric Multivariate Analysis, *Journal of Multivariate Analysis*, **83**, pp. 232–247.

[140] Serflng, R. (2004), Nonparametric Multivariate Descriptive Measures Based on Spatial Quantiles, *Journal of Statistical Planning and Inference*, **123**, pp. 259–278.

[141] Serfling, R. J., (2006), Multivariate Symmetry and Asymmetry, *Encyclopedia of Statistical Sciences*, Second Edition (S. Kotz, N. Balakrishnan, C. B. Read and B. Vidakovic, eds.), **8**, pp. 5338–5345. Wiley, New York.

[142] Serfling, R. (2006), Depth Functions in Nonparametric Multivariate Inference, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **73**.

[143] Serfling, R. (2010), Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation, *Journal of Nonparametric Statistics*, **22**, pp. 915–936.

[144] Shapiro, S. S., Wilk, M. B. (1965), An analysis of variance test for normality (complete samples). *Biometrika*, **65**, pp. 591–611.

[145] Singh, K. (1991), A Notion of Majority Depth, *Technical Report*, Rutgers University, Department of Statistics.

[146] Singh, K., Tyler, D.E., Zhang, J., and Mukherjee, S. (2009), Quantile scale curves, *Journal of Computational and Graphical Statistics*, **18**, pp. 92–105.

[147] Small, C. G. (1990), A Survey of Multidimensional Medians, *International Statistical Review*, **58**, pp. 263–277.

[148] Struyf, A., Rousseeuw, P. J. (2000), High-dimensional Computation of the Deepest Location, *Computational Statistics and Data Analysis*, **34**, pp. 415–426.

[149] Szekely, G. J., Rizzo, M. L. (2005) A new test for multivariate normality, *Journal of Multivariate Analysis*, **93**, pp. 58–80.

[150] Tukey, J. W. (1975), Mathematics and the picturing of data, *Proceedings of International Congress of Mathematicians, Vancouver*, **2**, pp. 523–531.

[151] Uspensky, J. V. (1937), Introduction to Mathematical Probability, McGraw-Hill, New York, pp. 204–205.

[152] Van Zwet, W.R. (1964), Convex Transformations of Random Variables, *Mathathematics Centre Tracts 7.* Mathematisch Centrum, Amsterdam.

[153] Wang, J., Serfling, R. (2006), On Scale Curves for Nonparametric Description of Dispersion, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **72**, pp. 37–48

[154] Weber, A. (1909), *Uber Den Stanford Der Industrien, Tubingen*, English translation by C. J. Freidrich (1929), *Alfred Weber's Theory of Location of Industries*, Chicago University Press, Chicago.

[155] Wilk, M.B., and Gnanadesikan, R. (1968), Probability plotting methods for the analysis of data, *Biometrika*, **55**, pp. 1–17.

[156] Zhu, L X., Neuhaus, G. (2003), Conditional tests for elliptical symmetry. *Journal of Multivariate Analysis*. **84**, pp. 284–298.

[157] Zuo, Y., Serfling, R. (2000), General Notions of Statistical Depth Function, *The Annals of Statistics*, **28**, pp. 461–482.