The Australian National University

Doctoral Thesis

Nonstandard Estimation for the von Mises Fisher Distribution

Author:
Maryam Ghodsi

Supervisors:
Professor Ross Arthur Maller
Professor Alan Hepburn Welsh

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in

Mathematical Science
Mathematical Sciences Institute
ANU College of Physical and Mathematical Sciences

Declaration of Authorship

I, Maryam Ghodsi, declare that this thesis titled 'Nonstandard Estimation for the von Mises Fisher Distribution' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____

THE AUSTRALIAN NATIONAL UNIVERSITY

Abstract

Mathematical Sciences Institute
ANU College of Physical and Mathematical Sciences

Doctor of Philosophy

Nonstandard Estimation for the von Mises Fisher Distribution

by Maryam Ghodsi

The main focus of this thesis is to provide a theoretical analysis of certain statistical models incorporating constraints of various kinds. The models are motivated with data examples using spherical distributions fitted to asset allocation data related to financial portfolios. We fit models consisting of several von Mises Fisher distributions of different dimensions to a sample of such financial data. Our analysis of the properties of the corresponding hypothesis test statistics under this model combines Silvey's approach towards constraints on the parameter space with Chernoff's innovations to find the asymptotic distributions of the likelihood ratio (deviance) statistics. Silvapulle and Sen detail situations when some constraints are imposed on either the parameter space or on the underlying distribution, where the information matrix is positive definite. But, in some examples, the expectation of the second derivative need not be a positive definite matrix. In such cases, methodology developed by Maller and others can be applied as it does not require the information matrix to be positive definite. Consequently, we apply this methodology to hypothesis tests for the equality of concentration parameters in the spherical subcomponent model. Properties of the tests are examined by simulations and a real data application is given.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Professor Ross Maller who devoted his time and immense knowledge providing me a supportive environment in order to conduct research. His positive attitude towards education pushed me to expand my understanding in science, specially in statistics. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

My sincere thanks goes to Kelly Wicks who provided me a suitable environment to do my research in a best way possible. Without her precious support it would not be possible to conduct this piece of research.

Besides my supervisor, I would like to thank my advisor, Professor Alan Welsh, for his insightful comments and encouragement.

Contents

List of Figures

List of Tables

Chapter 0

Preface

Rotnitzky, Cox, Bottai and Robins (2000) write in the first paragraph of their Introduction that "The asymptotic distributions of the maximum likelihood estimator (MLE) and of the likelihood ratio test statistic of a simple null hypothesis in parametric models have been extensively studied when the information matrix is non-singular. In contrast, the asymptotic properties of these statistics when the information matrix is singular have been studied only in certain specific problems, but no general theory has yet been developed". As a particular example we consider a d dimensional von Mises Fisher distribution on the sphere. In Section 2.4.1 of this thesis, we show that the expected second derivative of the log likelihood function of the von Mises Fisher distribution is singular at the true parameter point. The expectation of the score function vector of this distribution is not zero and hence the variance matrix of the first derivative of the log likelihood function is not equal to the expectation of the square of the score function vector. Maller and his coworkers (2015) have developed a theory which can be applied to the von Mises Fisher distribution. As an application, we analyze financial portfolios consisting of asset allocations from financial data. A combination of von Mises Fisher distributions from different dimensions is assumed to fit the data.

This thesis is organized as follows. Chapter 1 gives an overview of the methods which are applicable in nonstandard situations. Section 1.2.1 of Chapter 1 discusses the work of Wilks (1938) and Chernoff (1954). Wilks considers a special case of hypothesis tests in multivariate cases and proves, under certain assumptions, that the asymptotic distribution of the deviance statistic, under the null, depends on the dimension of the whole parameter space and the dimension of the parameter space under the null.

Chernoff (1954) considers the density of a multivariate normal distribution and shows that the supremum of this density (for its mean lying in an arbitrary set ϕ), when the covariance matrix is assumed to be an identity matrix, is equivalent to the (minimum) distance of a normal random variable from a cone approximating the set ϕ. Then he applies his arguments in three different examples covering the case when the covariance matrix is not an identity matrix. Chernoff in his paper on page 577 after defining the cone (Chernoff uses the word "positively homogeneous set" instead of "cone") states (while he considers the origin as the true value of the parameter) "we may remark that a set bounded by a smooth surface through the origin is approximated by the union of an open half-space with an optional positively homogeneous subset of the tangent hyperplane. It is also easy to see that if ϕ is approximated by a nonnull positively homogeneous set other than the whole space, then the origin is a boundary point of ϕ". Some other researchers including Feder (1968), Moran (1971), Chant (1974), Self and Liang (1987) reconsider Chernoff's methodology and take into account the case when the true parameter may be on the boundary of the parameter space. Silvapulle and Sen (2005) develop the method of finding an approximating cone even when the parameter space is defined by a set of nonlinear equality and inequality constraints. They approximate a set by a cone at the true value of the parameter.

Our interest in this thesis stems from a paper published by Vu and Zhou in 1997 discovering more aspects of Chernoff's theorem and method. They consider a general estimating function with neat and easy to handle notations where having an i.i.d sample is no longer crucial. Their methodology has recently been extended by Maller and his coworkers (2015). We apply their notations in our investigations throughout this thesis.

Silvapulle and Sen (2005) provide a full description of hypothesis testing in different situations where the null and alternative hypotheses might be a linear space, a closed convex cone, or a region defined by a continuously differentiable function. Our results are consistent with their methodology, where relevant, as described in Section 3.3.3.

Silvey (1959) considers another type of hypothesis test using distributions with constraints on the parameter space (such as the multinomial distribution) and introduces a new method of hypothesis testing. In Section 1.2.5 of Chapter 1 we explain how Neuenschwander and Flury (1997) classify different types of constraints and use Silvey's methodology in their examples.

Chapter 2 is devoted to a review of the von Mises Fisher distribution for data on a d dimensional sphere, its moments and properties, the variance matrix and proving that this variance matrix is a positive semidefinite matrix. Except for the variance matrix of a von Mises Fisher distribution theorem, all the other formulae in our exposition can be found in the book by Watson (1983). However, we treat the investigation with our particular application in mind, working with the density function and developing formulae based on its properties and presenting some new proofs of the theorems which can be observed in Watson (1983). For illustration we also graph contour plots of the distribution in 3 dimensions on a sphere. An exposition of the uniform distribution on a sphere is presented to investigate further results.

A suggestion for applying inferential statistics about spherical data is to substitute one parameter by writing it as a function of the other parameters. For example, we may eliminate μ_d using the relation $\mu_d = \sqrt{1 - \mu_1^2 - \ldots - \mu_{d-1}^2}$ in the density function of the spherical distribution. But this leads to unintuitive expressions and it is preferable to develop directly a methodology which can be applied for distributions with a singular expected second derivative matrix.

In the third and fourth chapters of this thesis, we apply our methodology to analyse portfolio or asset allocation data, having observed empirically that a model incorporating different dimensions of a von Mises distribution may be an appropriate model. We apply the aforementioned methodology, studying the theory in Chapter 3 and conducting the data analysis with Mathematica as reported in Chapter 4.

The financial data we used for the portfolio analysis come from various indices from different countries. Chapter 4 contains analyses of the asset allocations in portfolios related to the selected indices. Tables B.1 to B.16 contain the data which are analysed in this thesis. They are the square root of the asset allocations for the portfolios based on indices in Tables 4.1 and 4.2. The daily returns of the desired indices are taken from the Yahoo Finance website for further calculation in Mathematica to evaluate monthly asset allocations.

Chapter 5 describes Wood's method of simulation (1994) of data on the sphere. We compare the contour plots of the distribution in 2 and 3 dimensions and show how well Wood's algorithm works by comparing the simulated data on the circle or sphere with its equivalent contour plots. Wood's algorithm applies to a von Mises Fisher distribution with modal direction $(0, \cdots, 0, 1)^T$; we extend it to enable a known modal vector of the same dimension. To simulate data from a 3 dimensional von Mises Fisher distribution there is an explicit method of simulation due to its direct distribution function formula. Therefore, the results from this distribution are more reliable than for the other dimensions. We check the distribution of the deviance statistic derived from theory by simulations in this chapter and examine the theory applied in Chapter 3. Tables 5.1 to 5.8 and Figures 5.8 to 5.15 show these results.

Chapter 1

Introduction and Background

1.1 Introduction

Let $\mathbb{S}^d \subseteq \mathbb{R}^d$ be the sample space of a d dimensional random variable with the parameter space Θ. Denote the density function (assumed to exist) of this d dimensional random variable by $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ and $\mathbf{x} \in \mathbb{S}^d$ and define the likelihood function of a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of observations on independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from this distribution by

$$L_n(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; \boldsymbol{\theta}) = f(\mathbf{x}_1; \boldsymbol{\theta}) f(\mathbf{x}_2; \boldsymbol{\theta}) \ldots f(\mathbf{x}_n; \boldsymbol{\theta}). \tag{1.1}$$

Fisher (Kendall and Stuart (1979), page 8) identifies $L_n(\mathbf{x}, \boldsymbol{\theta})$ as the likelihood function when it is treated as a function of $\boldsymbol{\theta}$ for fixed $\mathbf{x}$ and the probability distribution of the sample when it is regarded as a function of $\mathbf{x}$ for fixed $\boldsymbol{\theta}$. Clearly, for a true likelihood, we have

$$\int_{\mathbf{x} \in \mathbb{S}^d} L_n(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = 1, \quad \boldsymbol{\theta} \in \Theta. \tag{1.2}$$

We use the notation $\mathcal{L}_n(\mathbf{x}, \boldsymbol{\theta})$ for the (natural) logarithm of the likelihood function and, when it is finite, we write

$$\mathcal{L}_n(\mathbf{x}, \boldsymbol{\theta}) = \log L_n(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\mathbf{x}_i, \boldsymbol{\theta}). \tag{1.3}$$

The likelihood can be used for estimation, hypothesis testing and in general for making inferences about the population under study through its parameters. For this we maximize $\mathcal{L}_n(\mathbf{x}, \boldsymbol{\theta})$ for variations in $\boldsymbol{\theta}$.

In this thesis, we consider a more general maximization problem when the likelihood function is replaced by an estimating function constructed from data which are neither necessarily independent nor from the same distribution. Therefore, we may dispense with the formula (1.1) and consider maximization of an arbitrary function $\mathcal{L}_n(\mathbf{x}, \boldsymbol{\theta})$ subject to some regularity conditions. Throughout we often use the symbol $\mathcal{L}_n(\boldsymbol{\theta})$ instead of $\mathcal{L}_n(\mathbf{x}, \boldsymbol{\theta})$ for the sake of simplicity.
We denote the first derivative vector of the log-likelihood function $\mathcal{L}_n(\boldsymbol{\theta})$ by $\mathbf{S}_n(\boldsymbol{\theta})$ and call it the score function vector and the negative of the second derivative matrix by $\mathbf{F}_n(\boldsymbol{\theta})$, assuming their existence. Thus,

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{1.4}$$

and

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

Let $\boldsymbol{\theta}_0 \in \Theta$ be the true value of the parameter. Under certain regularity conditions (assuming the existence of the first and second derivatives of $\mathcal{L}_n(\boldsymbol{\theta})$ and the interchangeability of the derivatives and integrals) and by following (1.2) and through some calculations, we can show that

$$\mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{S}_n(\boldsymbol{\theta}_0) = 0$$

and

$$\mathrm{Var}_{\boldsymbol{\theta}_0}(\mathbf{S}_n(\boldsymbol{\theta}_0)) = \mathrm{E}_{\boldsymbol{\theta}_0}(\mathbf{S}_n(\boldsymbol{\theta}_0)\mathbf{S}_n^T(\boldsymbol{\theta}_0)) = \mathrm{E}_{\boldsymbol{\theta}_0}(\mathbf{F}_n(\boldsymbol{\theta}_0)). \tag{1.5}$$

Standard proofs of the existence, uniqueness and consistency of a maximum estimator impose these as well as some other regularity conditions. These conditions should guarantee the existence of a maximum of the estimating function, at least in a neighborhood of the true parameter.

However, for some distributions these regularity conditions are not valid, for example for the von Mises Fisher distribution which contributes to a big part of this thesis. Our aim is to study maximum estimators, their distributions and hypothesis tests built upon them under nonstandard conditions when some of the assumptions such as (1.5) do not hold.

The present chapter is organized as follows. In the second section of this chapter, we review some early work done by Wilks (1938) and Chernoff (1954) and later contributions by Feder (1968), Moran (1971), Chant (1974), Self and Liang (1987), Vu and Zhou (1997) and Silvapulle and Sen (2005). Whereas Wilks' method is applicable for hypothesis tests which are in the form of equalities among parameters in a multivariate case (for example when we want to test that all the values of the parameters are exactly known ($\boldsymbol{\theta} = \boldsymbol{\theta}_0$) or in a more general case when we want to test that only some components of $\boldsymbol{\theta}$ are known), by contrast Chernoff's method works well with composite hypothesis tests concerning intervals when the true value of the parameter sits on the boundary of disjoint null and alternative sets.

Aitchison and Silvey (1958) consider a type of distribution where there can be some constraints on the parameter space. They assume the constraints can be expressed in the form of a well-behaved function and maximize the log likelihood function (1.3) using Lagrange multipliers. We discuss some basic properties of this method in Section 1.2.4 of this chapter.

A new approach towards constraints on the parameter space is offered by Neuenschwander and Flury (1997). We discuss their formulation of constraints in Section 1.2.5. Their viewpoints help us to handle hypothesis tests using Lagrangian multiplier tests by considering null hypotheses as constraints imposed on the parameter space.

1.2 Historical Overview

1.2.1 Wilks, Chernoff, Vu and Zhou

In the following sections we consider a genuine likelihood function $L_n(\boldsymbol{\theta})$ to be used for estimation and testing. Consider the hypothesis test

$$\begin{cases} H_0 & : \boldsymbol{\theta} \in \Omega \\ H_A & : \boldsymbol{\theta} \in \tau, \end{cases}$$

for two disjoint subsets Ω and τ of Θ. Consider the generalised likelihood ratio statistic for testing H_0, which is

$$LRT(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Omega} L_n(\mathbf{x}, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Omega \cup \tau} L_n(\mathbf{x}, \boldsymbol{\theta})}$$

and define the deviance statistic to be

$$d_n = -2 \log LRT(\mathbf{x}) = -2 \left(\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_\Omega) - \mathcal{L}_n(\widehat{\boldsymbol{\theta}}_{\Omega \cup \tau}) \right), \tag{1.6}$$

where $\widehat{\boldsymbol{\theta}}_\Omega$ is a value of $\boldsymbol{\theta}$ which maximizes $\mathcal{L}_n(\boldsymbol{\theta})$ on the set Ω and similarly for $\widehat{\boldsymbol{\theta}}_{\Omega \cup \tau}$. Clearly, $LRT(\mathbf{x}) \leq 1$ and $d_n \geq 0$.

An alternative statistic to consider is

$$LRT^*(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta}\in\Omega} L_n(\mathbf{x},\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\mathcal{T}} L_n(\mathbf{x},\boldsymbol{\theta})},$$

for which the corresponding deviance statistic is

$$d_n^* = -2\log LRT^*(\mathbf{x}) = -2\left(\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_\Omega) - \mathcal{L}_n(\widehat{\boldsymbol{\theta}}_\mathcal{T})\right).$$

Chernoff (1954) explained "since $LRT^*(\mathbf{x})$ is more expressive than $LRT(\mathbf{x})$ (that is, $LRT(\mathbf{x}) = LRT^*(\mathbf{x})$ if $LRT^*(\mathbf{x}) \leq 1$ and $LRT(\mathbf{x}) = 1$ if $LRT^*(\mathbf{x}) > 1$) it suffices to study the distribution of $LRT^*(\mathbf{x})$". So, he only used d_n^* in his examples. $LRT^*(\mathbf{x})$ may take any positive value and d_n^* may take any real value. While discussing Chernoff's approach we will consider both versions.

Assume the true value of the parameter (the parameter describing the distribution from which the sample is drawn) to be $\boldsymbol{\theta}_0$ which initially is considered to be an interior point of Θ ($\boldsymbol{\theta}_0$ is interior to Θ if there is a neighborhood of $\boldsymbol{\theta}_0$ contained in Θ). Denote the components of a typical point $\boldsymbol{\theta}$ and of $\boldsymbol{\theta}_0$ by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ and $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0d})$. Wilks (1938) considers the hypotheses

$$\begin{cases} H_0 & : \theta_i = \theta_{0i}; \quad i = m+1, m+2, \cdots, d, \\ H_A & : otherwise, \end{cases}$$

$0 \leq m \leq d-1$ and derives the asymptotic distribution of d_n when H_0 is true in an i.i.d. case. He proves that under certain regularity conditions the asymptotic distribution of d_n is chi square with $d-m$ degrees of freedom.

Chernoff (1954) extends the work of Wilks (1938) by considering subsets of Θ such as hyperplanes, i.e., subspaces of dimension $d-1$ or less. The hyperplanes in $\mathbb{R}$ are points, in $\mathbb{R}^2$ are lines and in $\mathbb{R}^3$ are planes. Every hyperplane divides the space in two parts. Under some regularity conditions Chernoff considers the hypotheses

$$\begin{cases} H_0 & : \boldsymbol{\theta} \text{ is on one side of a hyperplane} \\ H_A & : otherwise, \end{cases}$$

so his emphasis is quite different from Wilks'. He locates the true value of the parameter, $\boldsymbol{\theta}_0$, on the boundary of the two disjoint subsets defined by the null and the alternative hypotheses. In the one-dimensional case the null and alternative hypotheses are simply

$$\begin{cases} H_0 & : \theta \geq \theta_0 \\ H_A & : \theta < \theta_0. \end{cases}$$

Let $\mathcal{N}$ be a neighborhood of $\boldsymbol{\theta}_0$ in Θ and let $\mathbf{X}$ be a random variable in $\mathbb{R}^d$ with density $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathcal{N}$. Chernoff assumes

(CH1) for almost all $\mathbf{x}$, the first, second and third derivatives of $\log(f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}$ exist, for every $\boldsymbol{\theta} \in \mathcal{N}$;

(CH2) if $\boldsymbol{\theta} \in \mathcal{N}$, all of the first, second and third derivatives of $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ are bounded by finitely integrable functions where these functions are the same for the first and second derivatives and the expectation of the third one does not depend on $\boldsymbol{\theta}$.

(CH3) if $\boldsymbol{\theta} \in \mathcal{N}$, the matrix $\mathbf{S}_{\boldsymbol{\theta}}$ in (1.11) is finite and positive definite.

Throughout his proofs, he translates the origin so that $\boldsymbol{\theta}_0$ is zero and considers the hypotheses

$$\begin{cases} H_0 & : \boldsymbol{\theta} \in \Omega \subset \mathcal{N} \\ H_A & : \boldsymbol{\theta} \in \tau \subset \mathcal{N} \end{cases}$$

and illustrates his method of testing them through three examples.

Example 1.1 (CHERNOFF)**.** *Consider a bivariate normal distribution with mean* $\boldsymbol{\theta} = (\theta_1, \theta_2)$ *and a known variance matrix* $\boldsymbol{\Sigma}$ *and suppose we have one single observation* $\mathbf{X} = (X_1, X_2)$*. Consider the general hypotheses*

$$\begin{cases} H_0 & : a_1\theta_1 + a_2\theta_2 < 0 < b_1\theta_1 + b_2\theta_2 \\ H_A & : \textit{otherwise}, \end{cases} \tag{1.7}$$

where a_1*,* a_2*,* b_1 *and* b_2 *are given real numbers.* H_0 *specifies a certain region of the parameter space. Let* Ω *be the region under the null in* (1.7) *and* τ *be the complement of the set* Ω*. The maximum of the likelihood function can be written for this case as* $(n = 1)$*,*

$$\sup_{\boldsymbol{\theta}\in\Omega} L_n(\boldsymbol{\theta}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\{-Q_\Omega(\mathbf{x})/2\},$$

where

$$Q_\Omega(\mathbf{X}) = \inf_{\boldsymbol{\theta}\in\Omega} (\mathbf{X} - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\theta}).$$

Then

$$d_n^* = Q_\Omega(\mathbf{X}) - Q_\tau(\mathbf{X}).$$

First assume $\boldsymbol{\Sigma}$ to be $\mathbf{I}$, the identity matrix. Then $Q_\Omega(\mathbf{X})$ is the squared Euclidean distance from $\mathbf{X}$ to Ω. Ω is a cone with the vertex at $(0, 0)$ and angle ψ which is the

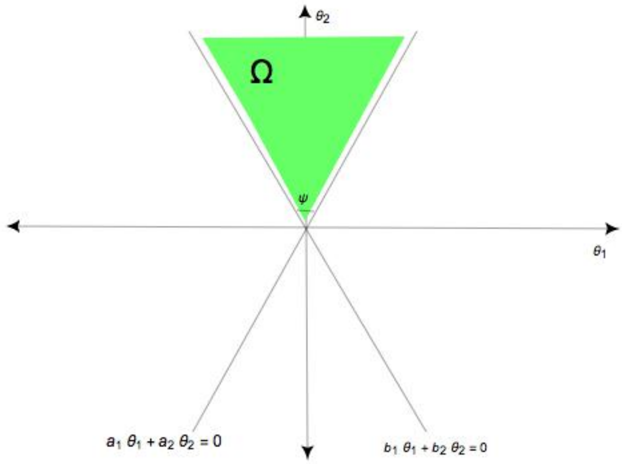

FIGURE 1.1: Cone described in Example 1 of Chernoff

angle between the two vectors $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ (Figure 1.1). Therefore,

$$Q_\Omega(\mathbf{X}) = \begin{cases} 0 & ; \mathbf{X} \in \Omega \\ \text{the squared distance from } \mathbf{X} \text{ to the boundary of } \Omega & ; \mathbf{X} \in \tau \end{cases}$$

and

$$Q_\tau(\mathbf{X}) = \begin{cases} \text{the squared distance from } \mathbf{X} \text{ to the boundary of } \Omega & ; \mathbf{X} \in \Omega \\ 0 & ; \mathbf{X} \in \tau. \end{cases}$$

Figure 1.1 shows this cone for given values of $\mathbf{a}$ and $\mathbf{b}$; see also Figure 1.2 where various regions are defined. We think of a random vector $\mathbf{X}$ being placed in the plane according to a bivariate standard normal distribution. Consider the regions in Figure 1.2 and by using some trigonometry we can compute

$$Q_\Omega(\mathbf{X}) = \begin{cases} 0 & ; \mathbf{X} \in \Omega \\ \dfrac{(a_1X_1 + a_2X_2)^2}{a_1^2 + a_2^2} & ; \mathbf{X} \in \text{Region } 3 \\ \dfrac{(b_1X_1 + b_2X_2)^2}{b_1^2 + b_2^2} & ; \mathbf{X} \in \text{Region } 4 \\ X_1^2 + X_2^2 & ; \mathbf{X} \in \text{Region } 5 \end{cases}$$

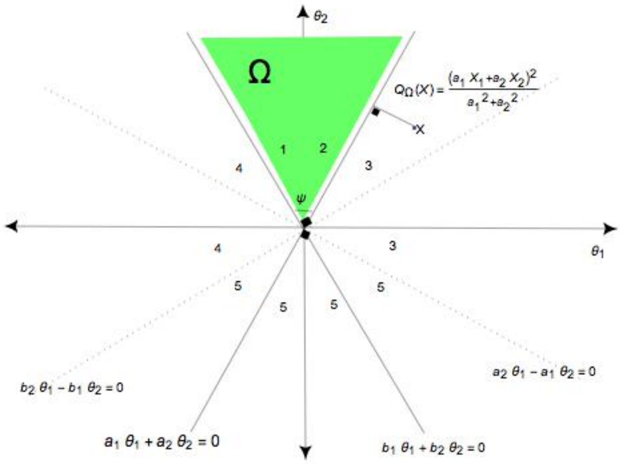

FIGURE 1.2: Cone described in Example 1 of Chernoff, observation $\mathbf{X}$ happens to fall in Region 3

and

$$Q_{\mathcal{T}}(\mathbf{X}) = \begin{cases} \dfrac{(a_1X_1 + a_2X_2)^2}{a_1^2 + a_2^2} & ; \mathbf{X} \in \text{Region 2} \\ \dfrac{(b_1X_1 + b_2X_2)^2}{b_1^2 + b_2^2} & ; \mathbf{X} \in \text{Region 1} \\ 0 & ; \mathbf{X} \notin \Omega. \end{cases}$$

As we see the distribution of d_n^* depends on the proportion of the plane covered by Ω and the angle ψ. If we assign probabilities to the regions in Figure 1.2 as

$$P(\mathbf{X} \in \text{region i}) = p_{\psi i}$$

where $\sum_{i=1}^{5} p_{\psi i} = 1$, then we have for $c \in \mathbb{R}$

$$\begin{aligned} P(d_n^* \leq c) &= \sum_{i=1}^{5} P(d_n^* \leq c | \mathbf{X} \in \text{region i}) P(\mathbf{X} \in \text{region i}) \\ &= p_{\psi 1}(1 - P_1(-c)) + p_{\psi 2}(1 - P_1(-c)) + p_{\psi 3}P_1(c) + p_{\psi 4}P_1(c) + p_{\psi 5}P_2(c), \end{aligned}$$

where P_1 is the cumulative distribution function (c.d.f) of the chi square distribution with one degree of freedom and P_2 is the c.d.f of the chi square distribution with two degrees of freedom. In terms of d_n for which we have $Q_{\Omega\cup\mathcal{T}}(\mathbf{X}) = 0$, the c.d.f of d_n is

$$P(d_n \leq c) = \begin{cases} 0 & ; c < 0 \\ p_{\psi 1} + p_{\psi 2} + (p_{\psi 3} + p_{\psi 4})P_1(c) + p_{\psi 5}P_2(c) & ; c \geq 0. \end{cases}$$

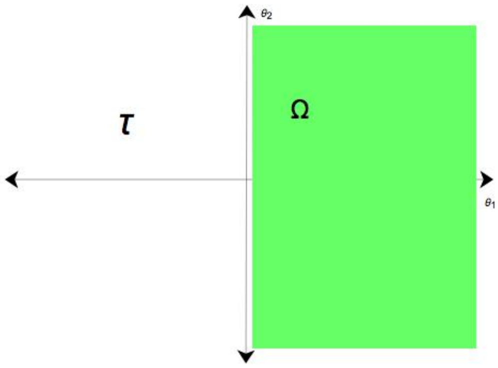

FIGURE 1.3: Regions for the hypothesis test (1.8)

Chernoff in the last paragraph of his Introduction and in a Remark at the end of his paper, mentions this situation and predicts that the distribution of d_n should be different from that of d_n^* as indeed it is according to our calculations.

Corollary 1.1. *In Example 1.1, we can show that the direction ($=\frac{\mathbf{X}}{||\mathbf{X}||}$) and the length ($= ||\mathbf{X}||$) of $\mathbf{X}$ are statistically independent. Therefore, the conditional distribution of $(X_1^2 + X_2^2)$, given that $\mathbf{X}$ falls in region i, is the same as that of its unconditional distribution (Silvapulle and Sen (2005), page 65).*

If we assume $a_1 = a_2 = b_2 = 0$ and $b_1 = 1$, then the hypothesis test in (1.7) is simplified to

$$\begin{cases} H_0 & : \theta_1 \geq 0, \quad \theta_2 \text{ is unknown} \\ H_A & : \theta_1 < 0, \end{cases} \tag{1.8}$$

where Figure 1.3 shows the regions under H_0. Now we have

$$Q_\Omega(\mathbf{X}) = \begin{cases} X_1^2 & ; X_1 \leq 0 \\ 0 & ; X_1 > 0 \end{cases}, \quad Q_\tau(\mathbf{X}) = \begin{cases} 0 & ; X_1 \leq 0 \\ X_1^2 & ; X_1 > 0, \end{cases}$$

and because $Q_{\Omega\cup\tau}(\mathbf{X}) = 0$, therefore

$$d_n^* = \begin{cases} X_1^2 & ; X_1 \leq 0 \\ -X_1^2 & ; X_1 > 0 \end{cases}, \quad d_n = \begin{cases} X_1^2 & ; X_1 \leq 0 \\ 0 & ; X_1 > 0. \end{cases}$$

We have from the example that X_1 has a $N(0,1)$ distribution (since we assume that $\boldsymbol{\theta}_0 = (0,0)$), therefore the distribution of d_n^*, for $c \in \mathbb{R}$, is

$$\begin{aligned} P(d_n^* \leq c) &= P(d_n^* \leq c|X_1 \leq 0)P(X_1 \leq 0) + P(d_n^* \leq c|X_1 > 0)P(X_1 > 0) \\ &= \frac{1}{2}P_1(c) + \frac{1}{2}(1 - P_1(-c)), \end{aligned}$$

having the density

$$f_{d_n^*}(y) = \frac{1}{2\sqrt{2\pi|y|}} \exp\{-\frac{|y|}{2}\}, \quad y \in \mathbb{R}. \tag{1.9}$$

By contrast, d_n has the distribution of a 50:50 mixture of a point mass at zero and a chi square distribution with one degree of freedom.

Example 1.2 (CHERNOFF)**.** *In the second example, Chernoff extends the first example to the case when* $\boldsymbol{\Sigma}$ *is not an identity matrix, but is a known positive definite symmetric matrix. He decomposes the inverse of this matrix as* $\boldsymbol{\Sigma}^{-1} = \mathbf{T}^T\mathbf{T}$*, writes*

$$(\mathbf{X} - \boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\theta}) = ||\mathbf{T}\mathbf{X} - \mathbf{T}\boldsymbol{\theta}||^2$$

and sets $\mathbf{Y} = \mathbf{TX}$*, where* $\mathbf{T}$ *is an upper triangular matrix and* $\mathbf{T}^T$ *is its transpose. Assuming as before that* $\boldsymbol{\theta}_0 = (0,0)$*, the distribution of* $\mathbf{Y}$ *is bivariate normal with mean* $\mathbf{0}$ *and variance matrix* $\mathbf{I}$*. He continues by considering two vectors* $\mathbf{a} = (a_1, a_2)$ *and* $\mathbf{b} = (b_1, b_2)$ *and lets* $\mathbf{Ta} = \mathbf{d}$ *and* $\mathbf{Tb} = \mathbf{e}$*. This problem, now, can be treated as in Example 1, when* $\mathbf{a}$ *and* $\mathbf{b}$ *are replaced by* $\mathbf{d}$ *and* $\mathbf{e}$*. The angle of the cone is here the angle between* $\mathbf{d}$ *and* $\mathbf{e}$*. We can calculate it via*

$$\cos\psi = \frac{\mathbf{d}^T\mathbf{e}}{||\mathbf{d}||||\mathbf{e}||} = \frac{\mathbf{a}^T\boldsymbol{\Sigma}^{-1}\mathbf{b}}{\sqrt{(\mathbf{a}^T\boldsymbol{\Sigma}^{-1}\mathbf{a})(\mathbf{b}^T\boldsymbol{\Sigma}^{-1}\mathbf{b})}},$$

and solve the example as for Example 1.1, replacing ψ *by a modified angle. In the case of an unknown variance* $\boldsymbol{\Sigma}$*, a consistent estimator of* $\boldsymbol{\Sigma}$ *can be used, but in that case the distribution of* $\mathbf{Y} = \mathbf{TX}$ *will no longer be normal.*

Example 1.3 (CHERNOFF)**.** *Chernoff in his third example considers the hypothesis test*

$$\begin{cases} H_0 & : \theta_1 > 0, \quad \theta_2 > 0 \\ H_A & : \theta_1 = 0, \quad \theta_2 = 0, \end{cases} \tag{1.10}$$

when $\boldsymbol{\Sigma} = \mathbf{I}$*. Figure 1.4 shows the regions* Ω *and* τ*. We have* $Q_\tau(\mathbf{X}) = X_1^2 + X_2^2$ *and*

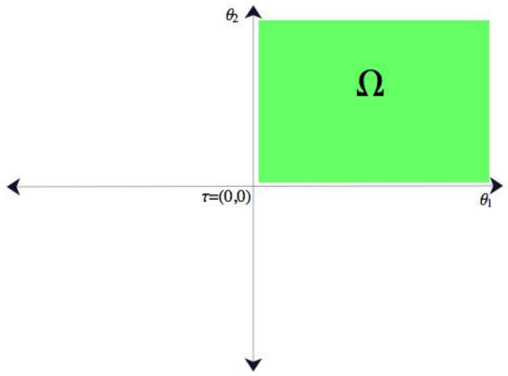

FIGURE 1.4: Regions for the hypothesis test (1.10)

$$Q_\Omega(\mathbf{X}) = \begin{cases} 0 & ; X_1 \geq 0, X_2 \geq 0 \\ X_2^2 & ; X_1 > 0, X_2 < 0 \\ X_1^2 & ; X_1 < 0, X_2 > 0 \\ X_1^2 + X_2^2 & ; X_1 < 0, X_2 < 0 \end{cases}, \quad d_n^* = \begin{cases} -(X_1^2 + X_2^2) & ; X_1 \geq 0, X_2 \geq 0 \\ -X_1^2 & ; X_1 > 0, X_2 < 0 \\ -X_2^2 & ; X_1 < 0, X_2 > 0 \\ 0 & ; X_1 < 0, X_2 < 0. \end{cases}$$

In this example, $d_n = 0$ *for all values of* $\mathbf{X}$*. The distribution of* d_n^* *is*

$$P(d_n^* \leq -c) = 1 - P(-d_n^* > c) = \begin{cases} 0 & ; c < 0 \\ \frac{1}{4} + \frac{1}{2}P_1(c) + \frac{1}{4}P_2(c) & ; c \geq 0. \end{cases}$$

Recall that a set $C \subseteq \mathbb{R}^d$ is a cone with vertex at 0 if $\boldsymbol{\theta} \in C$ implies $a\boldsymbol{\theta} \in C$ for all $a > 0$. Chernoff introduces the idea of a set $\phi \subset \mathcal{N}$ approximated by the cone C_ϕ at $\mathbf{0}$ if

$$\inf_{\mathbf{x} \in C_\phi} ||\mathbf{x} - \mathbf{y}|| = o(||\mathbf{y}||) \quad \text{for } \mathbf{y} \in \phi$$

and

$$\inf_{\mathbf{y} \in \phi} ||\mathbf{x} - \mathbf{y}|| = o(||\mathbf{x}||) \quad \text{for } \mathbf{x} \in C_\phi,$$

then proves the following theorem:
Suppose $\boldsymbol{\theta}_0 = \mathbf{0}$ and $\widehat{\boldsymbol{\theta}}_\phi$ is the maximum likelihood estimator in a set $\phi \subset \mathcal{N}$ and
1) the regularity conditions (CH1), (CH2), (CH3) are satisfied,
2) the origin is a boundary point of ϕ implies that $\widehat{\boldsymbol{\theta}}_\phi \xrightarrow{\text{P}} 0$, for any $\phi \subset \mathcal{N}$,
3) the sets Ω and τ are approximated by nonnull and disjoint cones C_Ω and C_τ.
Then the asymptotic distribution of d_n^* is the same as it would be for the test of $\boldsymbol{\theta} \in C_\Omega$ against $\boldsymbol{\theta} \in C_\tau$ based on one observation from a normal distribution with mean $\mathbf{0}$ and

variance $\mathbf{J}^{-1}$. In this setup $\mathbf{J} = \mathbf{S_0}$ with

$$\mathbf{S}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right]; \quad \boldsymbol{\theta} \in \Theta, \tag{1.11}$$

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathrm{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta})\right]; \quad \boldsymbol{\theta} \in \Theta. \tag{1.12}$$

Feder (1968), Moran (1971), Chant (1974) and Self and Liang (1987) extend Chernoff's result for when the true value of the parameter may be on the boundary of the parameter space (for example, when the inference is about testing $\kappa = 0$ in the von Mises Fisher distribution) with a rigorous analysis which considers some sub cases in detail.

Self and Liang (1987) prove the existence of a consistent maximum likelihood estimator, find the asymptotic distribution of that estimator and the deviance statistic when they consider the maximum likelihood estimator, in a loose sense, to be any point in the parameter space at which a local maximum of the likelihood function occurs. They allow the true parameter value to be on the boundary of the parameter space. They develop the previous work done by Moran (1971) and Chant (1974) in a rigorous way. They define a cone with vertex at $\boldsymbol{\theta}_0$, symbolised by C, to be a set of points such that if $\mathbf{x} \in C$ then $a(\mathbf{x} - \boldsymbol{\theta}_0) + \boldsymbol{\theta}_0 \in C$ for any real, nonnegative number a. They denote $C - \boldsymbol{\theta}_0$ to be a set obtained by translating a cone C with vertex at $\boldsymbol{\theta}_0$ so that its vertex lies at the origin.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n independent observations on $\mathbf{X}$. Self and Liang assume the existence of the first three derivatives of the log likelihood function with respect to $\boldsymbol{\theta}$ on the intersection of neighborhoods of the true parameter value and Θ. In the case of having the true parameter value, denoted by $\boldsymbol{\theta}_0$, on the boundary of Θ, the derivatives of the log likelihood function are taken from the appropriate side. Also, they assume that n^{-1} times of the absolute value of the third derivative of the log likelihood function is bounded by a function of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ whose expectation exists on the intersection of neighborhoods of $\boldsymbol{\theta}_0$ and Ω. Finally they assume that $n^{-1}\mathbf{F}_{\boldsymbol{\theta}}$ is a positive definite matrix on a neighborhood of $\boldsymbol{\theta}_0$ and at $\boldsymbol{\theta}_0$ is equal to the variance matrix of $n^{-1/2}\mathbf{S}_n(\boldsymbol{\theta}_0)$. To prove the consistency of the maximum likelihood estimator when $\boldsymbol{\theta}_0$ is on the boundary of Θ, Self and Liang assume that near $\boldsymbol{\theta}_0$, Θ behaves like a closed set and specifically, they assume that the intersection of Θ and the closure of the neighborhoods centered about $\boldsymbol{\theta}_0$ constitute closed subsets of $\mathbb{R}^d$.

Under the above assumptions, Self and Liang prove the existence and consistency of the maximum likelihood estimator represented by $\widehat{\boldsymbol{\theta}}_n$. To find the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$, they consider a random variable $\mathbf{Z}$ having normal distribution with mean $\boldsymbol{\theta}$ and variance matrix $\mathbf{S}_{\boldsymbol{\theta}_0}^{-1}$ where $\boldsymbol{\theta}$ is restricted to lie in $C_\Theta - \boldsymbol{\theta}_0$. Let F be the distribution of the maximum likelihood estimator of $\boldsymbol{\theta}$ based on a single observation of $\mathbf{Z}$ when $\boldsymbol{\theta} = \mathbf{0}$, symbolised by $\widehat{\boldsymbol{\theta}}$. Then under the above assumptions, the asymptotic distribution of $n^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is F. A question that arises here concerns the cone $C_\Theta - \boldsymbol{\theta}_0$ which is discussed in detail in the following three cases.

Case 1 in Self and Liang. If $\boldsymbol{\theta}_0$ is an interior point of Ω, then $C_\Theta - \boldsymbol{\theta}_0$ is $\mathbb{R}^d$.

Case 2 in Self and Liang. If $\Theta = \Theta_1 \times \ldots \times \Theta_d$ where the Θ_is are closed intervals in $\mathbb{R}^1$ and θ_{01} is a left endpoint of Θ_1 and the other θ_{0i}s are interior points of Θ_is for $2 \le i \le d$, then $C_\Theta - \boldsymbol{\theta}_0$ is $[0, \infty) \times \mathbb{R}^{d-1}$.

Case 3 in Self and Liang. Consider Θ to be similar to Case 2 and $\theta_{01}, \ldots, \theta_{0q}$ to be left endpoints of $\Omega_1, \ldots, \Omega_q$ and θ_{0i}s to be interior points of Ω_i for $q+1 \le i \le d$.

Cases 4 to 9 in Self and Liang derive the distribution of the deviance statistic, when the hypothesis is to test whether θ_0 lies in a subset of Θ symbolised by Ω, versus the alternative that $\boldsymbol{\theta}_0$ lies in the complement of Ω, symbolised by τ. The distribution of the deviance statistic is χ^2 with $d-r$ degrees of freedom when Ω is an r dimensional subset of Θ, $\boldsymbol{\theta}_0$ is a boundary point of both Ω and τ, but $\boldsymbol{\theta}_0$ is an interior point of Θ. When the dimension of Ω and τ is equal to the dimension of Θ and $\boldsymbol{\theta}_0$ is a boundary point of both Ω and τ, Chernoff (1954) presents a method for finding the asymptotic distribution of the deviance statistic. Self and Liang (1987) generalise this case to when $\boldsymbol{\theta}_0$ is a boundary point of Θ. They assume that Ω and τ are regular enough to be approximated by cones with vertices at $\boldsymbol{\theta}_0$ and consider six different cases that we discuss in the following. They assume Θ can be written as $\Theta_1 \times \ldots \Theta_d$ where Θ_is are either closed, half open, or open intervals in $\mathbb{R}^1$. They divide the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ into four categories and consider the hypothesis test $\theta_1 = \theta_{01}, \ldots, \theta_{q+s} = \theta_{0q+s}$, when there is no interest in $\theta_{q+s+1}, \ldots, \theta_d$ and they are left unknown so they call them nuisance parameters. The first q coordinates $\theta_1, \ldots, \theta_q$ have their true values on the boundary. The true parameter is not on the boundary for the next s coordinates $\theta_{q+1}, \ldots, \theta_{q+s}$. The next t coordinates in $\boldsymbol{\theta}$ are nuisance parameters and have their true values on the boundary. The remaining $d-q-s-t$ nuisance coordinates do not have the true value of the parameter on the boundary. Self and Liang present their methodology based on the four-tuple $(q, s, t, d-q-s-t)$. Let $\widetilde{C}$ and $\widetilde{C}_0$ be the approximating cones for Θ and Ω respectively.

Case 4 in Self and Liang. Consider $(0, s, 0, d-s)$ which has no parameters on the boundary. s parameters are set in the null and alternative hypotheses and $d-s$ remain unspecified. In this case, C_Θ is $\mathbb{R}^d$ and $\widetilde{C}_0$ is a $d-s$ dimensional subspace of $\mathbb{R}^d$.

Case 5 in Self and Liang. Let the parameter configuration be $(1, 0, 0, d-1)$. Therefore one parameter of interest has its true value on the boundary. In this case, $\widetilde{C} = [0, \infty) \times \mathbb{R}^{d-1}$ and $\widetilde{C}_0 = \{0\} \times \mathbb{R}^{d-1}$.

Case 6 in Self and Liang. Let the parameter configuration be $(1, 1, 0, d-2)$. Then $\widetilde{C}$ is $[0, \infty) \times \mathbb{R}^{d-1}$ and $\widetilde{C}_0$ is $\{0\} \times \{0\} \times \mathbb{R}^{d-2}$.

In Cases 5 and 6 there is only one parameter on the boundary and it occurs as a parameter of interest. Cases 7 and 8 cover the cases when this condition does not hold. Also, in these two cases the orthogonal structure of $C_\Theta - \boldsymbol{\theta}_0$ and $C_{\Theta_0} - \boldsymbol{\theta}_0$ is not preserved upon the transformation to $\widetilde{C}$ and $\widetilde{C}_0$.

Case 7 in Self and Liang. Let the parameter configuration be $(2, d-2, 0, 0)$. Therefore, $\widetilde{C}$ is as shown in Figure 1 on page 608 of Self and Liang and $\widetilde{C}_0$ is the origin. It is necessary to take into account that the angle in $\widetilde{C}$ is less than 180°, because of the convexity of the $C_\Theta - \boldsymbol{\theta}_0$ which is preserved by the linear mapping into $\widetilde{C}$. We note that if the orthogonal structure is retained then $\widetilde{C} = [0, \infty) \times [0, \infty)$ and $\widetilde{C}_0 = \{0\} \times \{0\}$.

Case 8 in Self and Liang. Let the parameter configuration be $(1, 0, 1, d-2)$. If the orthogonal structure is retained, then $\widetilde{C}$ is $[0, \infty) \times [0, \infty)$ and $\widetilde{C}_0$ is $\{0\} \times [0, \infty)$. However, in the case of nonorthogonality, Figure 2 in Self and Liang shows $\widetilde{C}$. The half line in the direction of θ_2 is $\widetilde{C}_0$.

When the information matrix is diagonal, the orthogonal structure of $C_\Theta - \boldsymbol{\theta}_0$ and $C_{\Theta_0} - \boldsymbol{\theta}_0$ is preserved upon transformation to $\widetilde{C}$ and $\widetilde{C}_0$ and Self and Liang write the following explicit case.

Case 9 in Self and Liang. Let $\widetilde{C} = \widetilde{\Omega}_1 \times \ldots \times \widetilde{\Omega}_d$. Subject to having the true value of the parameter on the boundary or not, $\widetilde{\Omega}_i$ is either $[0, \infty)$ or $(-\infty, \infty)$. Likewise, $\widetilde{C}_0$ is $\widetilde{\Omega}_{01} \times \widetilde{\Omega}_{0d}$, where $\widetilde{\Omega}_{0i} = \mathbb{R}^1$ for $q + s + 1 \leq i \leq d$ and $\widetilde{\Omega}_{0i} = \{0\}$ for $1 \leq i \leq q + s$.

A more general definition of a cone is in Silvapulle and Sen (2005). They use the Hausdorff distance between two sets in a metric space (see van der Vaart and Wellner (1996)) and summarise the work done by Chernoff (1954), Rockafellar and Wets (1998) Andrews (1999) and Geyer (1994), and provide a convenient way of calculating the approximating cone of ϕ at $\boldsymbol{\theta}_0$ when ϕ is defined by a set of nonlinear equality and inequality constraints, in many situations. The set ϕ might be the parameter space or any subset of it. Figure 4.3 on page 187 of Silvapulle and Sen (2005) confirms the argument discussed by Self and Liang about the approximated cone for a sphere.

The work done by Wilks and Chernoff is under the assumption that the sample independently comes from the same probability function $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$, the log likelihood function is in the form of (1.3) and the maximum likelihood estimator (MLE) is verified through maximizing (1.3). However, Vu and Zhou consider an estimating function $\mathcal{L}_n(\boldsymbol{\theta})$ rather than (1.3). They call the maximum of this function a maximum estimator (ME). The methodology described by Vu and Zhou (1997) does not require the existence and boundedness of the third derivative of $\mathcal{L}_n(\boldsymbol{\theta})$ and proves the existence and uniqueness of the ME on a neighborhood of $\boldsymbol{\theta}_0$ and its consistency for $\boldsymbol{\theta}_0$ as the sample size tends to infinity. They prove that the distribution of d_n (obtained based on MEs) is the same as the distribution of

$$\inf_{\boldsymbol{\theta}\in\widetilde{C}_\Omega} ||\mathbf{N}-\boldsymbol{\theta}||^2 - \inf_{\boldsymbol{\theta}\in\widetilde{C}_\tau} ||\mathbf{N}-\boldsymbol{\theta}||^2,$$

where $\widetilde{C}_\Omega$ and $\widetilde{C}_\tau$ are approximating cones for Ω and τ and satisfy some further assumptions. $\mathbf{N} = (N_1, N_2, \cdots, N_d)$ is a random vector which has a multivariate normal distribution with mean zero and a positive definite covariance matrix $\mathbf{V}$.

In the Vu and Zhou methodology, when (1.5) is satisfied (which need not necessarily hold), $\mathbf{V}$ is $\mathbf{I}$, the identity matrix. Their work generalizes that of Chernoff as it deals with more general estimating functions (not only an i.i.d. case) and makes less assumptions about $\mathcal{L}_n(\boldsymbol{\theta})$. The following example illustrates a sample application of Vu and Zhou (1997).

Example 1.4. *Consider a random sample* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *drawn from a bivariate normal distribution with mean* $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ *and known variance matrix* $\boldsymbol{\Sigma}$*. If we wish to test*

$$\begin{cases} H_0 & : \mu_1 = \mu_2 \\ H_A & : \mu_1 \neq \mu_2, \end{cases} \tag{1.13}$$

then $\mathbf{N} = (N_1, N_2)$*,* $\boldsymbol{\theta} = (\mu_1, \mu_2)$*,* $\Omega = \{(\mu_1, \mu_2)|\mu_1 = \mu_2 \in \mathbb{R}\}$ *and* $\tau = \{(\mu_1, \mu_2)|\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \mu_1 \neq \mu_2\}$ *(Figure 1.5). Therefore, we have* $\inf_{\boldsymbol{\theta}\in\Omega\cup\tau} ||\mathbf{N}-\boldsymbol{\theta}||^2 = \inf_{\boldsymbol{\theta}\in\tau} ||\mathbf{N}-\boldsymbol{\theta}||^2 = 0$ *and*

$$\inf_{\boldsymbol{\theta}\in\Omega} ||\mathbf{N}-\boldsymbol{\theta}||^2 = \frac{(N_1 - N_2)^2}{2}.$$

So the distribution of d_n^* *is the same as that of* d_n *and both have chi square distributions with one degree of freedom.*

Maller (2015) further develops the work of Vu and Zhou (1997) and relaxes the assumptions in order to allow the second derivative matrix of $\mathcal{L}_n(\boldsymbol{\theta})$ to be singular and to impose constraints on the parameter space. His work generalises that of Aitchison and Silvey (1958) which we discuss in Section 1.2.4.

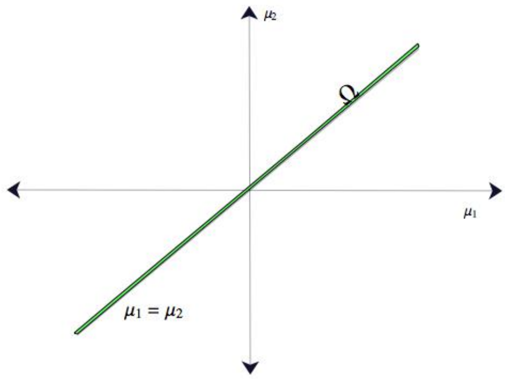

FIGURE 1.5: Regions for the hypothesis test (1.13)

1.2.2 Silvapulle and Sen

In this section we present some results of Silvapulle and Sen and discuss how they relate to ours. They consider two types of hypothesis testing:
Type A. Testing $H_0 : \boldsymbol{\theta} \in \mathcal{M}$ against $H_A : \boldsymbol{\theta} \in \mathcal{C}$,
Type B. Testing $H_0 : \boldsymbol{\theta} \in \mathcal{C}$ against $H_A : \boldsymbol{\theta} \notin \mathcal{C}$,
where $\mathcal{M}$ and $\mathcal{C}$ are subsets of a Euclidean space. $\mathcal{M}$ is a linear space and $\mathcal{C}$ is a closed convex cone and $\mathcal{M} \subset \mathcal{C}$.
Assume $\mathbf{V}$ to be a positive definite matrix and $\mathbf{Z}_{p\times 1}$ to be a normal random variable with mean $\mathbf{0}$ and variance $\mathbf{V}$ and $\overline{\chi}^2(\mathbf{V}, \mathcal{L})$ to have the same distribution as

$$\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z} - \min_{\boldsymbol{\theta}\in\mathcal{C}}(\mathbf{Z}-\boldsymbol{\theta})^T\mathbf{V}^{-1}(\mathbf{Z}-\boldsymbol{\theta}).$$

Silvapulle and Sen (2005) on page 75 name this distribution “Chi-Bar-Square” and provide a method to calculate it through simulation. They find the distribution of the deviance statistic for the foregoing Type A and Type B hypothesis testings based on this distribution, when there is a single observation from a normal distribution with mean $\boldsymbol{\theta}$ and variance $\mathbf{V}$.

Consider an i.i.d. case where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_d$ are from a common density function $f(\mathbf{X};\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$. In our notation, recall $\mathcal{L}_n(\boldsymbol{\theta})$ from (1.3). In a standard setup, the score function vector $\mathbf{S}_n(\boldsymbol{\theta}) = \dfrac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ from (1.4) has expectation 0 and the “information matrix” (the expected value of the negative of the second derivative matrix) for one observation from $\mathbf{S}_{\boldsymbol{\theta}}$ is given by (1.11). Assume $\boldsymbol{\theta}_0 \in A \subset \Theta$ and $\widehat{\boldsymbol{\theta}}_A$ a MLE over A where

$$\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_A) = \sup_{\boldsymbol{\theta}\in A} \mathcal{L}_n(\boldsymbol{\theta})$$

and A might be the parameter space Θ or the sets defined under null and alternative hypotheses. To find the asymptotic distribution of the deviance statistic for Type A and B hypothesis tests based on an estimating function $\mathcal{L}_n(\boldsymbol{\theta})$, the consistency of $\widehat{\boldsymbol{\theta}}_A$ is required. To establish the consistency of $\widehat{\boldsymbol{\theta}}_A$ for $\boldsymbol{\theta}_0$, Silvapulle and Sen (2005), page 146, require two strong conditions to be satisfied: there is a function $\mathcal{L}_0(\boldsymbol{\theta})$ (free of $\mathbf{x}$) such that

(1) $\sup_{\boldsymbol{\theta}\in A}\left|\frac{\mathcal{L}_n(\boldsymbol{\theta})}{n}-\mathcal{L}_0(\boldsymbol{\theta})\right| \xrightarrow{\text{P}} 0$ and

(2) $\sup_{\boldsymbol{\theta}\in A,||\boldsymbol{\theta}-\boldsymbol{\theta}_0||>\varepsilon} \mathcal{L}_0(\boldsymbol{\theta}) < \mathcal{L}_0(\boldsymbol{\theta}_0), \quad \varepsilon > 0.$

These two conditions ensure that there is a global maximiser $\boldsymbol{\theta}_0$ of $\mathcal{L}_0(\boldsymbol{\theta})$ which is unique with probability tending to one. To find the asymptotic null distribution of the deviance statistic for testing against inequality constraints, Silvapulle and Sen set out the following conditions named "Conditions Q":

(1) Distinct values of $\boldsymbol{\theta}$ correspond to distinct distributions.

(2) The first three partial derivatives of $\log f(x;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ exist almost everywhere.

(3) There exists a function $G(x)$ such that $\int G(y)dy < \infty$ and the absolute values of the first three partial derivatives of $\log f(x;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are bounded by $G(x)$ in a neighborhood of $\boldsymbol{\theta}_0$.

(4) The information matrix, $\mathbf{S}_{\boldsymbol{\theta}}$, in this case, is finite and positive definite.

An interesting situation arises in studying the von Mises Fisher distribution. It turns out that the variance matrix of the score function vector for one observation in a von Mises Fisher distribution is not equal to the Fisher information matrix of this distribution that is, the negative expectation of the second derivative matrix. If Condition (4) is not satisfied, the results of Proposition 4.2.1 of Silvapulle and Sen's book, which are fundamental for the rest of the book, cannot be applied for the von Mises Fisher distribution.

Proposition 1.2. *The expectation of the score function vector from the von Mises Fisher distribution is not zero and the variance matrix of the score function vector is not equal to n times the information matrix. Furthermore, the variance matrix of the score function vector is positive semidefinite. Therefore, Condition (4) of Silvapulle and Sen does not hold for the von Mises Fisher distribution.*

Proof of Proposition 1.2. For a brief discussion of this part, see Section 2.4.1 of Chapter 2. □

Often, the parameter of interest in a hypothesis is more complex than just the mean of a multivariate normal. For example, consider the hypotheses

$$\begin{cases} H_0 & : g(\boldsymbol{\theta}) = \mathbf{0} \\ H_A & : g(\boldsymbol{\theta}) \geq \mathbf{0}, \end{cases}$$

where g is a nonlinear function of $\boldsymbol{\theta}$. Such hypotheses are not in the form of Type A and Type B above. For a general hypothesis testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_A : \boldsymbol{\theta} \in \Theta_1$, the asymptotic null distribution of d_n depends on the local shapes of Θ_0 and Θ_1 at the assumed true value in the null parameter space. It turns out that the main results are unchanged when the null and alternative parameter spaces are replaced by cones that approximate them at the true value of the parameter provided that the boundaries of the parameter spaces are sufficiently smooth. Section 4.7 of Silvapulle and Sen introduces the basics which are necessary to study the asymptotic distributions.
Section 4.5.3 of Silvapulle and Sen considers a wider class of tests when $\mathcal{L}_n(\boldsymbol{\theta})$ is an estimation function. Condition A of this section assumes that the variance matrix of the vector of score function is nonsingular which is not applicable for the von Mises Fisher distribution (in the proof of Theorem 2.3, part (b), we show that the determinant of the variance matrix of the vector of score function is zero).

In the Vu and Zhou methodology, the conditions are milder than Silvapulle and Sen and there is no need to check for the existence of the third derivative of the log likelihood function, but still needs the positive definiteness of the expected negative second derivative to be valid.

1.2.3 Rotnitzky, Cox, Bottai and Robins

Rotnitzky, Cox, Bottai and Robins (2000) start with a simple case of an example from an “informative non- response” model which is discussed in Heckman (1976) and Diggle and Kenward (1994). The negative expectation of the second derivative of the log likelihood for a simple and informal case of this model is singular at a specific parameter point. On page 247 of their paper, they use the terminology “information matrix” for the negative expected second derivative of the log likelihood function. In view of Proposition 1.2, we will avoid using the terminology “information matrix” and specify which matrix we are using at any time.

They assume that exactly the first $s-1$ derivatives of $\log f(y;\theta)$ are, with probability 1, equal to 0 at the point $\theta = \theta^*$ and with probability greater than 0, $\partial^s \log f(Y;\theta^*)/\partial\theta^s \neq 0$. Here, we use their notation $\boldsymbol{\theta}^*$ for the true value of $\boldsymbol{\theta}$.

In a multidimensional case, consider the density function $f(\mathbf{y};\boldsymbol{\theta})$ in which $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ is a $d \times 1$ vector of parameters and the intention is to estimate this parameter vector. Assume the negative of the expectation of the second derivative of the log likelihood function of this density is singular and of rank $d-1$ and the derivatives of $\log f(\mathbf{y};\boldsymbol{\theta})$ exist up to a specific order. They denote by $S_j(\boldsymbol{\theta})$ the score with respect to θ_j, $S_j(\boldsymbol{\theta}) = \partial \log f(\mathbf{Y};\boldsymbol{\theta})/\partial\theta_j$, $1 \leq j \leq d$ and by S_j the value $S_j(\boldsymbol{\theta}^*)$. The rank of the negative of the expectation of the second derivative of the log likelihood function at $\boldsymbol{\theta}^*$ is $d-1$ if and only if $d-1$ elements of the score function vector, for instance, the last $d-1$ scores,

$$S_2, \ldots, S_d \quad \text{are linearly independent}$$

and the remaining score is equal to a linear combination of them, that is

$$S_1 = \mathbf{K}^T(S_2, \ldots, S_d) \tag{1.14}$$

for some $(d-1)$ dimensional vector of constants $\mathbf{K}$. Under this assumption, they show that the MLEs of some or all of the components of $\boldsymbol{\theta}^*$ will converge to $\boldsymbol{\theta}^*$ at rates slower than $O_p(n^{-1/2})$. Furthermore, they derive the asymptotic distribution of the MLE of $\boldsymbol{\theta}^*$ and of the likelihood ratio test statistic for testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^*$ versus $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ (Theorem 3 and 4 of the paper).

First, they consider a more specific case where the following two assumptions: (a) the score corresponding to θ_1 is zero at θ^*, that is $S_1 = 0$ and $K = 0$ in equation (1.14); and (b) higher-order partial derivatives of the log-likelihood with respect to θ_1 at θ^* are possibly also zero, but the first non-zero higher-order partial derivative of the log-likelihood with respect to θ_1 evaluated at θ_1^* is not a linear combination of the scores $S_2, \ldots, S_d$.

They prove that, under assumptions (a) and (b) and some other assumptions named (A1)-(A7) and (B1)-(B3), the MLE of $\boldsymbol{\theta}^*$ exists, is unique with probability tending to one and this MLE is a consistent estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}^*$. Also they derive the asymptotic distribution of $d_n = 2\left(\mathcal{L}_n(\widehat{\boldsymbol{\theta}}) - \mathcal{L}_n(\boldsymbol{\theta}^*)\right)$ for the test of $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^*$ against $H_A : \boldsymbol{\theta} \neq \boldsymbol{\theta}^*$. This distribution depends on whether s is odd or even. If s is odd the asymptotic distribution of the likelihood ratio test statistic is chi square with d degrees of freedom. If s is even, then the asymptotic distribution of d_n is a mixture of chi square distributions with d and $d-1$ degrees of freedom and mixing probabilities equal to $1/2$. Then they go further and reparameterize the likelihood function in a way to consider the cases in which the assumptions (a) and (b) do not hold. However, the equation (1.14) is required to be valid even after reparameterization as is said on page 265 of Rotnitzky, Cox, Bottai and Robins (2000).

Proposition 1.3. *In the von Mises Fisher distribution, because the last element of the score function vector is not a linear combination of the other elements, ((1.14) is not valid and* $x_d = \sqrt{1 - x_1^2 - \ldots - x_{d-1}^2}$*), so the distribution does not satisfy the assumptions of Rotnitzky, Cox, Bottai and Robins (2000).*

1.2.4 Aitchison and Silvey: Lagrange Multiplier Test

Aitchison and Silvey (1958) detail a method of dealing with hypothesis tests using Lagrange multipliers. They find the supremum of the likelihood function for an i.i.d. sample subject to constraints in a subset of the parameter space using Lagrange multipliers and consider the estimation of a parameter lying in a subset of a set of possible parameters. This subset is defined by a well-behaved function and the estimator is a solution of the likelihood function containing a Lagrange multiplier. They assume an independent sample from a distribution $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ and set some assumptions. They prove that, under their assumptions, a local maximum of the likelihood function exists and its asymptotic distribution is a multivariate normal with a specific mean vector and covariance matrix.

This setup is as follows. Assume Ω to be a subset of Θ and $\mathbf{h}(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}), \cdots, h_s(\boldsymbol{\theta})) = \mathbf{0}$ to denote s constraints on Θ. Let $\mathbf{h}(\boldsymbol{\theta})$ have first partial derivative $\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{h}(\boldsymbol{\theta})$. Consider the matrix

$$\begin{pmatrix} \mathbf{F}_{\boldsymbol{\theta}} & -\mathbf{H}(\boldsymbol{\theta}) \\ -\mathbf{H}^T(\boldsymbol{\theta}) & \mathbf{0} \end{pmatrix}^{-1}_{(d+s)\times(d+s)} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{pmatrix},$$

where $\mathbf{F}_{\boldsymbol{\theta}} = \mathrm{E}\{\mathbf{F}_1(\boldsymbol{\theta})\}$ is defined in (1.12). Assume $\mathbf{F}_{\boldsymbol{\theta}}$ is nonsingular and $\mathbf{H}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta})$ is invertible, then

$$\boldsymbol{\Sigma} = \mathbf{F}_{\boldsymbol{\theta}}^{-1} - \mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta})(\mathbf{H}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta}))^{-1}\mathbf{H}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}}^{-1}, \tag{1.15}$$

$$\mathbf{R} = -(\mathbf{H}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta}))^{-1} \tag{1.16}$$

and

$$\mathbf{Q} = \mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta}) \left(\mathbf{H}^T(\boldsymbol{\theta})\mathbf{F}_{\boldsymbol{\theta}}^{-1}\mathbf{H}(\boldsymbol{\theta})\right)^{-1}. \tag{1.17}$$

The formulae (1.15), (1.16) and (1.17) follow from standard formulae in matrix theory which can be found, for example, in Rao's book (1973), page 33. Since we are in the i.i.d. case, $\mathbf{Z}_n = \frac{\mathbf{S}_n(\boldsymbol{\theta})}{\sqrt{n}}$ can be assumed, under some regularity assumptions, to be distributed asymptotically normal with mean $\mathbf{0}$ and the variance matrix $\mathbf{S}_{\boldsymbol{\theta}}$ defined in (1.11).

Aitchison and Silvey (1958) are interested in the emergence of $\widehat{\boldsymbol{\theta}}$ as a solution of the equations

$$\frac{\mathbf{S}_n(\boldsymbol{\theta})}{n} + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda} = \mathbf{0}$$
$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0},$$

where $\boldsymbol{\lambda}$ is a Lagrange multiplier in $\mathbb{R}^s$. They prove that the distribution of maximum likelihood estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\lambda}}$ in a given subspace Ω is then given by

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\text{D}} \begin{pmatrix} \mathbf{F}_{\boldsymbol{\theta}} & -\mathbf{H}(\boldsymbol{\theta}) \\ -\mathbf{H}^T(\boldsymbol{\theta}) & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}_n \\ \mathbf{0} \end{pmatrix}, \quad \text{n} \to \infty, \tag{1.18}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\lambda}_0$ are the true values of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ in Θ and $\mathbb{R}^s$ and that, under their regularity assumptions, when $\mathbf{F}_{\boldsymbol{\theta}} = \mathbf{S}_{\boldsymbol{\theta}}$, we have

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\text{D}} \mathbf{N}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R} \end{pmatrix}\right), \quad n \to \infty, \tag{1.19}$$

This result is also in Neueschwander and Flury (1997), Theorem 1.1.

Proof. We have from the assumption and (1.18) that

$$\mathbf{Z}_n \xrightarrow{\text{D}} \mathbf{N}\left(\mathbf{0}, \mathbf{S}_{\boldsymbol{\theta}}\right), \quad \text{as } n \to \infty$$

therefore, the random vector on the left hand side of (1.19) is also asymptotically normal with mean $\mathbf{0}$ and variance matrix

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{Q}^T \\ \mathbf{Q} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{\boldsymbol{\theta}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}\mathbf{S}_{\boldsymbol{\theta}}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\mathbf{S}_{\boldsymbol{\theta}}\mathbf{Q} \\ \mathbf{Q}^T\mathbf{S}_{\boldsymbol{\theta}}\boldsymbol{\Sigma} & \mathbf{Q}^T\mathbf{S}_{\boldsymbol{\theta}}\mathbf{Q} \end{pmatrix}.$$

The relations

$$\begin{aligned}\mathbf{\Sigma}\mathbf{S}_{\boldsymbol\theta}\mathbf{\Sigma} &= \mathbf{\Sigma},\\ \mathbf{\Sigma}\mathbf{S}_{\boldsymbol\theta}\mathbf{Q} &= \mathbf{0},\\ \mathbf{Q}^T\mathbf{S}_{\boldsymbol\theta}\mathbf{\Sigma} &= \mathbf{0},\\ \mathbf{Q}^T\mathbf{S}_{\boldsymbol\theta}\mathbf{Q} &= -\mathbf{R}\end{aligned} \tag{1.20}$$

can be seen through (1.15), (1.16), (1.17) and the assumption that $\mathbf{S}_{\boldsymbol\theta} = \mathbf{F}_{\boldsymbol\theta}$ which comes from (1.5). We have

$$\begin{aligned}\mathbf{\Sigma}\mathbf{S}_{\boldsymbol\theta}\mathbf{\Sigma} = &\left(\mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\right)\mathbf{S}_{\boldsymbol\theta}\\ &\left(\mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\right)\\ &= \left(\mathbf{I} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\right)\\ &\left(\mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\right)\\ &= \mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\\ &- \left(\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\right)\\ &\left(\mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\right)\\ &= \mathbf{\Sigma} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\\ &+ \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\\ &= \mathbf{\Sigma};\end{aligned}$$

the last equality results from $\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right) = \mathbf{I}$. Also

$$\begin{aligned}\mathbf{Q}^T\mathbf{S}_{\boldsymbol\theta}\mathbf{Q} &= \left(\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\right)^T\mathbf{S}_{\boldsymbol\theta}\left(\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\right)\\ &= \left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-T}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\\ &= \left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-T}\\ &= -\mathbf{R}\end{aligned}$$

and

$$\begin{aligned}\mathbf{\Sigma}\mathbf{S}_{\boldsymbol\theta}\mathbf{Q} = &\left(\mathbf{F}_{\boldsymbol\theta}^{-1} - \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\right)\\ &\mathbf{S}_{\boldsymbol\theta}\left(\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\right)\\ &= \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\\ &- \mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta))^{-1}\mathbf{H}^T\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\left(\mathbf{H}^T(\boldsymbol\theta)\mathbf{F}_{\boldsymbol\theta}^{-1}\mathbf{H}(\boldsymbol\theta)\right)^{-1}\\ &= \mathbf{0}.\end{aligned}$$

□

Silvey (1959) considers the case when $\mathbf{F}_{\boldsymbol{\theta}}$ is not a positive definite matrix and suggests replacing $\mathbf{F}_{\boldsymbol{\theta}}$ by $\mathbf{F}^*_{\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}} + \mathbf{H}_1(\boldsymbol{\theta})^T \mathbf{H}_1(\boldsymbol{\theta})$, where $\mathbf{H}_1(\boldsymbol{\theta})$ is a sub matrix of $\mathbf{H}(\boldsymbol{\theta})$.

Maller (2015) develops this idea, adapting the Lagrangian multiplier test to consider the case when equations (1.5) are no longer valid so as to explore the behavior of a restricted maximum estimator; see Appendix A.

1.2.5 Neuenschwander and Flury Definition of Constraints

Neuenschwander and Flury (1997) classify the constraints imposed on a parameter space into three categories, named *model*, *identifiability* and *basic* constraints. Then they distinguish between these three types of constraints by considering those that are only there to simplify the parametric model (*model constraints*), or without them the parameter space is not identifiable (*identifiability constraints*), or without the constraints the family of statistical models is not defined (*basic constraints*). They define a parameter $\boldsymbol{\theta} \in \Theta$ to be identifiable, if (under the regularity conditions) it guarantees the nonsingularity of matrix $\mathbf{S}_{\boldsymbol{\theta}}$. We illustrate with some examples.

Example 1.5. *Let X have a multinomial distribution in which an individual can fall into any one of s classes with probability $\frac{\theta_i}{\theta_+}$, for $i = 1, \ldots, s$ and $\theta_+ = \sum_{i=1}^s \theta_i$. The log likelihood function while we have only one observation is*

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{s} x_i \log \theta_i - \log \theta_+,$$

where $x_i = 0$ or 1 and $\sum_{i=1}^s x_i = 1$. Therefore, we have

$$\mathbf{S}(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{x_1}{\theta_1}, \ldots, \frac{x_s}{\theta_s} \right)^T - \frac{1}{\theta_+} \mathbf{1}_s^T$$

and

$$\mathbf{F}(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = Diag \left(\frac{x_1}{\theta_1^2}, \ldots, \frac{x_s}{\theta_s^2} \right) - \frac{1}{\theta_+^2} \mathbf{1}_s \mathbf{1}_s^T,$$

with

$$E\left(\mathbf{F}(\boldsymbol{\theta})\right) = E\left(\mathbf{S}(\boldsymbol{\theta}) \mathbf{S}^T(\boldsymbol{\theta})\right) = \mathbf{S}_{\boldsymbol{\theta}} = \frac{1}{\theta_+} \left(Diag \left(\frac{1}{\theta_1}, \ldots, \frac{1}{\theta_s} \right) - \frac{1}{\theta_+} \mathbf{1}_s \mathbf{1}_s^T \right),$$

where $\mathbf{1}_s$ is a $s \times 1$ vector of ones. The matrix $E(\mathbf{F}(\boldsymbol{\theta}))$ has determinant zero.

Silvey (1959) comments on this example "it is clear that this determinant is zero because we have set in s-dimensional space a parameter that is really $(s-1)$-dimensional. However it is obvious that there is no difficulty about restricted estimation in the subset of Θ in which $\theta_+ = 1$".

Silvey represents the constraint of $\theta_+ = 1$ as an *identifiability* constraint, because it makes the following assumption valid:
(Silvey: Assumption 6B) $\boldsymbol{\theta}_0 \in \Omega$ (Ω is a non null subset of Θ, the parameter space) and for any other point $\boldsymbol{\theta} \in \Omega$, $F(t, \boldsymbol{\theta}) \neq F(t, \boldsymbol{\theta}_0)$ for at least one t. $F(t, \boldsymbol{\theta})$ is the distribution function and $f(t, \boldsymbol{\theta})$ is the density function of $\mathbf{X}$ at t.
Thus, two different values of $\boldsymbol{\theta}$ cannot give the same distribution for $\mathbf{X}$. For the multinomial distribution without the constraint $\theta_+ = 1$, we could choose $\boldsymbol{\theta}_1 = (1/2, 1/3)$ and $\boldsymbol{\theta}_2 = (2/3, 1/3)$. For $t = 0$, we have

$$f(0, \boldsymbol{\theta}_1) = \left(\frac{1}{2}\right)^0 \left(\frac{1}{3}\right)^1 = \frac{1}{3}$$
$$f(0, \boldsymbol{\theta}_2) = \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^1 = \frac{1}{3}.$$

Therefore, the constraint $\theta_+ = 1$ is needed to make the multinomial distribution *identifiable* according to Silvey (1959) and Neuenschwander and Flury (1997), pages 309 and 312.

Silvey (1959) suggests replacing the singular matrix $\mathbf{F}_{\boldsymbol{\theta}}$ by $\mathbf{F}_{\boldsymbol{\theta}} + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{H}_1^T(\boldsymbol{\theta})$, where $\mathbf{H}_1(\boldsymbol{\theta})$ is a submatrix of $\mathbf{H}(\boldsymbol{\theta})$. In Silvey's methodology the two matrices $\mathbf{F}_\theta$ and $\mathbf{S}_{\boldsymbol{\theta}}$ are assumed to be equal. However, as Theorem 2.3 shows, the negative expectation of second derivative of the log likelihood function and the variance matrix of the score function vector are not equal. Notice that $\mathbf{F}_{\boldsymbol{\theta}}$ is the negative expectation of the second derivative of the log likelihood function for one observation and $\mathbf{S}_{\boldsymbol{\theta}}$ is the expectation of the square of the score function vector for one observation.

Example 1.5. (Continued) *Let* $h(\boldsymbol{\theta}) = \theta_+ - 1$*. The first derivative of* $h(\boldsymbol{\theta})$ *with respect to* $\boldsymbol{\theta}$ *is*

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} h(\boldsymbol{\theta}) = \mathbf{1}_s^T,$$

so

$$\mathbf{S}_{\boldsymbol{\theta}} + \mathbf{H}(\boldsymbol{\theta})\mathbf{H}^T(\boldsymbol{\theta}) = Diag\left(\frac{1}{\theta_1}, \ldots, \frac{1}{\theta_s}\right),$$

which is a positive definite matrix.

There are other cases where constraints arise from the model. Silvey has another contingency table example.

Example 1.6 (Test of Homogeneity in a 2×2 table)**.** *Assume two different populations defined by A_1 and A_2 and take a random sample of size n_1 from A_1 and independently take a random sample of size n_2 from A_2. Also, assume the distribution of n_1 (and as a result n_2) to be binomial with probability*

$$\frac{n!}{n_1!n_2!}(\theta_1+\theta_2)^{n_1}(\theta_3+\theta_4)^{n_2}. \tag{1.21}$$

Consider a 2×2 contingency table 1.1 with a multinomial distribution with the probability

$$\frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\theta_1^{n_{11}}\theta_2^{n_{12}}\theta_3^{n_{21}}\theta_4^{n_{22}}. \tag{1.22}$$

Therefore, the conditional distribution of n_{ij} given n_1 and n_2 is (Cochran (1952))

Table 1.1: Numbers and their related probabilities for a contingency table in Example 1.6

	B_1	B_2	Total
A_1	$n_{11}(\theta_1)$	$n_{12}(\theta_2)$	$n_1(\theta_1+\theta_2)$
A_2	$n_{21}(\theta_3)$	$n_{22}(\theta_4)$	$n_2(\theta_3+\theta_4)$
Total	m_1	m_2	n

$$\left(\frac{n_1!}{n_{11}!n_{12}!}\left(\frac{\theta_1}{\theta_1+\theta_2}\right)^{n_{11}}\left(\frac{\theta_2}{\theta_1+\theta_2}\right)^{n_{12}}\right)\left(\frac{n_2!}{n_{21}!n_{22}!}\left(\frac{\theta_3}{\theta_3+\theta_4}\right)^{n_{21}}\left(\frac{\theta_4}{\theta_3+\theta_4}\right)^{n_{22}}\right).$$

The parameter space is $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^4 : \epsilon \le \theta_i \le \frac{1}{\epsilon}, for\ i = 1,2,3,4\}$, where ϵ is a small positive number and the true value of the parameter is $\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \theta_{03}, \theta_{04})$ which is known to belong to the set Θ. We are interested in whether the equality between the two conditional probabilities

$$P(B_1|A_1) = P(B_1|A_2),$$

holds or not. We have

$$P(B_1|A_1) = \frac{P(B_1 \cap A_1)}{P(A_1)} = \frac{\theta_1}{\theta_1+\theta_2}$$

and

$$P(B_1|A_2) = \frac{P(B_1 \cap A_2)}{P(A_2)} = \frac{\theta_3}{\theta_3+\theta_4}.$$

Therefore, the test for homogeneity in this example is

$$\begin{cases} H_0 & : \frac{\theta_1}{\theta_1+\theta_2} = \frac{\theta_3}{\theta_3+\theta_4} \\ H_A & : \frac{\theta_1}{\theta_1+\theta_2} \neq \frac{\theta_3}{\theta_3+\theta_4}. \end{cases} \tag{1.23}$$

The log likelihood function is

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\theta}) &= \log(\frac{n_1!}{n_{11}!n_{12}!}) + n_{11}\log\frac{\theta_1}{\theta_1+\theta_2} + n_{12}\log\frac{\theta_2}{\theta_1+\theta_2} \\ &+ \log(\frac{n_2!}{n_{21}!n_{22}!}) + n_{21}\log\frac{\theta_3}{\theta_3+\theta_4} + n_{22}\log\frac{\theta_4}{\theta_3+\theta_4} \\ &= constant + n_{11}\log\theta_1 + n_{12}\log\theta_2 - (n_{11}+n_{12})\log(\theta_1+\theta_2) \\ &+ n_{21}\log\theta_3 + n_{22}\log\theta_4 - (n_{21}+n_{22})\log(\theta_3+\theta_4), \end{aligned}$$

The first derivative vector is

$$\mathbf{S}_n(\boldsymbol{\theta}) =$$
$$\left(\frac{n_{11}}{\theta_1} - \frac{(n_{11}+n_{12})}{\theta_1+\theta_2}, \frac{n_{12}}{\theta_2} - \frac{(n_{11}+n_{12})}{\theta_1+\theta_2}, \frac{n_{21}}{\theta_3} - \frac{(n_{21}+n_{22})}{\theta_3+\theta_4}, \frac{n_{22}}{\theta_4} - \frac{(n_{21}+n_{22})}{\theta_3+\theta_4}\right)^T,$$

and the negative of the second derivative matrix is

$$\mathbf{F}_n(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n_{11}}{\theta_1^2} - \frac{(n_{11}+n_{12})}{(\theta_1+\theta_2)^2} & -\frac{(n_{11}+n_{12})}{(\theta_1+\theta_2)^2} & 0 & 0 \\ -\frac{(n_{11}+n_{12})}{(\theta_1+\theta_2)^2} & \frac{n_{12}}{\theta_2^2} - \frac{(n_{11}+n_{12})}{(\theta_1+\theta_2)^2} & 0 & 0 \\ 0 & 0 & \frac{n_{21}}{\theta_3^2} - \frac{(n_{21}+n_{22})}{(\theta_3+\theta_4)^2} & -\frac{(n_{21}+n_{22})}{(\theta_3+\theta_4)^2} \\ 0 & 0 & -\frac{(n_{21}+n_{22})}{(\theta_3+\theta_4)^2} & \frac{n_{22}}{\theta_4^2} - \frac{(n_{21}+n_{22})}{(\theta_3+\theta_4)^2} \end{pmatrix}.$$

We have from (1.22)

$$E(n_{11}) = n\theta_1, \quad E(n_{12}) = n\theta_2 \quad E(n_{21}) = n\theta_3, \quad E(n_{22}) = n\theta_4$$

and from (1.21)

$$E(n_1) = n(\theta_1+\theta_2), \quad E(n_2) = n(\theta_3+\theta_4).$$

Hence

$$E(\mathbf{F}_n(\boldsymbol{\theta})) = E(\mathbf{S}_n(\boldsymbol{\theta})\mathbf{S}_n^T(\boldsymbol{\theta})) = n\begin{pmatrix} \frac{1}{\theta_1} - \frac{1}{\theta_1+\theta_2} & -\frac{1}{\theta_1+\theta_2} & 0 & 0 \\ -\frac{1}{\theta_1+\theta_2} & \frac{1}{\theta_2} - \frac{1}{\theta_1+\theta_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\theta_3} - \frac{1}{\theta_3+\theta_4} & -\frac{1}{\theta_3+\theta_4} \\ 0 & 0 & -\frac{1}{\theta_3+\theta_4} & \frac{1}{\theta_4} - \frac{1}{\theta_3+\theta_4} \end{pmatrix}.$$

The determinant of this matrix is zero.
In this example

$$\mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \theta_1 + \theta_2 - 1 \\ \theta_3 + \theta_4 - 1 \\ \theta_1 - \theta_3 \end{pmatrix} = \mathbf{0}$$

and

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

We see that the first two columns of $\mathbf{H}(\boldsymbol{\theta})$ are *basic* constraints and the third is a consequence of the hypothesis test, thus a *model* constraint. Neuenschwander and Flury (1997) postulate that it is better to distinguish between these two different types of constraints. They state "suppose that in an estimation problem the parameter space $\Theta^* \subset \mathbb{R}^d$ is not identifiable, but specifying the values of s_1 constraints $\mathbf{h}_1(\boldsymbol{\theta})$ (typically $\mathbf{0}$) suffices to make it identifiable. Then put $\Theta = \{\boldsymbol{\theta} \in \Theta^* : \mathbf{h}_1(\boldsymbol{\theta}) = \mathbf{0}\}$. We will refer to Θ^* as the hyperspace. Furthermore, suppose that we wish to estimate $\boldsymbol{\theta}$ under s_2 additional constraints imposed on Θ, written as $\mathbf{h}_2(\boldsymbol{\theta}) = \mathbf{0}$. The constraints $h_2(\boldsymbol{\theta}) = \mathbf{0}$ are model constraints and none of them is implied by the identifiability constraints $h_1(\boldsymbol{\theta}) = \mathbf{0}$. The reduced parameter space is $\Theta_0 = \{\boldsymbol{\theta} \in \Theta : \mathbf{h}_2(\boldsymbol{\theta}) = \mathbf{0}\}$". Neuenschwander and Flury (1997) define

$$\mathbf{H}_1(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{h}_1(\boldsymbol{\theta}), \quad \mathbf{H}_2(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{h}_2(\boldsymbol{\theta}),$$

where $\mathbf{H}_1(\boldsymbol{\theta})$ is a $d \times s_1$ and $\mathbf{H}_2(\boldsymbol{\theta})$ is $d \times s_2$ matrices and put

$$\mathbf{H}(\boldsymbol{\theta}) = (\mathbf{H}_1(\boldsymbol{\theta}) \quad \mathbf{H}_2(\boldsymbol{\theta}))$$

and

$$\mathbf{F}^*_{\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}} + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{H}_1^T(\boldsymbol{\theta}).$$

Then they derive the asymptotic distribution of the maximum likelihood estimator $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ (under certain conditions) as the normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ from the leading $d \times d$ submatrix in

$$\begin{pmatrix} \mathbf{F}^*_{\boldsymbol{\theta}_0} & \mathbf{H}(\boldsymbol{\theta}_0) \\ \mathbf{H}^T(\boldsymbol{\theta}_0) & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{pmatrix},$$

that is (from (1.15))

$$\boldsymbol{\Sigma} = \mathbf{F}^{*-1}_{\boldsymbol{\theta}_0} - \mathbf{F}^{*-1}_{\boldsymbol{\theta}_0}\mathbf{H}(\boldsymbol{\theta}_0)\left(\mathbf{H}^T(\boldsymbol{\theta}_0)\mathbf{F}^{*-1}_{\boldsymbol{\theta}_0}\mathbf{H}(\boldsymbol{\theta}_0)\right)^{-1}\mathbf{H}^T(\boldsymbol{\theta}_0)\mathbf{F}^{*-1}_{\boldsymbol{\theta}_0}.$$

Example 1.6 (continued). *Now consider*

$$\mathbf{H}_1(\theta) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

and let $\mathbf{F}_{\boldsymbol{\theta}} = \dfrac{E(\mathbf{F}_n(\boldsymbol{\theta}))}{n}$. *Then we have*

$$\mathbf{F}^*_{\boldsymbol{\theta}} = \mathbf{F}_{\boldsymbol{\theta}} + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{H}_1^T(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\theta_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\theta_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\theta_3} & 0 \\ 0 & 0 & 0 & \frac{1}{\theta_4} \end{pmatrix},$$

which is positive definite. Therefore the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *is*

$$\boldsymbol{\Sigma} = \frac{\theta_{01}\theta_{02}\theta_{03}\theta_{04}}{\theta_{01}\theta_{02} + \theta_{03}\theta_{04}} \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}.$$

Consider $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T$ to be a vector of Lagrange multipliers. The equations

$$\frac{\mathbf{S}_n(\boldsymbol{\theta})}{n} - \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda} = 0,$$

give the restricted maximum likelihood estimators as

$$\begin{aligned}
\widehat{\theta}_1 = \widehat{\theta}_3 &= \frac{m_1}{n}, \\
\widehat{\theta}_2 = \widehat{\theta}_4 &= \frac{m_2}{n}, \\
\widehat{\lambda}_1 &= \frac{n}{m_2}\left(n_1\frac{m_2}{n} - n_{12}\right), \\
\widehat{\lambda}_2 &= \frac{n}{m_2}\left(n_2\frac{m_2}{n} - n_{22}\right), \\
\widehat{\lambda}_3 &= \frac{n}{2m_1}\left(n_{21} - n_{11}\right) + \frac{n}{2m_2}\left(n_{12} - n_{22}\right).
\end{aligned}$$

Hence

$$\mathbf{H}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}} = \begin{pmatrix} \widehat{\lambda}_1 + \widehat{\lambda}_3 \\ \widehat{\lambda}_1 \\ \widehat{\lambda}_2 - \widehat{\lambda}_3 \\ \widehat{\lambda}_2 \end{pmatrix} = \begin{pmatrix} \frac{n}{m_1}(n_1\frac{m_1}{n} - n_{11}) \\ \frac{n}{m_2}(n_1\frac{m_2}{n} - n_{12}) \\ \frac{n}{m_1}(n_2\frac{m_1}{n} - n_{21}) \\ \frac{n}{m_2}(n_2\frac{m_2}{n} - n_{22}) \end{pmatrix},$$

Silvey (1959) introduces the statistic $\widehat{\boldsymbol{\lambda}}^T\mathbf{H}^T(\widehat{\boldsymbol{\theta}})\left(\mathbf{J}_n\mathbf{F}^*_{\widehat{\boldsymbol{\theta}}}\right)^{-1}\mathbf{H}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}}$, where

$$\mathbf{J}_n = \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_1 & 0 & 0 \\ 0 & 0 & n_2 & 0 \\ 0 & 0 & 0 & n_2 \end{pmatrix},$$

for the hypothesis test (1.23) which has asymptotically a χ^2 distribution with 1 degree of freedom. This statistic is

$$\begin{aligned}
\widehat{\boldsymbol{\lambda}}^T\mathbf{H}^T(\widehat{\boldsymbol{\theta}})\left(\mathbf{J}_n\mathbf{F}^*_{\widehat{\boldsymbol{\theta}}}\right)^{-1}\mathbf{H}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\lambda}} &= \frac{(n_1\frac{m_1}{n} - n_{11})^2}{n_1m_1/n} + \frac{(n_1\frac{m_2}{n} - n_{12})^2}{n_1m_2/n} \\
&\quad + \frac{(n_2\frac{m_1}{n} - n_{21})^2}{n_2m_1/n} + \frac{(n_2\frac{m_2}{n} - n_{22})^2}{n_2m_2/n}
\end{aligned}$$

which is the usual statistic used in the chi square test of homogeneity in a 2×2 table.

Example 1.5 illustrates how the problem of a singular second derivative matrix can be overcome with Lagrange multipliers. This theme is developed further in Appendix A which lists recent results by Maller (2015). We use them in Chapter 3.

1.3 Composite Likelihood

Our spherical subcomponent model can be put in the context of the composite likelihood method.

The composite likelihood method is a way of constructing an estimating function. The composite function can be formed from products of marginals only, or from a mixture of marginals and conditionals, or totally by products of conditional distributions. The format of the estimating function is determined by the particular problem involved. Section 2.1 in Varin, Reid, and Firth (2011) explains a variety of possible formats and usages of composite likelihood.

Suppose the composite likelihood is made up of K elements of the form

$$L_c(\boldsymbol{\theta};\mathbf{y}) = \prod_{k=1}^{K} \left(L_k(\boldsymbol{\theta};\mathbf{y})\right)^{w_k}, \tag{1.24}$$

where the w_k are selected weights and are nonnegative. Let

$$\mathcal{CL}(\boldsymbol{\theta};\mathbf{y}) = \log\{L_c(\boldsymbol{\theta};\mathbf{y})\}$$

be the log-composite likelihood, let

$$u(\boldsymbol{\theta};\mathbf{y}) = \frac{\partial}{\partial\boldsymbol{\theta}}\mathcal{CL}(\boldsymbol{\theta};\mathbf{y})$$

be the composite score function and let $\widehat{\boldsymbol{\theta}}_{CL}$ be the maximum of the log-composite likelihood.

Because the composite likelihood is in general not the complete likelihood the derivative of the log-likelihood may not produce an unbiased estimating equation (Varin, Reid and Firth (2011)). Therefore, in general,

$$\mathrm{E}_{\boldsymbol{\theta}}\{u(\boldsymbol{\theta};\mathbf{Y})\} \neq 0$$

and we need to distinguish between

$$H(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}\left(-\frac{\partial}{\partial\boldsymbol{\theta}}u(\boldsymbol{\theta};\mathbf{Y})\right) = \int -\left(\frac{\partial}{\partial\boldsymbol{\theta}}u(\boldsymbol{\theta};\mathbf{y})\right) f(\boldsymbol{\theta};\mathbf{y})d\mathbf{y},$$

which is the expected second derivative of the composite score function, and the variance matrix

$$J(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}\left(u(\boldsymbol{\theta};\mathbf{Y})\right).$$

Thus, we replace the information matrix by the sandwich information matrix

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}H(\boldsymbol{\theta}).$$

Suppose we have n independent and identically distributed observations $\mathbf{y}_1, \ldots, \mathbf{y}_n$ from the model $f(\boldsymbol{\theta}; \mathbf{y})$ from which the components in (1.24) are formed. Under some regularity conditions it can be proved that $\widehat{\boldsymbol{\theta}}_{CL}$ is asymptotically normally distributed as

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{CL} - \boldsymbol{\theta}^*\right) \xrightarrow{\text{D}} \mathbf{N}\left(\mathbf{0}, G(\boldsymbol{\theta}^*)^{-1}\right),$$

where the function $\mathrm{E}\{\mathcal{CL}(\boldsymbol{\theta}; \mathbf{y})\}$ achieves its maximum at $\boldsymbol{\theta}^*$, assumed to be an interior point of the parameter space.

If we define

$$I(\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\theta}}\left(\frac{\partial}{\partial\boldsymbol{\theta}} \log f(\boldsymbol{\theta}; \mathbf{Y})\right)$$

to be the expected Fisher information, then the ratio of $G(\boldsymbol{\theta})$ to the expected Fisher information $I(\boldsymbol{\theta})$ gives the asymptotic efficiency of $\widehat{\boldsymbol{\theta}}_{CL}$ relative to the maximum likelihood estimator from the full model (Varin, Reid, and Firth (2011)). Our application to the spherical subcomponent model does not fit into the standard situation as we have to allow for constraints as well as the difference between the expected second derivative of the composite score function and its variance matrix.

Chapter 2

von Mises Fisher Distribution on the Sphere

2.1 Introduction

In this chapter, we study the von Mises Fisher distribution, analysing its behavior on the sphere using contour plots and calculating its normalizing constant and its moments. We also introduce some properties of the variance matrix of this distribution which are of interest and useful for our analysis.

We can find formulae for the normalizing constant and the moments of the von Mises Fisher distribution in almost all books related to the spherical distributions. However, we calculated these values via a direct method which is illuminating. Theorem 2.1 concerns the properties of the normalising constant and its derivatives which we can find in Watson (1983), but the method of proving presented in this chapter is different from his. The results of this theorem are used throughout the thesis.
Theorems 2.2 and 2.3 investigate the properties of the variance matrix and score function of this distribution and are what motivated us to write this thesis. Examining the uniform distribution on the sphere in Section 2.5 through considering the value of zero for κ in the density function of the von Mises Fisher distribution helped us to calculate the moments of the uniform distribution on the sphere and to present Theorem 2.4.

2.2 Spherical Data

Angular and spherical data are observations that appear in a wide variety of random experiments. As indicated by the name, they are related to angles. In 2 dimensions observations are on a unit circle with origin in the centre and can be represented by a unit vector which shows the position of the observation on the circumference of the circle. For three or more dimensions, observations are on the surface of a sphere or hypersphere. In the case of a unit circle, the cartesian coordinates of the vectors can be represented by $(\cos\theta^\circ, \sin\theta^\circ)$ and in polar coordinates they are $(1, \theta^\circ)$.

In this thesis, we work with the density of a distribution which involves a random vector $\mathbf{X}$ in Cartesian coordinates. Therefore, we write $x_1 = \cos\theta$ and $x_2 = \sin\theta$ in 2 dimensional coordinates and $x_1 = \sin\theta\cos\phi$, $x_2 = \sin\theta\sin\phi$ and $x_3 = \cos\theta$ in 3 dimensions. In d dimensions we have

$$\begin{aligned}
x_1 &= \prod_{i=1}^{d-1} \sin\theta_i, \\
x_j &= \prod_{i=j}^{d-1} \sin\theta_i \cos\theta_{j-1}, \quad j = 2, \dots, d-1, \\
x_d &= \cos\theta_d.
\end{aligned}$$

Spherical data are employed in many scientific disciplines. We study an example in finance.

Example 2.1 (Portfolio with no short sales (NSS))**.** *In portfolio analysis we mean to present a method to enable us to invest optimally in a group of assets. Consider d assets. The intention is to allocate some proportion of our capital to each of them in order to have the highest return for a given level of risk. If we define X_i to be the fraction of a unit amount of capital invested in asset i and μ_i to be the expected return on asset i, for $i = 1, \dots, d$, then,*

$$\sum_{i=1}^{d} \mu_i X_i,$$

is the expected return on the portfolio. Moreover, we have the total investment condition

$$\sum_{i=1}^{d} X_i = 1. \tag{2.1}$$

Let $\mathbf{\Sigma} = (\sigma_{ij})$ be covariance matrix of the returns, where σ_{ii} is the variance of the return on the asset i and σ_{ij} is the covariance between the returns on asset i and asset j, while

$i \neq j$. Then the risk of the portfolio is given by the square root of its variance

$$\left[\sum_{i=1}^{d} X_i^2\sigma_{ii}^2 + \sum_{\substack{i=1\\i\neq j}}^{d}\sum_{j=1}^{d} X_iX_j\sigma_{ij}\right]^{1/2} = \left[\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X}\right]^{1/2}, \tag{2.2}$$

where $\mathbf{X} = [X_1, X_2, \cdots, X_d]^T$. When short sales are not allowed (NSS), each invested amount must be nonnegative, so we have another condition

$$X_i \geq 0, \quad i = 1, 2, \cdots, d. \tag{2.3}$$

Portfolio analysis helps us to determine the allocation of a unit amount of capital to maximize return in a specific time interval subject to a given level of risk, or, equivalently, to minimize risk for a given level of return. The X_i's are called the asset allocations. To choose a portfolio of assets with the lowest risk and highest return when short sales are not allowed, we minimize (2.2) under the constraints (2.1) and (2.3). Equivalently, (Sharpe (1994)), we maximize the Sharpe ratio

$$\frac{\mathbf{X}^T\boldsymbol{\mu}}{[\mathbf{X}^T\boldsymbol{\Sigma}\mathbf{X}]^{1/2}}, \tag{2.4}$$

under the constraints (2.1) and (2.3), where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)^T$ is the vector of expected returns. The vector of asset allocations $\mathbf{X} = (X_1, X_2, \cdots, X_d)^T$ where each X_i is greater than zero and $\sum_{i=1}^{d} X_i = 1$, can be transformed to vector $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_d)^T$ $= (\sqrt{X_1}, \sqrt{X_2}, \cdots, \sqrt{X}_d)^T$. Then $Y_1^2 + Y_2^2 + \cdots Y_d^2 = 1$, so $\mathbf{Y}$ lies on the unit sphere. With this motivation in mind we go on to study the von Mises Fisher distribution on the sphere.

2.3 von Mises Fisher Distribution and its Moments

Let $\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^d : x_1^2 + x_2^2 + \cdots + x_d^2 = 1\}$ denote a unit sphere in $\mathbb{R}^d$. The von Mises Fisher distribution is defined on this sphere by the following density function:

$$f_{\mathbf{X}}(\mathbf{x}) = c(\kappa)^{-1}\exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\}, \quad \mathbf{x} \in \mathbb{S}^d, \kappa \geq 0, \boldsymbol{\mu} \in \mathbb{S}^d, \tag{2.5}$$

where $f_{\mathbf{X}}(x_1, x_2, \cdots, x_d)$ is the density of a d dimensional random vector $\mathbf{X}$ at the point $\mathbf{x} = (x_1, x_2, \cdots, x_d)$ on the surface of the sphere.

In this distribution $\kappa \geq 0$ represents the *concentration* and $\boldsymbol{\mu}$ is the *mean direction* or the *pole* such that $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_d)$ with $\mu_1^2 + \mu_2^2 + \cdots + \mu_d^2 = 1$. κ is a measure of

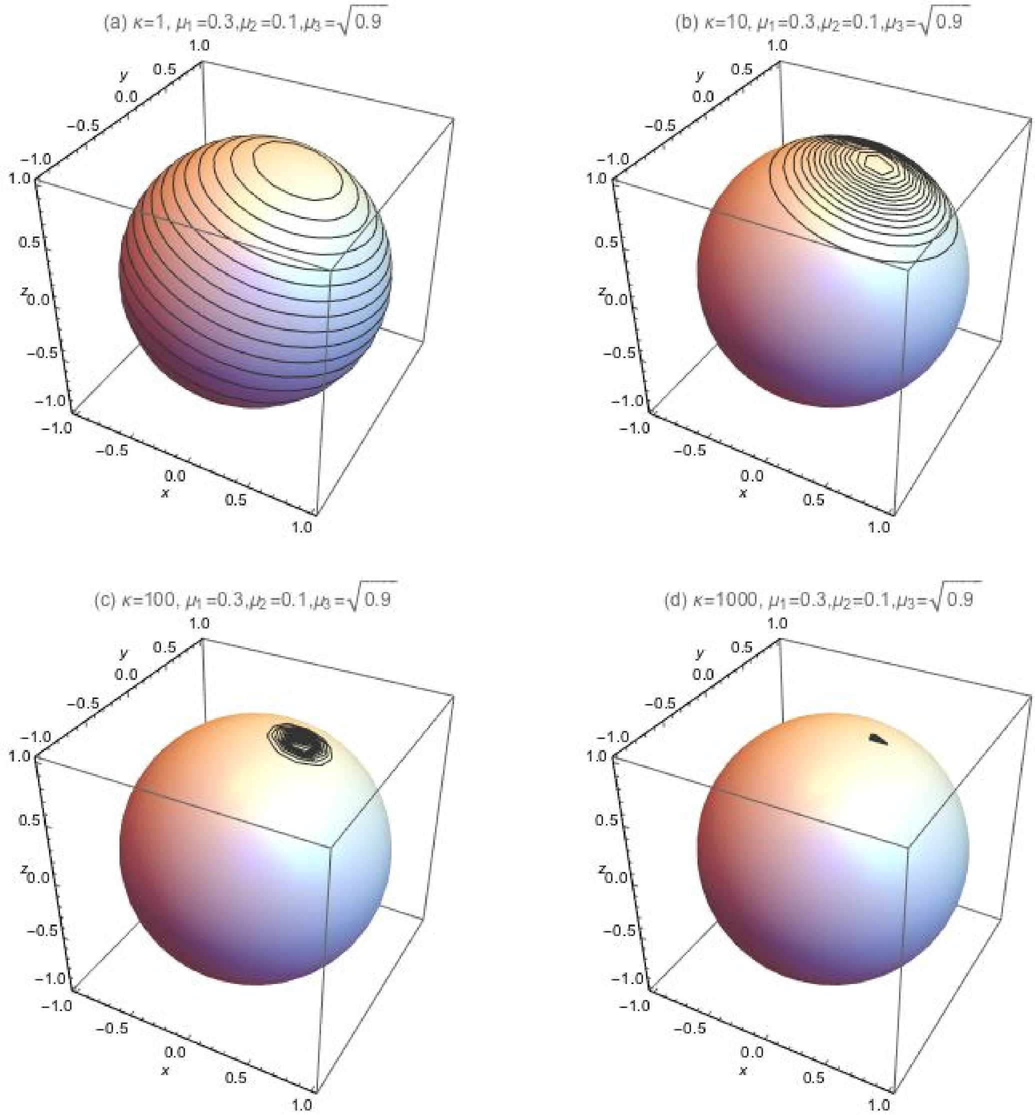

FIGURE 2.1: Comparing contour plots of von Mises Fisher distribution for different values of $\kappa = 1$, $\kappa = 10$, $\kappa = 100$ and $\kappa = 1000$ while $\mu_1 = 0.3$, $\mu_2 = 0.1$, $\mu_3 = \sqrt{0.9}$

precision. If $\kappa = 0$, then the data are distributed uniformly over the sphere. When κ is large, the distribution is concentrated on a small portion of the sphere. $\boldsymbol{\mu}$ is called the modal or mean vector of the distribution and locates the density on the sphere. Some contour plots of the density for different values of κ is shown in Figure 2.1. Figure 2.2 shows the effect of different values of the modal vector.

We assume throughout this section that κ is not zero. Therefore, we choose the parameter space to be $\Theta = \{(\kappa, \mu_1, \cdots, \mu_d) \in (0, \infty) \times (-1, 1)^d\}$ and the true value of the parameter to be $\boldsymbol{\theta}_0 = (\kappa_0, \mu_{10}, \cdots, \mu_{d0}) \in \Theta$.

The first derivative of $c(\kappa)$ with respect to κ will be denoted by c_κ and the second derivative by $c_{\kappa\kappa}$. These notations will be used throughout the thesis.

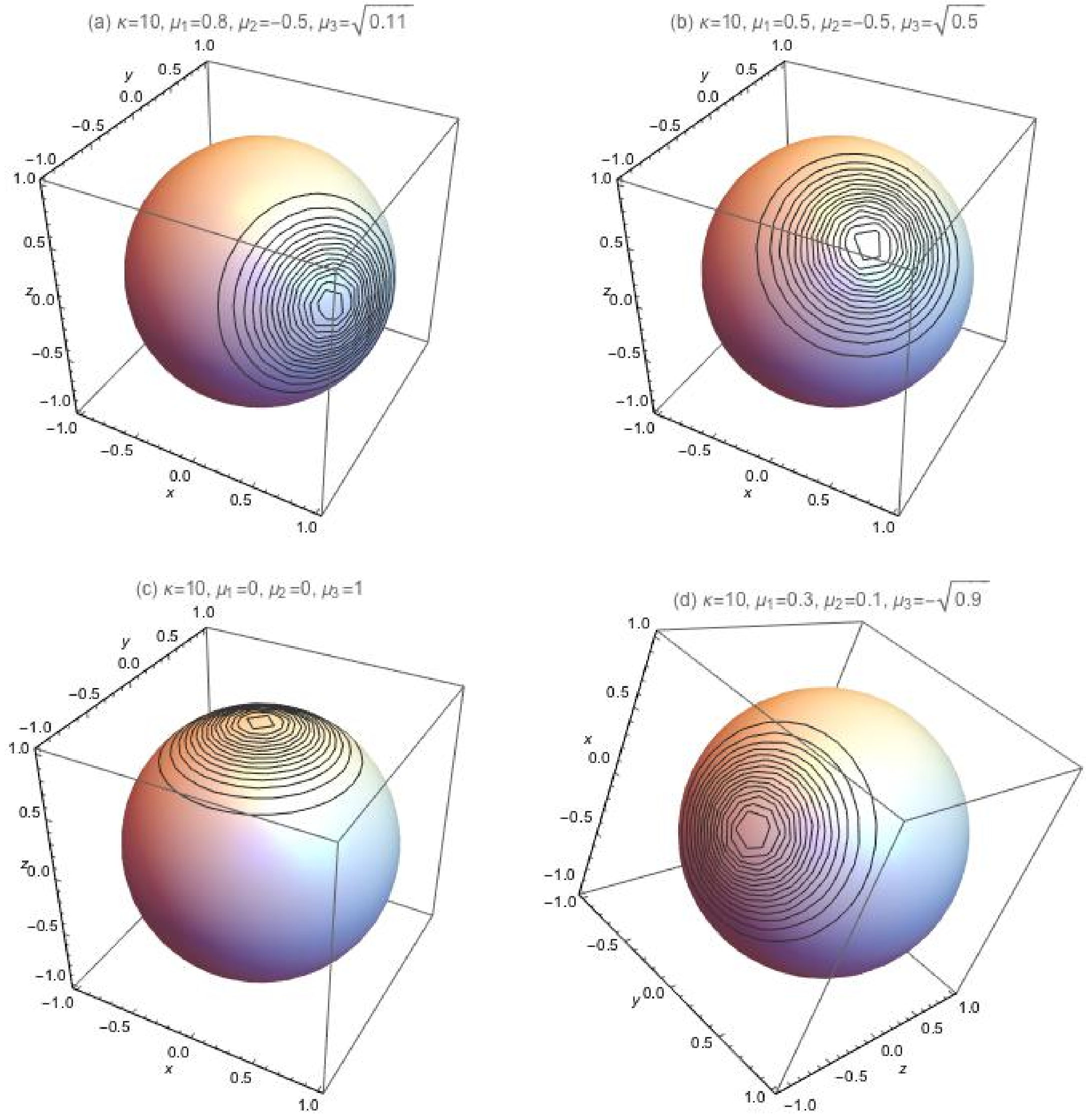

FIGURE 2.2: Comparing contour plots of von Mises Fisher distribution for different modal vectors while $\kappa = 10$

2.3.1 Properties of $c(\kappa)$

Let $\mathbf{X} = (X_1, X_2, \cdots, X_d)^T \in \mathbb{S}^d$ be a d dimensional random vector with the distribution (2.5). The integral of the density function over the sample space must be equal to one. So

$$c(\kappa) = \int_{\mathbf{x}\in\mathbb{S}^d} \exp\{\kappa\boldsymbol{\mu}^T\mathbf{x}\}dw(\mathbf{x}) = \int_{\mathbf{x}\in\mathbb{S}^d} \exp\{\kappa(\mu_1 x_1 + \cdots + \mu_d x_d)\}dw(\mathbf{x}), \tag{2.6}$$

where $dw(\cdot)$ is the area element on $\mathbb{S}^d$ and is calculated as follows.

If we parameterize a 3 dimensional sphere and consider the vector $(i, j, k)^T$ to be the standard basis vector of the Euclidean space, we get $\vec{r}(x, y) = xi + yj + \sqrt{1 - x^2 - y^2}k$ and the area of the parallelogram defined by two vectors $\frac{\partial \vec{r}(x,y)}{\partial x}$ and $\frac{\partial \vec{r}(x,y)}{\partial y}$ is the cross product of these two vectors, that is

$$\frac{\partial \vec{r}(x,y)}{\partial x} \times \frac{\partial \vec{r}(x,y)}{\partial y} = \begin{vmatrix} i & j & k \\ 1 & 0 & \frac{-x}{\sqrt{1-x^2-y^2}} \\ 0 & 1 & \frac{-y}{\sqrt{1-x^2-y^2}} \end{vmatrix}$$
$$= \frac{x}{\sqrt{1-x^2-y^2}} i + \frac{y}{\sqrt{1-x^2-y^2}} j + k,$$

and we have

$$dw(\mathbf{x}) = ||\frac{\partial \vec{r}(x,y)}{\partial x} \times \frac{\partial \vec{r}(x,y)}{\partial y}|| = \sqrt{\left(\frac{x}{\sqrt{1-x^2-y^2}}\right)^2 + \left(\frac{y}{\sqrt{1-x^2-y^2}}\right)^2 + 1}.$$

$c(\kappa)$ in (2.6) appears to depend on $\boldsymbol{\mu}$ as well as κ. However, $c(\kappa)$ depends on κ only. To see this, let

$$\mathbf{M} = \begin{bmatrix} \mu_1 & \mu_d & 0 & \cdots & 0 \\ \mu_2 & 0 & \mu_d & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{d-1} & 0 & 0 & \cdots & \mu_d \\ \mu_d & -\mu_1 & -\mu_2 & \cdots & -\mu_{d-1} \end{bmatrix}_{d\times d} = \begin{bmatrix} \mathbf{a} & \mu_d \mathbf{I}_{(d-1)\times(d-1)} \\ \mu_d & -\mathbf{a}^T \end{bmatrix}, \tag{2.7}$$

where

$$\mathbf{a}^T = (\mu_1, \mu_2, \cdots, \mu_{d-1}), \tag{2.8}$$

and we assume $\mu_d \neq 0$ (at least one μ_j must be nonzero). We look for an orthogonal matrix that satisfies $\boldsymbol{\mu}^T\mathbf{Q} = [1, 0, \cdots, 0]$. We claim that such a $\mathbf{Q}$ does exist. Let $\mathbf{Q} = [q_1, q_2, \cdots, q_d]$. From the orthogonality assumption, we verify that $q_1 = \boldsymbol{\mu}$ and $q_i^T q_j = 0$, for all $i \neq j$, $2 \geq i, j \geq d$. These give us $d + \frac{(d-1)\times(d-2)}{2}$ equations and $(d-1)^2$ variables. Because the number of variables, for $d \geq 2$, is more than the number of equations, we can conclude that $\mathbf{Q}$ exists. Such a matrix can be built upon $\mathbf{M}$.
One of the options for $\mathbf{Q}$ is to build an orthonormal matrix by the Gram-Schmidt method. For $d = 3$, for example,

$$\mathbf{Q} = \begin{bmatrix} \mu_1 & \frac{\mu_3}{\sqrt{\mu_1^2+\mu_3^2}} & \frac{-\mu_1\mu_2}{\sqrt{\mu_1^2+\mu_3^2}} \\ \mu_2 & 0 & \sqrt{\mu_1^2+\mu_3^2} \\ \mu_3 & \frac{-\mu_1}{\sqrt{\mu_1^2+\mu_3^2}} & \frac{-\mu_2\mu_3}{\sqrt{\mu_1^2+\mu_3^2}} \end{bmatrix}. \tag{2.9}$$

Make the transformation $\mathbf{x} = \mathbf{Q}\mathbf{y}$. We can write (2.6) as

$$c(\kappa) = \int_{\mathbf{x}\in\mathbb{S}^d} \exp\{\kappa y_1\} dw(\mathbf{y}), \tag{2.10}$$

on noting that

$$\boldsymbol{\mu}^T\mathbf{x} = \boldsymbol{\mu}^T\mathbf{Q}\mathbf{y} = [1, 0, 0, \cdots, 0]\,\mathbf{y} = y_1, \tag{2.11}$$

where $\mathbf{y} \in \mathbb{S}^d$. (2.10) shows that $c(\kappa)$ depends on κ only.

Using the formula of surface integrals (Silverman (1985)), (2.10) can be written as

$$\begin{aligned} c(\kappa) &= \int_{\mathbb{S}^d} \exp\{\kappa y_1\} dw(\mathbf{y}) \\ &= 2\int_0^1\int_0^1\cdots\int_0^1 \\ &\sqrt{\left(\frac{y_1}{\sqrt{1-y_1^2-\cdots-y_{d-1}^2}}\right)^2+\cdots+\left(\frac{y_{d-1}}{\sqrt{1-y_1^2-\cdots-y_{d-1}^2}}\right)^2+1} \\ &\exp\{\kappa y_1\}dy_{d-1}dy_{d-2}\cdots dy_1 \\ &= 2\int_0^1\int_0^1\cdots\int_0^1 \frac{\exp\{\kappa y_1\}}{\sqrt{1-y_1^2-\cdots-y_{d-1}^2}} dy_{d-1}dy_{d-2}\cdots dy_1 \end{aligned} \tag{2.12}$$

Now make the transformation to polar coordinates:

$$\begin{aligned} y_1 &= \cos\theta_1 \\ y_2 &= \sin\theta_1\cos\theta_2 \\ &\vdots \\ y_i &= \left(\prod_{j=1}^{i-1}\sin\theta_i\right)\cos\theta_i \qquad \text{for} \quad i = 2, 3, \cdots, d-2 \\ y_{d-1} &= \left(\prod_{j=1}^{d-2}\sin\theta_j\right)\cos\theta_{d-1}, \end{aligned} \tag{2.13}$$

so (2.12) is

$$c(\kappa) = 2\int_0^\pi d\theta_{d-1}\int_0^\pi \exp\{\kappa\cos\theta_1\}\sin^{d-2}\theta_1 d\theta_1 \prod_{j=1}^{d-3}\int_0^\pi \sin^j\theta_{d-j-1}d\theta_{d-j-1}. \tag{2.14}$$

Using the equation

$$\int_0^\pi \sin^j x dx = \frac{\sqrt{\pi}\Gamma(\frac{j+1}{2})}{\Gamma(\frac{j+2}{2})}, \quad j \geq 0, \tag{2.15}$$

we can write the above integral as:

$$\begin{aligned} c(\kappa) &= 2 \times \pi \times \pi^{\frac{d-3}{2}} \left(\frac{\Gamma(\frac{d-2}{2})}{\Gamma(\frac{d-1}{2})} \frac{\Gamma(\frac{d-3}{2})}{\Gamma(\frac{d-2}{2})} \cdots \frac{\Gamma(1)}{\Gamma(\frac{3}{2})} \right) \int_0^{\pi/2} \exp\{\kappa \cos\theta_1\} \sin^{d-2}\theta_1 d\theta_1 \\ &= 2\pi^{\frac{d}{2}-\frac{1}{2}} \frac{1}{\Gamma(\frac{d-1}{2})} \frac{\sqrt{\pi}\Gamma(\frac{d-1}{2}) I_{\frac{d-2}{2}}(\kappa)}{(\frac{\kappa}{2})^{\frac{d-2}{2}}} \\ &= (2\pi)^{d/2} \kappa^{1-d/2} I_{d/2-1}(\kappa), \end{aligned} \tag{2.16}$$

where $I_{d/2-1}(\cdot)$ is the modified Bessel function of first kind (Abramowitz and Stegun (1964), Chapter 9, page 376). $I_{d/2-1}(\kappa)$ satisfies the ordinary differential equation

$$\kappa^2 \frac{d^2 W(\kappa)}{d\kappa^2} + \kappa \frac{dW(\kappa)}{d\kappa} - (\kappa^2 + (\frac{d-2}{2})^2) W(\kappa) = 0.$$

In our case, we have $W(\kappa) = (2\pi)^{d/2} \kappa^{d/2-1} c(\kappa)$, so $c(\kappa)$ satisfies

$$\kappa c_{\kappa\kappa} + (d-1) c_\kappa - \kappa c(\kappa) = 0, \tag{2.17}$$

where $c_\kappa = \frac{\partial c(\kappa)}{\partial \kappa}$ and $c_{\kappa\kappa} = \frac{\partial^2 c(\kappa)}{\partial \kappa^2}$.

2.3.2 Moments

From (2.6) we have

$$c_\kappa = \frac{\partial c(\kappa)}{\partial \kappa} = \int_{\mathbf{x} \in \mathbb{S}^d} (\boldsymbol{\mu}^T \mathbf{x}) \exp\{\kappa \boldsymbol{\mu}^T \mathbf{x}\} dw(\mathbf{x}) = c(\kappa) E(\boldsymbol{\mu}^T \mathbf{X}).$$

Thus

$$E(\boldsymbol{\mu}^T \mathbf{X}) = \frac{c_\kappa}{c(\kappa)}. \tag{2.18}$$

By considering the second derivative we see that

$$c_{\kappa\kappa} = \frac{\partial^2 c(\kappa)}{\partial \kappa^2} = \int_{\mathbf{x} \in \mathbb{S}^d} (\boldsymbol{\mu}^T \mathbf{x})^2 \exp\{\kappa \boldsymbol{\mu}^T \mathbf{x}\} dw(\mathbf{x}) = c(\kappa) E(\boldsymbol{\mu}^T \mathbf{X})^2.$$

Thus

$$E(\boldsymbol{\mu}^T \mathbf{X})^2 = \frac{c_{\kappa\kappa}}{c(\kappa)}. \tag{2.19}$$

Put $\mu_d = \sqrt{1 - \mu_1^2 - \mu_2^2 - \cdots - \mu_{d-1}^2} \neq 0$ in (2.6) and consider the first and second derivative of $c(\kappa)$ with respect to μ_i. Let $\tilde{\boldsymbol{\mu}}_i^T = (\mu_d, -\mu_i)^T$ and $\mathbf{X}_{i,d} = (X_i, X_d)^T$ for $1 \leq i \leq d-1$. We find that

$$E(\widetilde{\boldsymbol{\mu}}_i^T \mathbf{X}_{i,d}) = 0 \tag{2.20}$$

$$E(\widetilde{\boldsymbol{\mu}}_i^T \mathbf{X}_{i,d})^2 = \frac{\mu_d^2 + \mu_i^2}{\kappa \mu_d} E(X_d)$$

Solving equations (2.18) and (2.20), we obtain

$$E(X_i) = \frac{c_\kappa}{c(\kappa)} \mu_i, \quad \mathrm{i} = 1, 2, \cdots, \mathrm{d}, \tag{2.21}$$

hence

$$\mathrm{E}\mathbf{X} = \frac{c_\kappa}{c(\kappa)} \boldsymbol{\mu}. \tag{2.22}$$

Therefore, we have

$$E(\widetilde{\boldsymbol{\mu}}_i^T \mathbf{X}_{i,d})^2 = \frac{c_\kappa}{\kappa c(\kappa)} \left(\mu_d^2 + \mu_i^2\right). \tag{2.23}$$

In the next step, the derivative of $c(\kappa)$ is first taken with respect to κ and then with respect to μ_i:

$$\frac{\partial^2 c(\kappa)}{\partial \mu_i \partial \kappa} = \int_{|\mathbf{x}|=1} \left((x_i - \frac{\mu_i}{\mu_d} x_d) + \kappa (x_i - \frac{\mu_i}{\mu_d} x_d)(\mu^T \mathbf{x}) \right) \exp\{\kappa \mu^T \mathbf{x}\} dw(\mathbf{x}).$$

The result is the following equation:

$$E\left((\widetilde{\boldsymbol{\mu}}_i^T \mathbf{X}_{i,d})(\mu^T \mathbf{X}) \right) = 0. \tag{2.24}$$

Alternatively when the first derivative is taken with respect to μ_i and then μ_j, we have

$$\frac{\partial^2 c(\kappa)}{\partial \mu_i \partial \mu_j} = \int_{|\mathbf{x}|=1} \left(-\kappa \frac{\mu_i \mu_j}{\mu_d^3} x_d + \kappa^2 (x_i - \frac{\mu_i}{\mu_d} x_d)(x_j - \frac{\mu_j}{\mu_d} x_d) \right) \exp\{\kappa \boldsymbol{\mu}^T \mathbf{x}\} dw(\mathbf{x}),$$

which gives

$$E(\widetilde{\boldsymbol{\mu}}_i^T \mathbf{X}_{i,d})(\widetilde{\boldsymbol{\mu}}_j^T \mathbf{X}_{j,d}) = \frac{c_\kappa}{\kappa c(\kappa)} \mu_i \mu_j. \tag{2.25}$$

Consider (2.19), (2.23), (2.24) and (2.25). These are the elements of $\mathrm{E}(\mathbf{MXX}^T\mathbf{M}^T)$ for matrix $\mathbf{M}$ satisfying (2.7). Therefore we have

$$E(\mathbf{MXX}^T\mathbf{M}^T) = \frac{c_\kappa}{\kappa c(\kappa)} \begin{bmatrix} \frac{\kappa c_{\kappa\kappa}}{c_\kappa} & 0 & 0 & \cdots & 0 \\ 0 & \mu_1^2+\mu_d^2 & \mu_1\mu_2 & \cdots & \mu_1\mu_{d-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mu_1\mu_{d-1} & \mu_2\mu_{d-1} & \cdots & \mu_{d-1}^2+\mu_d^2 \end{bmatrix}. \tag{2.26}$$

Another representation for this matrix is

$$E(\mathbf{MXX}^T\mathbf{M}^T) = \frac{c_\kappa}{\kappa c(\kappa)} \begin{bmatrix} \frac{\kappa c_{\kappa\kappa}}{c_\kappa} & \mathbf{0}_{d-1}^T \\ \mathbf{0}_{d-1} & \mathbf{F} \end{bmatrix},$$

where

$$\mathbf{F} = \left(\mu_d^2\mathbf{I} + \mathbf{aa}^T\right) \tag{2.27}$$

and $\mathbf{a}$ is defined in (2.8). Then

$$\mathbf{a}^T\mathbf{F} = \mathbf{a}^T\left(\mu_d^2 + \mathbf{aa}^T\right) = \mathbf{a}^T$$

and

$$\begin{aligned} \frac{1}{\mu_d}\left(\mathbf{I} - \mathbf{aa}^T\right)F &= \frac{1}{\mu_d}\left(\mathbf{I} - \mathbf{aa}^T\right)\left(\mu_d^2\mathbf{I} + \mathbf{aa}^T\right) \\ &= \frac{1}{\mu_d}\left(\mu_d^2\mathbf{I} + \mathbf{aa}^T - \mu_d^2\mathbf{aa}^T - \mathbf{aa}^T\mathbf{aa}^T\right) \\ &= \mu_d\mathbf{I}. \end{aligned}$$

So

$$\begin{aligned} \mathrm{E}\mathbf{XX}^T &= \mathbf{M}^{-1}\left(\mathrm{E}\mathbf{MXX}^T\mathbf{M}^T\right)\mathbf{M}^{-T} \\ &= \frac{c_\kappa}{\kappa c(\kappa)} \begin{bmatrix} \mathbf{a} & \frac{1}{\mu_d}(\mathbf{I} - \mathbf{aa}^T) \\ \mu_d & -\mathbf{a}^T \end{bmatrix} \begin{bmatrix} \frac{\kappa c_{\kappa\kappa}}{c_\kappa} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{a}^T & \mu_d \\ \frac{1}{\mu_d}(\mathbf{I} - \mathbf{aa}^T) & -\mathbf{a} \end{bmatrix} \\ &= \frac{c_\kappa}{\kappa c(\kappa)} \begin{bmatrix} \frac{\kappa c_{\kappa\kappa}}{c_\kappa}\mathbf{a} & \mu_d\mathbf{I} \\ \frac{\kappa c_{\kappa\kappa}}{c_\kappa}\mu_d & -\mathbf{a}^T \end{bmatrix} \begin{bmatrix} \mathbf{a}^T & \mu_d \\ \frac{1}{\mu_d}(\mathbf{I} - \mathbf{aa}^T) & -\mathbf{a} \end{bmatrix} \\ &= \frac{c_\kappa}{\kappa c(\kappa)} \begin{bmatrix} \mathbf{I} + \left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right)\mathbf{aa}^T & \mu_d\left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right)\mathbf{a} \\ \mu_d\left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right)\mathbf{a}^T & 1 + \mu_d^2\left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right) \end{bmatrix}. \end{aligned}$$

Or,

$$
\begin{aligned}
\mathrm{E}\mathbf{X}\mathbf{X}^T &= \frac{c_\kappa}{\kappa c(\kappa)}\{\mathbf{I} + \left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right)\begin{bmatrix}\mathbf{a}\mathbf{a}^T & \mu_d\mathbf{a}\\ \mu_d\mathbf{a}^T & \mu_d^2\end{bmatrix}\}\\
&= \frac{c_\kappa}{\kappa c(\kappa)}\mathbf{I}_d + \frac{c_\kappa}{\kappa c(\kappa)}\left(\frac{\kappa c_{\kappa\kappa}}{c_\kappa} - 1\right)\boldsymbol{\mu}\boldsymbol{\mu}^T\\
&= \frac{c_\kappa}{\kappa c(\kappa)}\left(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^T\right) + \frac{c_{\kappa\kappa}}{c(\kappa)}\boldsymbol{\mu}\boldsymbol{\mu}^T. \qquad (2.28)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
trace(\mathrm{E}\mathbf{X}\mathbf{X}^T) &= \frac{dc_\kappa}{\kappa c(\kappa)} + \frac{\kappa c_{\kappa\kappa} - c_\kappa}{\kappa c(\kappa)}\\
&= \frac{(d-1)c_\kappa}{\kappa c(\kappa)} + \frac{c_{\kappa\kappa}}{c(\kappa)}\\
&= \sum_{i=1}^{d}\mathrm{E}X_i^2 = 1.
\end{aligned}
$$

As a result the following relation is verified:

$$\kappa c_{\kappa\kappa} + (d-1)c_\kappa - \kappa c(\kappa) = 0. \qquad (2.29)$$

Note that this is the same as (2.17).

From (2.22) and (2.28), we have

$$\mathrm{Var}(\mathbf{X}) = \frac{c_\kappa}{\kappa c(\kappa)}\left(\mathbf{I} - \boldsymbol{\mu}\boldsymbol{\mu}^T\right) + \left(\frac{c_{\kappa\kappa}c(\kappa) - c_\kappa^2}{c^2(\kappa)}\right)\boldsymbol{\mu}\boldsymbol{\mu}^T. \qquad (2.30)$$

Also from (2.19) and (2.18), we have

$$\mathrm{Var}(\boldsymbol{\mu}^T\mathbf{X}) = \mathrm{E}\left(\boldsymbol{\mu}^T\mathbf{X} - \frac{c_\kappa}{c(\kappa)}\right)^2 = \frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}, \qquad (2.31)$$

from (2.20) and (2.23),

$$\mathrm{Var}(\tilde{\boldsymbol{\mu}}_i^T\mathbf{X}_{i,d}) = \frac{c_\kappa}{\kappa c(\kappa)}\left(\mu_i^2 + \mu_d^2\right), \qquad (2.32)$$

from (2.18), (2.20) and (2.24),

$$\mathrm{Cov}(\boldsymbol{\mu}^T\mathbf{X}, \tilde{\boldsymbol{\mu}}_i^T\mathbf{X}_{i,d}) = 0, \qquad (2.33)$$

and finally from (2.20) and (2.25)

$$\mathrm{Cov}(\tilde{\boldsymbol{\mu}}_i^T\mathbf{X}_{i,d}, \tilde{\boldsymbol{\mu}}_j^T\mathbf{X}_{j,d}) = \frac{c_\kappa}{\kappa c(\kappa)}\mu_i\mu_j. \tag{2.34}$$

These formulae will be needed for the asymptotic analyses to follow later.

2.4 Properties of the Variance Matrix in a von Mises Fisher Distribution

Define

$$A(\kappa) = \frac{c_\kappa}{c(\kappa)}, \tag{2.35}$$

then

$$A'(\kappa) = \frac{c_{\kappa\kappa}}{c(\kappa)} - \frac{c_\kappa^2}{c^2(\kappa)}, \tag{2.36}$$

which we denote as $a(\kappa)$ throughout. Then we have the following theorem:

Theorem 2.1. *(i) For all values of $\kappa > 0$, we have $A(\kappa) > 0$ and $a(\kappa) > 0$. Therefore, $A(\kappa)$ is an increasing function of $\kappa > 0$.*
(ii) For all values of $\kappa > 0$, we have $c_\kappa > 0$ and $c_{\kappa\kappa} > 0$. And $c(\kappa)$ is an increasing function of κ for all $\kappa > 0$.

Proof of Theorem 2.1. (i) Recalling that a variance is always positive, hence $A(\kappa) > 0$ follows from (2.32) and $a(\kappa) > 0$ follows from (2.31).
(ii) Of course $c(\kappa)$ is positive. From Part (i) we have $c_\kappa > 0$. From (2.19) we conclude that $c_{\kappa\kappa}$ is also positive, all for $\kappa > 0$. □

The variance matrix of a von Mises Fisher distribution, which is verified from (2.30), can be rewritten as

$$\boldsymbol{\Sigma} = \frac{A(\kappa)}{\kappa}\left(\mathbf{I}_d - \boldsymbol{\mu}\boldsymbol{\mu}^T\right) + a(\kappa)\boldsymbol{\mu}\boldsymbol{\mu}^T. \tag{2.37}$$

Suppose $\mathbf{B}^{1/2}$ ($\mathbf{B}^{T/2}$) is a left (right) square root of a positive definite matrix $\mathbf{B}$. These are any matrices which satisfy $\mathbf{B}^{1/2}\mathbf{B}^{T/2} = \mathbf{B}$. The matrix $\mathbf{B}^{1/2}$ is clearly not unique. We define $\mathbf{B}^{T/2}$ to be $(\mathbf{B}^{1/2})^T$. The Cholesky square root and the symmetric positive definite square root are the common methods for deriving a square root. In the Cholesky square root method, the left and right square roots $\mathbf{B}^{1/2}$ and $\mathbf{B}^{T/2}$ are denoted as the lower

and upper triangular matrices with positive diagonal elements satisfying $\mathbf{B}^{1/2}\mathbf{B}^{T/2} = \mathbf{B}$ and $\mathbf{B}^{T/2} = (\mathbf{B}^{1/2})^T$.
We have the following results.

Theorem 2.2. *For the variance matrix of a von Mises Fisher distribution represented in (2.37) we have*

$$
\begin{aligned}
(i) \quad & \boldsymbol{\Sigma}^{-1} = \frac{\kappa}{A(\kappa)}\mathbf{I}_d + \left(\frac{1}{a(\kappa)} - \frac{\kappa}{A(\kappa)}\right)\boldsymbol{\mu}\boldsymbol{\mu}^T, \\
(ii) \quad & Det(\boldsymbol{\Sigma}) = \left(\frac{A(\kappa)}{\kappa}\right)^{d-1} a(\kappa), \\
(iii) \quad & \boldsymbol{\Sigma}^{1/2} = \sqrt{\frac{A(\kappa)}{\kappa}}\mathbf{I}_d + \left(\sqrt{a(\kappa)} - \sqrt{\frac{A(\kappa)}{\kappa}}\right)\boldsymbol{\mu}\boldsymbol{\mu}^T, \\
(iv) \quad & \textit{Eigenvalues of } \boldsymbol{\Sigma} \textit{ are } \underbrace{\frac{A(\kappa)}{\kappa}, \ldots, \frac{A(\kappa)}{\kappa}}_{d-1}, \quad \textit{and} \quad a(\kappa).
\end{aligned}
$$

Proof of Theorem 2.2. (i) From the Sherman-Morrison determinant formula (C. Radhakrishna Rao (1973), page 33), the inverse and determinant of a matrix in the form of $\mathbf{B} = \mathbf{A} + \mathbf{U}\mathbf{V}^T$ are

$$
\begin{aligned}
\mathbf{B}^{-1} &= \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{U}\mathbf{V}^T\mathbf{A}^{-1}}{1 + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}} \\
Det(\mathbf{B}) &= (1 + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}).Det(\mathbf{A})
\end{aligned}
$$

where $\mathbf{U}$ and $\mathbf{V}$ are two vectors and $\mathbf{A}$ is an invertible matrix. Therefore, if we consider $\mathbf{A}$ to be $\frac{A(\kappa)}{\kappa}\mathbf{I}_d$, $\mathbf{U} = \left(a(\kappa) - \frac{A(\kappa)}{\kappa}\right)\boldsymbol{\mu}$ and $\mathbf{V} = \boldsymbol{\mu}$, then

$$
\begin{aligned}
\mathbf{A}^{-1} &= \frac{\kappa}{A(\kappa)}\mathbf{I}_d, \\
1 + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U} &= 1 + \boldsymbol{\mu}^T(\frac{\kappa}{A(\kappa)})(a(\kappa) - \frac{A(\kappa)}{\kappa})\boldsymbol{\mu} = \frac{\kappa}{A(\kappa)}a(\kappa), \\
\mathbf{A}^{-1}\mathbf{U}\mathbf{V}^T\mathbf{A}^{-1} &= (\frac{\kappa}{A(\kappa)})^2\left(a(\kappa) - \frac{A(\kappa)}{\kappa}\right)\boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}
$$

and the result (i) comes straightforwardly.

(iii) A suggestion for the square root of the variance matrix can be checked via $\boldsymbol{\Sigma}^{T/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$ on noting that $\boldsymbol{\Sigma}^{1/2}$ is a symmetric matrix.

(iv) For the eigenvalues, we solve the equation

$$Det(\boldsymbol{\Sigma} - \lambda \mathbf{I}_d) = (\frac{A(\kappa)}{\kappa} - \lambda)^d (1 + \frac{a(\kappa) - \frac{A(\kappa)}{\kappa}}{\frac{A(\kappa)}{\kappa} - \lambda}) = 0$$

to get the answer. □

2.4.1 Covariance Matrix of the Score Function Vector of the von Mises Fisher Distribution

In this section, we calculate the covariance matrix of the first derivative of the logarithm of the likelihood function and the negative of the second derivative of the logarithm of the likelihood function of the von Mises Fisher distribution in order to emphasize that they are not equal.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be n identically independent random variables from the von Mises Fisher distribution represented in (2.5). In this case, the log likelihood function is

$$\mathcal{L}_n(\boldsymbol{\theta}) = -n \log c(\kappa) + n\kappa \boldsymbol{\mu}^T \overline{\mathbf{X}},$$

where $\overline{\mathbf{X}} = \sum_{i=1}^{n} \mathbf{X}_i$. We define the score function vector to be

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = n \begin{pmatrix} -A(\kappa) + \boldsymbol{\mu}^T \overline{\mathbf{X}} \\ \kappa \overline{\mathbf{X}} \end{pmatrix},$$

where $A(\kappa) = \frac{c_\kappa}{c(\kappa)}$ is defined in (2.35). The expectation of the score function vector is

$$\mathrm{E}\mathbf{S}_n(\boldsymbol{\theta}) = n \begin{pmatrix} 0 \\ \kappa \mathrm{E}\overline{\mathbf{X}} \end{pmatrix}, \tag{2.38}$$

which is not zero and

$$\{\mathrm{E}\mathbf{S}_n(\boldsymbol{\theta})\}\{\mathrm{E}\mathbf{S}_n(\boldsymbol{\theta})\}^T = n^2 \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \kappa^2 \mathrm{E}\overline{\mathbf{X}}\mathrm{E}\overline{\mathbf{X}}^T \end{pmatrix}. \tag{2.39}$$

From the properties of the von Mises Fisher distribution we have

$$\begin{aligned} \mathrm{E}\mathbf{X} = \mathrm{E}\overline{\mathbf{X}} &= A(\kappa)\boldsymbol{\mu}, \\ \mathrm{E}\boldsymbol{\mu}^T\mathbf{X} = \mathrm{E}\boldsymbol{\mu}^T\overline{\mathbf{X}} &= A(\kappa), \\ \mathrm{Var}(\boldsymbol{\mu}^T\mathbf{X}) = n\mathrm{Var}(\boldsymbol{\mu}^T\overline{\mathbf{X}}) &= \frac{dA(\kappa)}{d\kappa} = a(\kappa), \end{aligned}$$

therefore $\mathbf{S}_n(\boldsymbol{\theta})\mathbf{S}_n^T(\boldsymbol{\theta})$ is

$$\mathbf{S}_n(\boldsymbol{\theta})\mathbf{S}_n^T(\boldsymbol{\theta}) = n^2 \begin{pmatrix} (-A(\kappa) + \boldsymbol{\mu}^T\overline{\mathbf{X}})^2 & \kappa(-A(\kappa) + \boldsymbol{\mu}^T\overline{\mathbf{X}})\overline{\mathbf{X}}^T \\ \kappa(-A(\kappa) + \boldsymbol{\mu}^T\overline{\mathbf{X}})\overline{\mathbf{X}} & \kappa^2\overline{\mathbf{X}}\overline{\mathbf{X}}^T \end{pmatrix}$$

$$= n^2 \begin{pmatrix} (\boldsymbol{\mu}^T\overline{\mathbf{X}} - \mathrm{E}\boldsymbol{\mu}^T\overline{\mathbf{X}})^2 & \kappa\boldsymbol{\mu}^T(\overline{\mathbf{X}} - \mathrm{E}\overline{\mathbf{X}})\overline{\mathbf{X}}^T \\ \kappa\overline{\mathbf{X}}(\overline{\mathbf{X}} - \mathrm{E}\overline{\mathbf{X}})^T\boldsymbol{\mu} & \kappa^2\overline{\mathbf{X}}\overline{\mathbf{X}}^T \end{pmatrix}$$

and has the expectation

$$\mathrm{E}\{\mathbf{S}_n(\boldsymbol{\theta})\mathbf{S}_n^T(\boldsymbol{\theta})\} = n^2 \begin{pmatrix} \mathrm{Var}(\boldsymbol{\mu}^T\overline{\mathbf{X}}) & \kappa\boldsymbol{\mu}^T\mathrm{Var}(\overline{\mathbf{X}}) \\ \kappa\mathrm{Var}(\overline{\mathbf{X}})\boldsymbol{\mu} & \kappa^2\left(\mathrm{Var}(\overline{\mathbf{X}}) + \mathrm{E}\overline{\mathbf{X}}\mathrm{E}\overline{\mathbf{X}}^T\right) \end{pmatrix}. \tag{2.40}$$

From (2.39), the variance matrix of the score function vector is thus

$$\mathrm{Var}\{\mathbf{S}_n(\boldsymbol{\theta})\} = n^2 \begin{pmatrix} \mathrm{Var}(\boldsymbol{\mu}^T\overline{\mathbf{X}}) & \kappa\boldsymbol{\mu}^T\mathrm{Var}(\overline{\mathbf{X}}) \\ \kappa\mathrm{Var}(\overline{\mathbf{X}})\boldsymbol{\mu} & \kappa^2\mathrm{Var}(\overline{\mathbf{X}}) \end{pmatrix}$$

$$= n \begin{pmatrix} a(\kappa) & \kappa\boldsymbol{\mu}^T\mathrm{Var}(\mathbf{X}) \\ \kappa\mathrm{Var}(\mathbf{X})\boldsymbol{\mu} & \kappa^2\mathrm{Var}(\mathbf{X}) \end{pmatrix}$$

$$= n \begin{pmatrix} a(\kappa) & \kappa a(\kappa)\boldsymbol{\mu}^T \\ \kappa a(\kappa)\boldsymbol{\mu} & \kappa^2\mathrm{Var}(\mathbf{X}) \end{pmatrix}. \tag{2.41}$$

It is proved that this matrix is not a positive definite matrix (this proof is done in Theorem 2.3, part (b) in the following). The minus second derivative of the log likelihood is

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial}{\partial\boldsymbol{\theta}^T}\mathbf{S}_n(\boldsymbol{\theta}) = n \begin{pmatrix} a(\kappa) & -\overline{\mathbf{X}}^T \\ -\overline{\mathbf{X}} & 0 \end{pmatrix}$$

with the expectation

$$\mathrm{E}\mathbf{F}_n(\boldsymbol{\theta}) = n \begin{pmatrix} a(\kappa) & -\mathrm{E}\overline{\mathbf{X}}^T \\ -\mathrm{E}\overline{\mathbf{X}} & 0 \end{pmatrix} = n \begin{pmatrix} a(\kappa) & -A(\kappa)\boldsymbol{\mu}^T \\ -A(\kappa)\boldsymbol{\mu} & \mathbf{0} \end{pmatrix} \tag{2.42}$$

which is not equal to $\mathrm{Var}\{\mathbf{S}_n(\boldsymbol{\theta})\}$ (the two matrices (2.41) and (2.42) are not equal) and $\mathrm{E}\mathbf{F}_n(\boldsymbol{\theta})$ is singular.

Theorem 2.3. *(a) The expectation of the score function vector in the von Mises Fisher distribution is not zero (Matrix (2.38)).*

(b) The variance matrix of the score function vector for the von Mises Fisher distribution is not a positive definite matrix (Matrix (2.41)).

(c) The variance matrix of the score function vector for the von Mises Fisher distribution is not equal to the negative of the expectation of the second derivative of the log likelihood function (the two matrices ((2.41) and (2.42)).

(d) The negative expectation of the second derivative of the log likelihood function of the von Mises Fisher distribution is singular (matrix (2.42) has determinant zero).

Proof of Part (b). From (2.37), we have

$$\text{Var}(\mathbf{S}_n(\boldsymbol{\theta})) = n \begin{pmatrix} a(\kappa) & \kappa a(\kappa)\boldsymbol{\mu}^T \\ \kappa a(\kappa)\boldsymbol{\mu} & \kappa^2 \text{Var}(\mathbf{X}) \end{pmatrix} = n \begin{pmatrix} a(\kappa) - \frac{\kappa^2 a^2(\kappa)}{d(\kappa)} & \mathbf{0}^T \\ \mathbf{0} & \kappa A(\kappa)\mathbf{I}_d \end{pmatrix} + \mathbf{U}\mathbf{V}^T \tag{2.43}$$

$$= n(\mathbf{B} + \mathbf{U}\mathbf{V}^T) \tag{2.44}$$

where

$$\mathbf{U} = \begin{pmatrix} \frac{\kappa a(\kappa)}{d(\kappa)} \\ \boldsymbol{\mu} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \kappa a(\kappa) \\ d(\kappa)\boldsymbol{\mu} \end{pmatrix}$$

and

$$d(\kappa) = \kappa^2 \left(a(\kappa) - \frac{A(\kappa)}{\kappa} \right).$$

From the Sherman-Morrison determinant formula (C. Radhakrishna Rao (1973), page 33), we have

$$Det\left(\text{Var}(\mathbf{S}_n(\boldsymbol{\theta}))\right) = (1 + \mathbf{V}^T\mathbf{B}^{-1}\mathbf{U})Det(\mathbf{B}).$$

The determinant of $\mathbf{B}$ is

$$\begin{aligned} Det(\mathbf{B}) &= (\kappa A(\kappa))^d \left(a(\kappa) - \frac{\kappa^2 a^2(\kappa)}{d(\kappa)} \right), \\ &= (\kappa A(\kappa))^d \frac{A(\kappa)a(\kappa)}{A(\kappa) - \kappa a(\kappa)} \end{aligned}$$

and

$$1 + \mathbf{V}^T\mathbf{B}^{-1}\mathbf{U} = 0,$$

therefore, $\text{Var}(\mathbf{S}_n(\boldsymbol{\theta}))$ is singular and is not a positive definite matrix. □

2.5 Uniform Distribution on the Sphere

We make a brief aside to introduce the uniform distribution.
If in the von-Mises distribution (2.5) κ becomes zero, the density collapses to

$$f_{\mathbf{X}}(\mathbf{x}) = c(0)^{-1}, \quad \mathbf{x} \in \mathbb{S}^d, \tag{2.45}$$

which is called the uniform distribution on the sphere. $c(0)$ is $\frac{2\pi^{d/2}}{\Gamma(d/2)}$ which is calculated through the following process:

$$\begin{aligned}
c(0) = \int_{\mathbf{x}\in\mathbb{S}^d} dw(\mathbf{x}) &= 2\int_0^1 \cdots \int_0^1 \frac{1}{\sqrt{1 - x_1^2 - x_2^2 - \cdots - x_{d-1}^2}} dx_1 dx_2 \cdots dx_{d-1} \\
&= 2\int_0^{\pi} \cdots \int_0^{\pi} \sin^{d-2}\theta_1 \sin^{d-3}\theta_2 \cdots \sin\theta_{d-2} d\theta_1 d\theta_2 \cdots d\theta_{d-1} \\
&= 2\pi \left(\prod_{j=1}^{d-2} \int_0^{\pi/2} \sin^j \theta_{d-j-1} d\theta_{d-j-1} \right) \\
&= 2\pi \left(\prod_{j=1}^{d-2} \frac{\sqrt{\pi}\Gamma(\frac{j+1}{2})}{\Gamma(\frac{j+2}{2})} \right) \\
&= \frac{2\pi^{d/2}}{\Gamma(d/2)}.
\end{aligned}$$

We used (2.13) and (2.15) in lines 2 and 6. We have the obvious equality

$$\begin{aligned}
\mathrm{E}X_i = \frac{1}{c(0)} \int_{\mathbf{x}\in\mathbb{S}^d} x_i dw(\mathbf{x}) &= \frac{2}{c(0)} \int_0^1 \cdots \int_0^1 \frac{x_i}{\sqrt{1 - x_1^2 - x_2^2 - \cdots - x_{d-1}^2}} \\
& dx_1 dx_2 \cdots dx_{d-1} \\
&= \frac{2}{c(0)} \int_0^{\pi} \cdots \int_0^{\pi} \sin^{d-1}\theta_1 \sin^{d-2}\theta_2 \cdots \sin^{d-i}\theta_i \cos\theta_i \sin^{d-i-2}\theta_{i+1} \cdots \\
& \sin\theta_{d-2} d\theta_1 d\theta_2 \cdots d\theta_{d-1} \\
&= \frac{2}{c(0)} \left(\int_0^{\pi} d\theta_{d-1} \right) \left(\int_0^{\pi} \sin\theta_{d-2} d\theta_{d-2} \right) \cdots \left(\int_0^{\pi} \sin^{d-i-2}\theta_{i+1} d\theta_{i+1} \right) \\
& \left(\int_0^{\pi} \sin^{d-i}\theta_i \cos\theta_i d\theta_i \right) \\
& \cdots \left(\int_0^{\pi} \sin^{d-2}\theta_2 d\theta_2 \right) \left(\int_0^{\pi} \sin^{d-1}\theta_1 d\theta_1 \right) \\
&= 0; \quad 1 \le i \le d-1,
\end{aligned} \tag{2.46}$$

on noting that $\int_0^\pi \sin^{d-i}\theta_i \cos\theta_i d\theta_i = 0$. We can apply the same method to see that $\mathrm{E}X_iX_j = 0$ for $i \neq j$. To calculate $\mathrm{E}X_i^2$, we have

$$\begin{aligned}
\mathrm{E}X_i^2 &= \frac{1}{c(0)}\int_{\mathbf{x}\in\mathbb{S}^d} x_i^2 dw(\mathbf{x}) = \frac{2}{c(0)}\int_0^1 \cdots \int_0^1 \frac{x_i^2}{\sqrt{1-x_1^2-x_2^2-\cdots-x_{d-1}^2}} dx_1 dx_2 \cdots dx_{d-1} \\
&= \frac{2}{c(0)}\left(\int_0^\pi d\theta_{d-1}\right)\left(\int_0^\pi \sin\theta_{d-2}d\theta_{d-2}\right)\cdots\left(\int_0^\pi \sin^{d-i-2}\theta_{i+1}d\theta_{i+1}\right) \\
&\left(\int_0^\pi \sin^{d-i-1}\theta_i \cos^2\theta_i d\theta_i\right)\left(\int_0^\pi \sin^{d-i+2}\theta_{i-1}d\theta_2\right)\cdots \\
&\cdots\left(\int_0^\pi \sin^{d-1}\theta_2 d\theta_2\right)\left(\int_0^\pi \sin^d\theta_1 d\theta_1\right) \\
&= \frac{1}{d},
\end{aligned} \tag{2.47}$$

because

$$\begin{aligned}
&\left(\int_0^\pi \sin^{d-i+2}\theta_{i-1}d\theta_{i-1}\right)\left(\int_0^\pi \sin^{d-i-1}\theta_i\cos^2\theta_i d\theta_i\right)\left(\int_0^\pi \sin^{d-i-2}\theta_{i+1}d\theta_{i+1}\right) \\
&= \frac{\sqrt{\pi}\Gamma(\frac{d-i+1}{2}+1)}{\Gamma(\frac{d-i}{2}+2)}\left(\frac{\sqrt{\pi}\Gamma(\frac{d-i}{2})}{\Gamma(\frac{d-i+1}{2})} - \frac{\sqrt{\pi}\Gamma(\frac{d-i}{2}+1)}{\Gamma(\frac{d-i+1}{2}+1)}\right)\frac{\sqrt{\pi}\Gamma(\frac{d-i-1}{2})}{\Gamma(\frac{d-i}{2})} \\
&= \frac{\sqrt{\pi}(\frac{d-i+1}{2})\Gamma(\frac{d-i+1}{2})}{\Gamma(\frac{d-i}{2}+2)}\left(\frac{\sqrt{\pi}\Gamma(\frac{d-i}{2})(\frac{d-i+1}{2}-\frac{d-i}{2})}{(\frac{d-i+1}{2})\Gamma(\frac{d-i+1}{2})}\right)\frac{\sqrt{\pi}\Gamma(\frac{d-i-1}{2})}{\Gamma(\frac{d-i}{2})} \\
&= \frac{\sqrt{\pi}\Gamma(\frac{d-i+1}{2})}{\Gamma(\frac{d-i}{2}+2)}\left(\frac{\sqrt{\pi}\Gamma(\frac{d-i}{2})(\frac{1}{2})}{\Gamma(\frac{d-i+1}{2})}\right)\frac{\sqrt{\pi}\Gamma(\frac{d-i-1}{2})}{\Gamma(\frac{d-i}{2})}.
\end{aligned}$$

Also by symmetry, $\mathrm{E}X_d = 0$ and $\mathrm{E}X_d^2 = \frac{1}{d}$ (since $\mathrm{E}(X_1^2 + \cdots + X_d^2) = 1$). Therefore, in a uniform distribution on a sphere we have

$$\begin{aligned}
\mathrm{E}\mathbf{X} &= 0, \\
\mathrm{Var}(\mathbf{X}) &= \frac{1}{d}\mathbf{I}_d,
\end{aligned} \tag{2.48}$$

where $\mathbf{I}_d$ is a $d \times d$ identity matrix. Now, we can introduce the following theorem:

Theorem 2.4. *In a von Mises Fisher distribution where* $A(\kappa) = \dfrac{c_\kappa}{c(\kappa)}$, *we have (i)* $A(0) = 0$ *and (ii)* $a(0) = \frac{1}{d}$.

Proof of Theorem 2.4. In (2.18) and (2.19) substitute $\kappa = 0$ in the derivatives. Let $\mathrm{E}_{uniformity}(\mathbf{X})$ be the expectation of $\mathbf{X}$ while the density is a uniform distribution on the sphere given in (2.45). From 2.18 which is $A(\kappa) = \mathrm{E}(\boldsymbol{\mu}^T\mathbf{X})$ and putting $\kappa = 0$ in it, we have

$$\mathrm{E}_{uniformity}\left(\boldsymbol{\mu}^T\mathbf{X}\right) = A(0),$$

or equivalently,

$$\boldsymbol{\mu}^T \mathrm{E}_{uniformity}(\mathbf{X}) = A(0).$$

From (2.46) we have $\mathrm{E}_{uniformity}\mathbf{X} = 0$, so the proof of part (i) is done.

For the second part, we apply the same method as used to prove Part (i) in the following form

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{\mu}^T\mathbf{X}\right)^2 &= \frac{c_{\kappa\kappa}}{c(\kappa)} = a(\kappa) + A^2(\kappa) \\
\mathrm{E}_{uniformity}\left(\boldsymbol{\mu}^T\mathbf{X}\right)^2 &= a(0) + A^2(0) \\
\mathrm{E}_{uniformity}\left(\sum_{i=1}^{d}\mu_i^2 X_i^2 + \sum\sum\nolimits_{i\neq j}\mu_i\mu_j X_i X_j\right) &= a(0) + A^2(0) \\
\sum_{i=1}^{d}\mu_i^2 \mathrm{E}_{uniformity}\left(X_i^2\right) + \sum\sum\nolimits_{i\neq j}\mu_i\mu_j \mathrm{E}_{uniformity}\left(X_i X_j\right) &= a(0) + 0 \\
\frac{1}{d} + 0 &= a(0).
\end{aligned}$$

The last equation comes from (2.47) and $\mathrm{E}_{uniformity}(X_iX_j) = 0$. □

2.6 Maximum Likelihood Estimators

We now return to the main development for the von Mises Fisher distribution.
Let $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ be n independent random variables from the von Mises Fisher distribution and $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ observations on them. The log likelihood function is then:

$$\mathcal{L}_n(\kappa, \mu_1, \cdots, \mu_d) = -n\log c(\kappa) + \kappa n \boldsymbol{\mu}^T \overline{\mathbf{x}}_n,$$

where $\overline{\mathbf{x}}_n = (\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n)/n$. Let $\overline{\mathbf{x}}_n = (\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_d)^T$, then we have

$$\mathcal{L}_n(\kappa, \mu_1, \cdots, \mu_{d-1}) = -n\log c(\kappa) + \exp\{n\kappa(\mu_1\overline{x}_1 + \cdots + \sqrt{1 - \mu_1^2 - \cdots - \mu_{d-1}^2}\overline{x}_d)\}.$$

Taking the first derivatives with respect to κ and μ_i gives the following equations:

$$\frac{\partial \mathcal{L}_n}{\partial \kappa} = \frac{-nc_\kappa}{c(\kappa)} + n(\mu_1\overline{x}_1 + \cdots + \sqrt{1 - \mu_1^2 - \cdots - \mu_{d-1}^2}\overline{x}_d) = 0$$

and

$$\frac{\partial \mathcal{L}_n}{\partial \mu_i} = n\kappa\left(\overline{x}_i - \frac{\mu_i}{\mu_d}\overline{x}_d\right) = 0.$$

Solving these equations give the MLEs $\widehat{\mu}_i$ and $\widehat{\kappa}$ as:

$$\widehat{\mu}_i = \frac{\overline{x}_i}{||\overline{\mathbf{x}}_n||} \tag{2.49}$$

and $\widehat{\kappa}$ is to satisfy the following equation:

$$\frac{c_{\widehat{\kappa}}}{c(\widehat{\kappa})} = ||\overline{\mathbf{x}}_n||. \tag{2.50}$$

Chapter 3

A Model for Spherical Data with Lower Dimensional Subcomponents

3.1 Introduction

The aim of this chapter is to introduce a methodology which was motivated by a problem arising when analysing asset allocations in finance, when the issue is to make a portfolio based upon d assets. Because of the nature and structure of a portfolio, it is reasonable to hypothesise a spherical distribution for the square roots of the allocation proportions. But, as the dimension of the data will be determined following the calculation of the portfolio, a combination of spherical distributions with dimensions from 1 to d may be needed.

The von Mises Fisher distribution has special features which make it necessary to apply the strategy and methodology introduced by Maller (2015) to handle this kind of model.

Two real data examples, one containing three and the other ten indices from different countries provide practical case studies for the procedure.

To understand how to build a portfolio with no short sales from d assets and transform the data to a spherical form, we refer the reader to Chapter 2, Example 2.1.

The proportion of 1 total unit invested in Asset i is denoted by $y_i \geq 0$, $1 \leq i \leq d$, $\sum_{i=1}^{d} y_i = 1$. So a portfolio can be represented by the vector of allocations $(y_1, y_2, \ldots, y_d)$, with corresponding observation on the unit sphere $(x_1, x_2, \ldots, x_d) = (\sqrt{y_1}, \sqrt{y_2}, \ldots, \sqrt{y_d})$.

In practice we often observe some of the x_i to be exact zeros (Asset i is not included in the portfolio) in which case the observations fall on a sphere of lower dimensions.

3.2 Setting up the Model

From observations on d assets, $2^d - 1$ possible nontrivial portfolios can be formed. An observed number m_1 of these will contain one asset, $m_1 \in \{0, 1, \ldots, \binom{d}{1}\}$, m_2 of these contain two assets, $m_2 \in \{0, 1, \ldots, \binom{d}{2}\}$. In general, m_i portfolios contain i assets, $m_i \in \{0, 1, \ldots, \binom{d}{i}\}$. Let $m = \sum_{i=2}^{d} m_i$, and $k = m + \sum_{i=2}^{d} i m_i$.

Consider $\mathbb{S}^d$ to be the sample space relevant to d dimensional spherical data, that is, $\mathbb{S}^d = \{\mathbf{x} : \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d, x_1^2 + x_2^2 + \ldots + x_d^2 = 1\}$, and denote

$$\begin{aligned} S_{ij} = \{\mathbf{x} : \mathbf{x} \in \mathbb{S}^d; \quad & d - i \text{ elements of } \mathbf{x} \text{ are equal to zero according to} \\ & \text{reordering no. } j \text{ of zero and nonzero elements}\} \end{aligned} \tag{3.1}$$

to be a subcomponent of $\mathbb{S}^d$, for $j = 1, \ldots, m_i$ and $i = 2, \ldots, d$. Factor i indicates how many assets occur in the nonzero weight in a portfolio. Index j numbers the portfolios within subgroup i, in a certain order. For example, suppose we have 3 assets, $d = 3$. When the portfolio contains only one asset, we have $i = 1$ and three different corner points, so

$$\begin{aligned} S_{11} &= (x_1, 0, 0), \\ S_{12} &= (0, x_2, 0), \\ S_{13} &= (0, 0, x_3), \end{aligned}$$

are the three different possibilities for this occurrence. Here the x_j denote nonzero elements. In this case, in fact, $x_1 = x_2 = x_3 = 1$. As we see $0 \le j \le \binom{3}{1} = 3$. For $i = 2$, we have two assets and $0 \le j \le \binom{3}{2}$, corresponding to

$$\begin{aligned} S_{21} &= (x_1, x_2, 0), \\ S_{22} &= (x_1, 0, x_3), \\ S_{23} &= (0, x_2, x_3). \end{aligned}$$

For $i = 3$ the only possibility is $S_{33} = (x_1, x_2, x_3)$. We treat S_{21}, S_{22} and S_{23} as distinct cases. Later, Theorem 3.8 and Section 3.5 refer to the case when we do not distinguish between them.

Inference will be done conditional on the observed values of the m_i. Portfolios which contain only one asset have no diversification effect and are eliminated from the analysis. So as we see in the following we do not have f_{11} in the structure of the distribution (3.2). Consider the function

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) &= (f_{21}(\mathbf{x}))^{I_{\mathbf{x}}(S_{21})}\,(f_{22}(\mathbf{x}))^{I_{\mathbf{x}}(S_{22})}\ldots(f_{dm_d}(\mathbf{x}))^{I_{\mathbf{x}}(S_{dm_d})} \\ &= \prod_{i=2}^{d}\prod_{j=1}^{m_i}(f_{ij}(\mathbf{x}))^{I_{\mathbf{x}}(S_{ij})}, \end{aligned} \tag{3.2}$$

where

$$I_{\mathbf{x}}(S_{ij}) = \begin{cases} 1 & : \mathbf{x} \in S_{ij} \\ 0 & : \mathbf{x} \notin S_{ij}. \end{cases}$$

For $j = 1, \ldots, m_i$ and $i = 2, \ldots, d$, f_{ij} will be assumed to be the density of an i dimensional von Mises Fisher distribution with parameters $\boldsymbol{\mu}_{ij} \in \mathbb{S}^i$ and $\kappa_{ij} > 0$. Therefore,

$$f_{ij}(\mathbf{x}) = \{c^{(i)}(\kappa_{ij})\}^{-1}\exp\{\kappa_{ij}\boldsymbol{\mu}_{ij}^T\mathbf{x}\}, \quad \mathbf{x} \in S_{ij}, \tag{3.3}$$

where $c^{(i)}(\kappa)$ is the normalizing constant (See (2.5)).

Let n independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be distributed according to the density functions in (3.2). For $k = 1, 2, \ldots, n$ define

$$z_{ijk} = \begin{cases} 1 & : \mathbf{x}_k \in S_{ij} \\ 0 & : \mathbf{x}_k \notin S_{ij}, \end{cases} \tag{3.4}$$

where $\mathbf{x}_k$ is the observed value of $\mathbf{X}_k$. Corresponding to (3.2), form the function

$$L_n(\boldsymbol{\theta}) = \prod_{k=1}^{n}\prod_{i=2}^{d}\prod_{j=1}^{m_i}(f_{ij}(x_k))^{z_{ijk}},$$

and take logs to get

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{k=1}^{n}\{z_{21k}\log(f_{21}(\mathbf{x}_k)) + z_{22k}\log(f_{22}(\mathbf{x}_k)) + \ldots + z_{dm_dk}\log(f_{dm_d}(\mathbf{x}_k))\},$$

where $\boldsymbol{\theta}^T = (\kappa_{21}, \boldsymbol{\mu}_{21}, \kappa_{22}, \boldsymbol{\mu}_{22}, \ldots, \kappa_{dm_d}, \boldsymbol{\mu}_{dm_d})$. The function $\mathcal{L}_n(\boldsymbol{\theta})$ is not a true log likelihood but we will see that as a criterion function it will lead to consistent estimation under some conditions. The function $\mathcal{L}_n(\boldsymbol{\theta})$ can be viewed as a version of a "composite likelihood" as discussed in Section 1.3. As in Vu and Zhou (1997) we call the estimators that are calculated through maximizing the function $\mathcal{L}_n(\boldsymbol{\theta})$ "maximum estimators".

Assume n_{ij} samples are drawn from population (i, j). Therefore we can write $\sum_{k=1}^{n} z_{ijk} = n_{ij}$. Assume that the n_{ij} depend on n such that

$$\lim_{n\to\infty} \frac{n_{ij}}{n} = p_{ij}.$$

From (3.3), we get

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\sum_{j=1}^{m_2} n_{2j} \log c^{(2)}(\kappa_{2j}) - \ldots - \sum_{j=1}^{m_d} n_{dj} \log c^{(d)}(\kappa_{dj}) + \sum_{i=2}^{d}\sum_{j=1}^{m_i} n_{ij}\kappa_{ij}\boldsymbol{\mu}_{ij}^T\overline{\mathbf{X}}_{ij}. \quad (3.5)$$

The first derivative of the estimating function with respect to $\boldsymbol{\theta}$ is a vector whose elements, for $i = 2, \ldots, d$ and $j = 1, \ldots, m_i$, are

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -n_{21}A^{(2)}(\kappa_{21}) + n_{21}\boldsymbol{\mu}_{21}^T\overline{\mathbf{X}}_{21} \\ n_{21}\kappa_{21}\overline{\mathbf{X}}_{21} \\ \vdots \\ -n_{ij}A^{(i)}(\kappa_{ij}) + n_{ij}\boldsymbol{\mu}_{ij}^T\overline{\mathbf{X}}_{ij} \\ n_{ij}\kappa_{ij}\overline{\mathbf{X}}_{ij} \\ \vdots \\ -n_{dm_d}A^{(d)}(\kappa_{dm_d}) + n_{dm_d}\boldsymbol{\mu}_{dm_d}^T\overline{\mathbf{X}}_{dm_d} \\ n_{dm_d}\kappa_{dm_d}\overline{\mathbf{X}}_{dm_d} \end{pmatrix}_{k\times 1}, \quad (3.6)$$

where we write

$$A^{(i)}(\kappa_{ij}) = \frac{c_{\kappa_{ij}}^{(i)}}{c^{(i)}(\kappa_{ij})} \quad (3.7)$$

for the i dimensional case and recall $m = \sum_{i=2}^{d} m_i$. The negative of the second derivative of the estimating function is

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = \begin{pmatrix} \mathbf{F}_{21} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_{dm} \end{pmatrix}_{k\times k}, \quad (3.8)$$

where

$$\mathbf{F}_{ij} = n_{ij}\begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & \mathbf{0} \end{pmatrix}_{(i+1)\times(i+1)}, \quad j = 1, 2, \ldots, m_i, \quad i = 2, 3, \ldots, d.$$

The eigenvalues of $\mathbf{F}_{ij}$ are

$$\begin{cases} 0; \quad (i-1 \text{ times}), \\ \frac{1}{2}\left(a(\kappa_{ij}) + \sqrt{a(\kappa_{ij})^2 + 4||\overline{\mathbf{X}}_{ij}||^2}\right), \\ \frac{1}{2}\left(a(\kappa_{ij}) - \sqrt{a(\kappa_{ij})^2 + 4||\overline{\mathbf{X}}_{ij}||^2}\right), \end{cases}$$

so each $\mathbf{F}_{ij}$ is singular. The number of zero eigenvalues for each $\mathbf{F}_{ij}$ is $i-1$.

From (3.6), (2.18) and (2.21) we see that the expectation of $\mathbf{S}_n(\boldsymbol{\theta})$ is

$$\mathrm{E}\mathbf{S}_n(\boldsymbol{\theta}) = \begin{pmatrix} 0 \\ n_{21}\kappa_{21}A^{(2)}(\kappa_{21})\boldsymbol{\mu}_{21} \\ \vdots \\ 0 \\ n_{ij}\kappa_{ij}A^{(i)}(\kappa_{ij})\boldsymbol{\mu}_{ij} \\ \vdots \\ 0 \\ n_{dm_d}\kappa_{dm_d}A^{(d)}(\kappa_{dm_d})\boldsymbol{\mu}_{dm_d} \end{pmatrix}, \quad j = 1, 2, \ldots, m_i, \quad i = 2, 3, \ldots, d.$$

This is not zero for any $\boldsymbol{\theta} \in \Theta$, and since $\mathbf{F}_n(\boldsymbol{\theta})$ is singular for all $\boldsymbol{\theta} \in \Theta$, we see that "standard" asymptotic theory, as was discussed in (1.5) and the paragraph after that, for maximum estimators does not apply. We use the notation ME for the maximum estimator all throughout.

Consider the vector of restrictions as

$$\mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}_{21}^T\boldsymbol{\mu}_{21} - 1 \\ \vdots \\ \boldsymbol{\mu}_{dm_d}^T\boldsymbol{\mu}_{dm_d} - 1 \end{pmatrix} = \mathbf{0}_{m\times 1}.$$

The derivative of this vector with respect to $\boldsymbol{\theta}$ is

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ 2\boldsymbol{\mu}_{21} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & 2\boldsymbol{\mu}_{ij} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & 2\boldsymbol{\mu}_{dm_d} \end{pmatrix}_{k\times m}. \tag{3.9}$$

Define a vector of Lagrange Multipliers

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_{21} \\ \lambda_{22} \\ \vdots \\ \lambda_{dm_d} \end{pmatrix}_{m\times 1},$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$. Assume there is a constant matrix $\mathbf{C}_n$ and a constant $a > 0$

$$\mathbf{C}_n = a \begin{pmatrix} n_{21} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & n_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & n_{ij} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & n_{dm_d} \end{pmatrix}_{m\times m}, \quad j = 1, 2, \ldots, m_i, \quad i = 2, 3, \ldots, d,$$

such that

$$\boldsymbol{\lambda}_n = \mathbf{C}_n \boldsymbol{\lambda}. \tag{3.10}$$

Here $\boldsymbol{\lambda}_n \in \mathbb{R}^m$ are Lagrange Multipliers which depend on n_{ij} for $j = 1, 2, \ldots, m_i$ and $i = 2, 3, \ldots, d$. We have

$$\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n = a \begin{pmatrix} 0 \\ 2n_{21}\lambda_{21}\boldsymbol{\mu}_{21} \\ 0 \\ 2n_{22}\lambda_{22}\boldsymbol{\mu}_{22} \\ \vdots \\ 0 \\ 2n_{dm_d}\lambda_m\boldsymbol{\mu}_{dm_d} \end{pmatrix}_{k\times 1}.$$

To find expressions for the MEs, prove their consistency, and derive their asymptotic distributions we apply the Maller (2015) methodology outlined in Appendix A. Define

$$\mathbf{S}_n^{\lambda}(\boldsymbol{\theta}) = \mathbf{S}_n(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n = \begin{pmatrix} \mathbf{S}_{21}^{\lambda} \\ \mathbf{S}_{22}^{\lambda} \\ \vdots \\ \mathbf{S}_{dm_d}^{\lambda} \end{pmatrix},$$

where for $a > 0$,

$$\mathbf{S}_{ij}^{\lambda} = n_{ij} \begin{pmatrix} \boldsymbol{\mu}_{ij}^T(\overline{\mathbf{X}}_{ij} - \mathrm{E}\mathbf{X}_{ij}) \\ \kappa_{ij}\overline{\mathbf{X}}_{ij} + 2a\lambda_{ij}\boldsymbol{\mu}_{ij} \end{pmatrix}_{(i+1)\times(i+1)}, \quad j = 1, 2, \ldots, m_i, \quad i = 2, 3, \ldots, d. \tag{3.11}$$

The negative derivative of $\mathbf{S}_n^{\lambda}(\boldsymbol{\theta})$ is

$$\mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) = -\frac{\partial \mathbf{S}_n^{\lambda}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{F}_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n = \begin{pmatrix} \mathbf{F}_{21n}^{\lambda} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22n}^{\lambda} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_{dm_dn}^{\lambda} \end{pmatrix},$$

where

$$\mathbf{F}_{ijn}^{\lambda} = n_{ij} \begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & -2a\lambda_{ij}\mathbf{I}_{i\times i} \end{pmatrix}_{(i+1)\times(i+1)}, \tag{3.12}$$

By the weak law of large numbers, we have

$$\frac{1}{n}\mathbf{F}_{ijn}^{\lambda} \xrightarrow{\mathrm{P}} \mathbf{F}_{ij}^{\lambda}, \quad \text{as } n \to \infty,$$

where

$$\mathbf{F}_{ij}^{\lambda} = p_{ij} \begin{pmatrix} a(\kappa_{ij}) & -\boldsymbol{\mu}_{ij}^T||\mathrm{E}\mathbf{X}_{ij}|| \\ -\boldsymbol{\mu}_{ij}||\mathrm{E}\mathbf{X}_{ij}|| & -2a\lambda_{ij}\mathbf{I}_{i\times i}. \end{pmatrix}_{(i+1)\times(i+1)}. \tag{3.13}$$

Therefore,

$$\frac{1}{n}\mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) \xrightarrow{\mathrm{P}} \mathbf{F}^{\lambda}(\boldsymbol{\theta}),$$

where

$$\mathbf{F}^{\lambda}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{F}_{21}^{\lambda} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22}^{\lambda} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_{dm_d}^{\lambda} \end{pmatrix}$$

and $\mathbf{F}_{ij}^{\lambda}$ is given in (3.13). The eigenvalues of this matrix are, for $i = 2, \ldots, d$ and $j = 1, 2, \ldots, m_i$,

$$\begin{cases} -2a\lambda_{ij}; \quad (i-1 \text{ times}), \\ \frac{1}{2}\left(a(\kappa_{ij}) - 2a\lambda_{ij} + \sqrt{4||\mathrm{E}\mathbf{X}_{ij}||^2 + (a(\kappa_{ij}) + 2a\lambda_{ij})^2}\right), \\ \frac{1}{2}\left(a(\kappa_{ij}) - 2a\lambda_{ij} - \sqrt{4||\mathrm{E}\mathbf{X}_{ij}||^2 + (a(\kappa_{ij}) + 2a\lambda_{ij})^2}\right). \end{cases}$$

The number of eigenvalues which are equal to $-2a\lambda_{ij}$ are $i-1$. The last eigenvalue is always negative for λ_{ij} near λ_{0ij}. Because the function

$$g(\kappa_{ij}) = \kappa_{ij} a(\kappa_{ij}) - A(\kappa_{ij}) = \kappa_{ij} A'(\kappa_{ij}) - A(\kappa_{ij})$$

is a decreasing function of $\kappa > 0$, and from (2.4), $g(0) = A(0) = 0$ (refer to Watson (1983), $g'(\kappa_{ij}) = \kappa_{ij} A''(\kappa_{ij}) < 0$), therefore $\mathbf{F}^{\lambda}(\boldsymbol{\theta})$ is not a positive definite matrix near $\boldsymbol{\lambda}_0$, and, as discussed in Appendix A, we need to introduce a matrix which is to substitute for $\mathbf{F}^{\lambda}(\boldsymbol{\theta})$.

Define for $j = 1, 2, \ldots, m_i$ and $i = 2, 3, \ldots, d$,

$$\mathbf{G}_n = \frac{\sqrt{b}}{2} \begin{pmatrix} \sqrt{n_{21}} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{n_{22}} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{n_{ij}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \sqrt{n_{dm_d}} \end{pmatrix}_{m \times m}, \tag{3.14}$$

where $b > 0$. Then we have

$$\mathbf{F}_n^{\lambda *}(\boldsymbol{\theta}) = \mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\mathbf{G}_n\mathbf{G}_n^T\mathbf{H}^T(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{F}_{21n}^{\lambda *} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22n}^{\lambda *} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_{dm_dn}^{\lambda *} \end{pmatrix},$$

where

$$\mathbf{F}_{ijn}^{\lambda *} = n_{ij} \begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & -2a\lambda_{ij}\mathbf{I}_{i \times i} + b\boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^T \end{pmatrix}_{(i+1)\times(i+1)}.$$

By the weak law of large numbers, we have

$$\frac{1}{n}\mathbf{F}_n^{\lambda*}(\boldsymbol{\theta}) \xrightarrow{\mathrm{P}} \mathbf{F}^{\lambda*}(\boldsymbol{\theta}) := \begin{pmatrix} \mathbf{EF}_{21n}^{\lambda*} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{EF}_{22n}^{\lambda*} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{EF}_{dm_dn}^{\lambda*} \end{pmatrix},$$

as $n \to \infty$, where

$$\mathbf{EF}_{ij}^{\lambda*} = p_{ij} \begin{pmatrix} a(\kappa_{ij}) & -\boldsymbol{\mu}_{ij}^T ||\mathrm{E}\mathbf{X}_{ij}|| \\ -\boldsymbol{\mu}_{ij}||\mathrm{E}\mathbf{X}_{ij}|| & -2a\lambda_{ij}\mathbf{I}_{i\times i} + b\boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^T \end{pmatrix}_{(i+1)\times(i+1)}. \tag{3.15}$$

The eigenvalues of $\mathbf{F}^{\lambda*}(\boldsymbol{\theta})$ are, for $i = 2, \ldots, d$, and $j = 1, 2, \ldots, m_i$,

$$\begin{cases} -2a\lambda_{ij}; \quad (i-1 \text{ times}), \\ \frac{1}{2}\left(a(\kappa_{ij}) - 2a\lambda_{ij} + b + \sqrt{4||\mathrm{E}\mathbf{X}_{ij}||^2 + (a(\kappa_{ij}) - 2a\lambda_{ij} - b)^2}\right), \\ \frac{1}{2}\left(a(\kappa_{ij}) - 2a\lambda_{ij} + b - \sqrt{4||\mathrm{E}\mathbf{X}_{ij}||^2 + (a(\kappa_{ij}) - 2a\lambda_{ij} - b)^2}\right). \end{cases}$$

The number of eigenvalues which are equal to $-2a\lambda_{ij}$ are $i-1$. If we choose a value of b such that

$$b > \frac{||\mathrm{E}\mathbf{X}_{ij}||^2}{a(\kappa_{ij})}, \quad \forall j = 1, 2, \ldots, m_i, \quad \text{and} \quad \forall i = 2, 3, \ldots, d,$$

then $\mathbf{F}^{\lambda*}(\boldsymbol{\theta})$ will be a positive definite matrix near λ_0. Therefore equation (A.13) in Appendix A is satisfied.

3.2.1 Consistency of the Maximum Estimators

We can now apply Theorem A.1 of Appendix A to prove the consistency of the maximum estimators.

Theorem 3.1. *Define $N_n(A)$ and $N_n^h(A)$ to be the neighborhoods*

$$N_n(A) = \{\boldsymbol{\theta} \in \Theta : (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \mathbf{D}_n \mathbf{D}_n^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \leq A^2\} \quad \textit{and} \quad N_n^h(A) = N_n(A) \cap \Theta^h,$$

for each $n = 1, 2, \cdots$ and $A > 0$, and for a given $d \times d$ nonsingular matrix $\mathbf{D}_n$. Then there is an estimator $\widehat{\boldsymbol{\theta}}_n \in \Theta^h$ which, with probability approaching 1 as $n \to \infty$ then $A \to \infty$, satisfies $h(\widehat{\boldsymbol{\theta}}_n) = 0$ and maximizes $\mathcal{L}_n(\boldsymbol{\theta})$ uniquely on $N_n^h(A)$. This estimator $\widehat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}_0$ and does not depend on $\boldsymbol{\lambda}_0$ or on the choice of $\boldsymbol{\lambda}_n$ (in (3.10)) , $\mathbf{G}_n$ (in (3.14)) or $\mathbf{D}_n$.

Proof of Theorem 3.1. Referring to Theorem A.1, it is sufficient to verify assumption (A1). This follows immediately from (3.6) and (3.8). The next step is to define the $d \times d$ nonsingular matrix $\mathbf{D}_n$. In the present situation we can take $\mathbf{D}_n = \sqrt{n}\mathbf{I}_k$ and we see that $\lambda_{min}(\mathbf{D}_n\mathbf{D}_n^T) = n$ which tends to ∞ as $n \to \infty$. Also, from (3.6) and the weak law of large numbers we have

$$\mathbf{D}_n^{-1}\mathbf{S}_n(\boldsymbol{\theta}) \xrightarrow{\text{P}} \mathbf{L} := \begin{pmatrix} 0 \\ p_{21}\kappa_{21}A^{(2)}(\kappa_{21})\boldsymbol{\mu}_{21} \\ \vdots \\ 0 \\ p_{ij}\kappa_{ij}A^{(i)}(\kappa_{ij})\boldsymbol{\mu}_{ij} \\ \vdots \\ 0 \\ p_{dm_d}\kappa_{dm_d}A^{(d)}(\kappa_{dm_d})\boldsymbol{\mu}_{dm_d} \end{pmatrix}_{k\times 1}, \quad \text{as } n \to \infty, \tag{3.16}$$

where $A^{(i)}(\kappa_{ij})$ is defined in (3.7). In order to calculate $\boldsymbol{\lambda}_0$, we solve the system of equations

$$\mathbf{L}_0 + \mathbf{H}(\boldsymbol{\theta}_0)\mathbf{C}\boldsymbol{\lambda} = \mathbf{0},$$

where

$$\mathbf{C} = \lim_{n\to\infty}\frac{\mathbf{C}_n}{n} = a\begin{pmatrix} p_{21} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & p_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & p_{ij} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & p_{dm_d} \end{pmatrix} \tag{3.17}$$

and we have

$$-2a\lambda_{0ij} = \kappa_{ij0}A^{(i)}(\kappa_{ij}) = \kappa_{ij0}||\text{E}_0\mathbf{X}_{ij}||, \quad i = 2, \ldots, d, \quad j = 1, \ldots, m_i. \tag{3.18}$$

Hence

$$\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{S}_{21}^{\lambda_0} \\ \vdots \\ \mathbf{S}_{ij}^{\lambda_0} \\ \vdots \\ \mathbf{S}_{dm_d}^{\lambda_0} \end{pmatrix},$$

where $A^{(i)}(\kappa_{ij})$ is defined in (3.7) and

$$\mathbf{S}_{ij}^{\lambda_0} = n_{ij} \begin{pmatrix} \boldsymbol{\mu}_{0ij}^T(\overline{\mathbf{X}}_{ij} - \mathrm{E}_0\mathbf{X}_{ij}) \\ \kappa_{ij0}(\overline{\mathbf{X}}_{ij} - \mathrm{E}_0\mathbf{X}_{ij}) \end{pmatrix}_{(i+1)\times 1}, \quad i = 2, \ldots, d, \quad j = 1, \ldots, m_i. \tag{3.19}$$

We have

$$\begin{aligned}\mathrm{Var}_0\{\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)\} = \mathrm{E}_0\{\left(\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)\right)\left(\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)\right)^T\} &= n \begin{pmatrix} p_{21}\mathbf{v}_{21}^{\lambda} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & p_{22}\mathbf{v}_{22}^{\lambda} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & p_{dm_d}\mathbf{v}_{dm_dn}^{\lambda} \end{pmatrix} \\ &= n\mathbf{V}(\boldsymbol{\theta}_0) = n\mathbf{V}_0, \end{aligned} \tag{3.20}$$

where

$$\mathbf{v}_{ij}^{\lambda} = \begin{pmatrix} a(\kappa_{0ij}) & \kappa_{0ij}a(\kappa_{0ij})\boldsymbol{\mu}_{0ij}^T \\ \kappa_{0ij}a(\kappa_{0ij})\boldsymbol{\mu}_{0ij} & \kappa_{0ij}^2\mathrm{Var}_0(\mathbf{X}_{ij}) \end{pmatrix}_{(i+1)\times(i+1)}. \tag{3.21}$$

Therefore, since $\mathbf{D}_n = \sqrt{n}\mathbf{I}_k$,

$$\mathrm{Var}_0\{\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)\} = \mathbf{V}_0. \tag{3.22}$$

By considering a unit vector $\mathbf{u}_{d\times 1}$, and the fact that $\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)$ has a finite expectation and variance, we can apply the multidimensional Chebychev's inequality and write

$$P\left(\mathbf{u}^T\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) > \varepsilon\right) < \frac{\mathbf{u}^T\mathbf{V}_0\mathbf{u}}{\varepsilon^2},$$

for any $\varepsilon > 0$. This gives us

$$\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) = O_p(1), \quad \text{as} \quad \mathrm{n} \to \infty$$

and the expression (A.12) in Appendix A is satisfied. The assumption (A.13) of Appendix A comes from (3.15) and the discussions following that. This completes the proof of Theorem 3.1. □

3.2.2 Consistency of $\widehat{\boldsymbol{\lambda}}_n$

From equation (3.11) and $\mathbf{S}_n^{\lambda}(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$ we have

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_{ij}^T\left(\overline{\mathbf{X}}_{ij} - \mathrm{E}\mathbf{X}_{ij}\right) &= 0, \\ \widehat{\kappa}_{ij}\overline{\mathbf{X}}_{ij} + 2a\widehat{\lambda}_{ij}\boldsymbol{\mu}_{ij} &= \mathbf{0}, \end{aligned}$$

for $j = 1, 2, \ldots, m_i$ and $i = 2, 3, \ldots, d$. From the second equations in the above pair we get

$$\widehat{\kappa}\overline{\mathbf{X}}_{ij} = -2a\widehat{\lambda}_{ij}\boldsymbol{\mu}_{ij}, \tag{3.23}$$
$$\widehat{\kappa}||\overline{\mathbf{X}}_{ij}|| = -2a\widehat{\lambda}_{ij} \tag{3.24}$$

and

$$\widehat{\kappa}\overline{\mathbf{X}}_{ij} + 2a\widehat{\lambda}_{ij}\boldsymbol{\mu}_{ij} = 0,$$
$$\widehat{\boldsymbol{\mu}}_{ij} = \frac{\widehat{\kappa}\overline{\mathbf{X}}_{ij}}{-2a\widehat{\lambda}_{ij}}.$$

Therefore,

$$\widehat{\boldsymbol{\mu}}_{ij} = \frac{\overline{\mathbf{X}}_{ij}}{||\overline{\mathbf{X}}_{ij}||}.$$

Also

$$\widehat{\boldsymbol{\mu}}_{ij}^T \left(\overline{\mathbf{X}}_{ij} - \mathrm{E}\mathbf{X}_{ij}\right) = 0, \tag{3.25}$$
$$\widehat{\boldsymbol{\mu}}_{ij}^T \left(\overline{\mathbf{X}}_{ij} - A^{(i)}(\widehat{\kappa}_{ij})\boldsymbol{\mu}_{ij}\right) = 0$$

and

$$A^{(i)}(\widehat{\kappa}_{ij}) = ||\overline{\mathbf{X}}_{ij}||.$$

Theorem 3.2. *There exists a consistent estimator $\widehat{\boldsymbol{\lambda}}_n$, obtained from equation (3.11), for the true value of $\boldsymbol{\lambda}_0$, verified by (3.18).*

Proof of Theorem 3.2. To prove this theorem, it is sufficient to follow Theorem A.2 in Appendix A. As we see from (3.6) and (3.8), assumption (A1) holds. We have $\boldsymbol{\lambda}_n$ in the form of (3.10). There is a consistent estimator $\widehat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ satisfying $\mathbf{h}(\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}$ from Theorem 3.1. We calculated a unique $\widehat{\boldsymbol{\lambda}}_n$ in (3.23). We set $a_n = \frac{1}{n}$ and calculated (3.17), so we have (A.17). The existence of $\mathbf{L}_0$ can be verified by considering (3.16) which gave the true values of $\boldsymbol{\lambda}_0$ in (3.18).
It only remains to check the assumption

$$\mathbf{D}_n^{-1}\left(\mathbf{S}_n^{\widehat{\lambda}}(\widehat{\boldsymbol{\theta}}_n) - \mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)\right) \xrightarrow{\mathrm{P}} 0, \quad \text{as} \quad \mathrm{n} \to \infty.$$

If we apply the weak law of large numbers in (3.19) and notice that $\mathbf{S}_n^{\widehat{\lambda}}(\widehat{\boldsymbol{\theta}}_n) = 0$, then the proof is done. □

3.2.3 Asymptotic Distribution of the MEs

Define

$$\mathbf{U}_n^{\lambda}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{F}_n^{\lambda*}(\boldsymbol{\theta}) & -\mathbf{H}(\boldsymbol{\theta})\mathbf{C}_n \\ -\mathbf{C}_n^T\mathbf{H}^T(\boldsymbol{\theta}) & \mathbf{0} \end{pmatrix}_{(k+m)\times(k+m)}$$

and

$$\mathbf{J}_n = \sqrt{n}\mathbf{I}_{(k+m)}.$$

As $n \to \infty$, assume

$$\mathbf{J}_n^{-1}\mathbf{U}_n^{\lambda_0}(\boldsymbol{\theta}_0)\mathbf{J}_n^{-T} \xrightarrow{\text{P}} \mathbf{U}_0, \tag{3.26}$$

where we denote

$$\mathbf{U}_0^{-1} = \begin{pmatrix} \mathbf{P}_{k\times k} & \mathbf{Q}_{k\times m} \\ \mathbf{Q}_{m\times k}^T & \mathbf{R}_{m\times m} \end{pmatrix}. \tag{3.27}$$

Lemma 3.3. *We have*

$$\mathbf{D}_n^{-1}\mathbf{F}_n^{\lambda_0^*}(\boldsymbol{\theta}_0)\mathbf{D}_n^{-1} \xrightarrow{\text{P}} \mathbf{F}_0^*, \quad \text{as } n \to \infty,$$

where $\mathbf{F}_0^*$ *exists under* (3.26) *and is nonsingular. Further, referring to* (1.15), (1.16) *and* (1.17), *we have*

$$\begin{aligned}
\mathbf{R} &= -\left(\mathbf{H}(\boldsymbol{\theta})^T(\mathbf{F}_0^*)^{-1}\mathbf{H}(\boldsymbol{\theta})\right)^{-1}, \\
\mathbf{Q} &= (\mathbf{F}_0^*)^{-1}\mathbf{H}(\boldsymbol{\theta})\left(\mathbf{H}^T(\boldsymbol{\theta})(\mathbf{F}_0^*)^{-1}\mathbf{H}(\boldsymbol{\theta})\right)^{-1}, \\
\mathbf{P} &= (\mathbf{F}_0^*)^{-1} - (\mathbf{F}_0^*)^{-1}\mathbf{H}(\boldsymbol{\theta})\left(\mathbf{H}^T(\boldsymbol{\theta})(\mathbf{F}_0^*)^{-1}\mathbf{H}(\boldsymbol{\theta})\right)^{-1}\mathbf{H}^T(\boldsymbol{\theta})(\mathbf{F}_0^*)^{-1}.
\end{aligned}$$

Applying Theorem A.3 in Appendix A we have

$$\mathbf{J}_n\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\text{D}} \mathbf{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{P}\mathbf{V}_0\mathbf{P} & \mathbf{P}\mathbf{V}_0\mathbf{Q} \\ \mathbf{Q}^T\mathbf{V}_0\mathbf{P} & \mathbf{Q}^T\mathbf{V}_0\mathbf{Q} \end{pmatrix}\right)$$

Theorem 3.4. *If* $\mathbf{F}_0^* = \mathbf{V}_0$*, where* $\mathbf{V}_0$ *is introduced in* (3.22)*, then*

$$\begin{aligned}\mathbf{P}\mathbf{V}_0\mathbf{P} &= \mathbf{P},\\ \mathbf{P}\mathbf{V}_0\mathbf{Q} &= \mathbf{0},\\ \mathbf{Q}^T\mathbf{V}_0\mathbf{P} &= \mathbf{0},\\ \mathbf{Q}^T\mathbf{V}_0\mathbf{Q} &= -\mathbf{R}.\end{aligned}$$

Proof of Theorem 3.4. This proof comes from (1.20) in Section 1.2.4. □

Assume $\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) \xrightarrow{\text{D}} \mathbf{Z}$ for an a.s. finite random vector $\mathbf{Z} \in \mathbb{R}^k$ as $n \to \infty$. We have $\mathbf{Z} \sim \mathbf{N}(\mathbf{0}, \mathbf{V}_0)$, where $\mathbf{V}_0$ is $\text{Var}_0(\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0))$ and is given in (3.22). Therefore,

$$\mathbf{N} = \mathbf{P}^{T/2}\mathbf{Z} \sim \mathbf{N}(\mathbf{0}, \mathbf{P}^{T/2}\mathbf{V}_0\mathbf{P}^{1/2}), \tag{3.28}$$

where $\mathbf{P}$ is calculated through (3.27).

Theorem 3.5. *Assume* $\mathbf{P}\mathbf{V}_0\mathbf{P} = \mathbf{P}$*. If* $\mathbf{P}_{k\times k}$ *is invertible, we have*

$$\mathbf{P}^{T/2}\mathbf{V}_0\mathbf{P}^{1/2} = \mathbf{I}.$$

If $\mathbf{P}$ *is not invertible with rank* ℓ*,* $\ell < k$*, we have*

$$\mathbf{P}^{T/2}\mathbf{V}_0\mathbf{P}^{1/2} = \begin{pmatrix}\mathbf{I}_\ell & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.$$

Let $\mathbf{N} = (N_1, N_2, \ldots, N_k)^T$ *be the vector in* (3.28)*. It follows that* $N_1, N_2, \ldots, N_\ell$ *are independent* $N(0,1)$ *random variables, while* $N_{\ell+1} = \ldots = N_k = 0$ *(Silvey (1959), page 403).*

Proof of Theorem 3.5. We can write

$$\mathbf{P}^{T/2}\mathbf{V}_0\mathbf{P}^{1/2} = \mathbf{P}^{T/2}\mathbf{P}^{-1}\mathbf{P}\mathbf{V}_0\mathbf{P}\mathbf{P}^{-1}\mathbf{P}^{1/2}.$$

When $\mathbf{P}$ is invertible and $\mathbf{P}\mathbf{V}_0\mathbf{P} = \mathbf{P}$, the following result is obtained:

$$\begin{aligned}\mathbf{P}^{T/2}\mathbf{P}^{-1}\mathbf{P}\mathbf{V}_0\mathbf{P}\mathbf{P}^{-1}\mathbf{P}^{1/2} = \mathbf{P}^{T/2}\mathbf{P}^{-1}\mathbf{P}^{1/2} &= \mathbf{P}^{T/2}(\mathbf{P}^{1/2}\mathbf{P}^{-T/2})^{-1}\mathbf{P}^{1/2}\\ &= \mathbf{P}^{T/2}\mathbf{P}^{-T/2}\mathbf{P}^{-1/2}\mathbf{P}^{1/2}\\ &= \mathbf{I}.\end{aligned}$$

But the matrix $\mathbf{P}$ in our setup is not positive definite; its determinant is zero and consequently it is a singular matrix. In fact it is a non-negative definite symmetric matrix. In this case, we proceed as follows. Because $\mathbf{P}_{k\times k}$ is a symmetric and non-negative definite matrix, then from Schur's Decomposition theorem (C. Radhakrishna Rao (1973), pages 42 to 45) there exists an orthogonal matrix $\mathbf{U}$ (that is $\mathbf{U}\mathbf{U}^T = \mathbf{I}$) whose columns are the eigenvectors of $\mathbf{P}$ and a diagonal matrix $\mathbf{S}$ whose diagonal elements are the eigenvalues of $\mathbf{P}$, denoted $s_1, s_2, \ldots, s_k$ with $s_1 \geq s_2 \geq \ldots \geq s_k \geq 0$, such that $\mathbf{P} = \mathbf{U}\mathbf{S}\mathbf{U}^T$. The rank of $\mathbf{P}$ is determined by the number of non-zero values of s_j. The square root of this matrix can be written as

$$\mathbf{P}^{1/2} = \mathbf{U}\mathbf{S}^{1/2}\mathbf{U}^T,$$

where

$$\mathbf{S}^{1/2} = Diag(s_1^{1/2}, s_2^{1/2}, \ldots, s_k^{1/2}).$$

We define the pseudo-inverse of $\mathbf{P}$ by

$$\mathbf{P}^+ = \mathbf{U}\mathbf{R}\mathbf{U}^T,$$

where $\mathbf{R}$ is a diagonal matrix with the elements

$$r_j = \begin{cases} \frac{1}{s_j} & ; s_j \neq 0 \\ 0 & ; s_j = 0. \end{cases}$$

Therefore, we can write

$$(\mathbf{P}^+)^{1/2} = \mathbf{U}\mathbf{R}^{1/2}\mathbf{U}^T.$$

Applying this result to simplify $\mathbf{P}^{T/2}\mathbf{P}^+\mathbf{P}^{1/2}$ gives

$$\begin{aligned} \mathbf{P}^{T/2}\mathbf{P}^+\mathbf{P}^{1/2} &= (\mathbf{U}\mathbf{S}^{1/2}\mathbf{U}^T)^T(\mathbf{U}\mathbf{R}\mathbf{U}^T)(\mathbf{U}\mathbf{S}^{1/2}\mathbf{U}^T) \\ &= (\mathbf{U}\mathbf{S}^{T/2}\mathbf{U}^T)(\mathbf{U}\mathbf{R}\mathbf{U}^T)(\mathbf{U}\mathbf{S}^{1/2}\mathbf{U}^T) \\ &= \mathbf{U}\mathbf{S}^{T/2}\mathbf{R}\mathbf{S}^{1/2}\mathbf{U}^T \\ &= \begin{pmatrix} \mathbf{I}_\ell & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \end{aligned}$$

where ℓ is the rank of $\mathbf{P}$. □

Theorem 3.6. *For a d dimensional von Mises Fisher distribution with the parameters κ and $\boldsymbol{\mu}$ we have $\mathbf{P}\mathbf{V}_0\mathbf{P} = \mathbf{P}$, and the dimension of random variable $\mathbf{Z}$ in (3.28) is equal to the dimension of $\mathbf{P}$.*

Proof of Theorem 3.6. The proof is achieved through noticing that $\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^T$ is idempotent and $\boldsymbol{\mu}^T(\mathbf{I}-\boldsymbol{\mu}\boldsymbol{\mu}^T)=0$ for a von Mises Fisher distribution. □

3.3 Hypothesis Testing

Typically, there are two different kinds of hypothesis tests about the concentration parameters in a model whose subcomponents are von Mises Fisher distributions of different dimensions which might be interesting. We introduce them in this section and derive the corresponding asymptotic distributions. The proof of Theorem 3.7 is in Section 3.3.1 and that of Theorem 3.8 is in Section 3.3.2.

Theorem 3.7. *We wish to test whether the κ's for a certain dimension in the model (3.2) are equal or not. This test can be formulated as*

$$\begin{cases} H_0 & : \kappa_{\ell 1} = \kappa_{\ell 2} = \ldots = \kappa_{\ell m_\ell} \\ H_A & : \textit{at least one equality is not satisfied,} \end{cases} \tag{3.29}$$

where ℓ is the dimension of interest, $2 \leq \ell \leq d$. Then the asymptotic distribution of d_n for testing this hypothesis is chi square with $m_\ell - 1$ degrees of freedom.

Next, consider

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=2}^{d} (f_i(\mathbf{x}))^{I_{\mathbf{x}}(S_i)} \tag{3.30}$$

where for $i = 2, \ldots, d$,

$$S_i = \{\mathbf{x} : \mathbf{x} \in \mathbb{S}^d \quad \text{with } d-i \text{ elements zero}\},$$

is a special case of S_{ij} in (3.1) in which the order of zeros is not important and

$$I_{\mathbf{x}}(S_i) = \begin{cases} 1 & ; \mathbf{x} \in S_i \\ 0 & ; \mathbf{x} \notin S_i. \end{cases}$$

The function $f_i(\mathbf{x})$ for $i = 2, \ldots, d$ is an i dimensional von Mises Fisher distribution with the density

$$f_i(\mathbf{x}) = \{c^{(i)}(\kappa^{(i)})\}^{-1} \exp\{\kappa^{(i)} \boldsymbol{\mu}_i^T \mathbf{x}\}, \quad \mathbf{x} \in \mathbb{S}^i, ||\boldsymbol{\mu}_i|| = 1, \kappa^{(i)} > 0.$$

We are going to test the hypothesis

$$\begin{cases} H_0 & : \kappa^{(2)} = \kappa^{(3)} = \ldots = \kappa^{(d)} \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{3.31}$$

where $\kappa^{(i)}$ is the concentration parameter, $c^{(i)}(\kappa^{(i)})$ is the normalising constant, and $\boldsymbol{\mu}_i$ is the modal vector of an i dimensional von Mises Fisher distribution.

Theorem 3.8. *The asymptotic distribution of the deviance statistic for the hypothesis test* (3.31) *is chi square with* $d-2$ *degrees of freedom.*

3.3.1 Proof of Theorem 3.7

We follow the formulation of Neuenshawander and Flurry (1997) for doing hypothesis tests based on Lagrange Multiplier tests. Section 1.2.5 of Chapter 1 explains this method. Based on this method to do the hypothesis test (3.29), we need to add the constraints

$$\mathbf{h}_2(\boldsymbol{\theta}) = \begin{pmatrix} \kappa_{l1} - \kappa_{lm_\ell} \\ \kappa_{l2} - \kappa_{lm_\ell} \\ \vdots \\ \kappa_{lm_{\ell-1}} - \kappa_{lm_\ell} \end{pmatrix} = \mathbf{0}_{m_{\ell-1}\times 1},$$

to the process of making inference. Therefore, $\mathbf{S}_n^{\lambda}(\boldsymbol{\theta})$ becomes

$$\mathbf{S}_n^{\lambda}(\boldsymbol{\theta}) = \mathbf{S}_n(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n},$$

where $\mathbf{H}(\boldsymbol{\theta})$ is defined in (3.9) and

$$\mathbf{H}_2(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{0}_{i_1\times 1} & \mathbf{0}_{i_1\times 1} & \mathbf{0}_{i_1\times 1} & \cdots & \mathbf{0}_{i_1\times 1} & \mathbf{0}_{i_1\times 1} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \cdots & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \cdots & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \cdots & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} \\ -1 & -1 & -1 & \cdots & -1 & -1 \\ \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \cdots & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} \\ \mathbf{0}_{i_2\times 1} & \mathbf{0}_{i_2\times 1} & \mathbf{0}_{i_2\times 1} & \cdots & \mathbf{0}_{i_2\times 1} & \mathbf{0}_{i_2\times 1} \end{pmatrix}_{k\times m_{\ell-1}}. \tag{3.32}$$

Here $i_1 = (\sum_{i=2}^{l-1} m_i) + \sum_{i=2}^{l-1} im_i$ and $i_2 = (m - \sum_{i=2}^{l} m_i) + \sum_{i=l+1}^{d} im_i$. Also

$$\boldsymbol{\lambda}_2 = \begin{pmatrix} \lambda_{m+1} \\ \lambda_{m+2} \\ \vdots \\ \lambda_{m+m_{\ell-1}} \end{pmatrix}_{m_{\ell-1}\times 1}, \quad \mathbf{C}_{2n} = a \begin{pmatrix} n_{l1} & 0 & \dots & 0 \\ 0 & n_{l2} & \dots & 0 \\ 0 & 0 & \dots & n_{lm_{\ell-1}} \end{pmatrix}_{m_{\ell-1}\times m_{\ell-1}}$$

and set $\boldsymbol{\lambda}_{2n} = \mathbf{C}_{2n}\boldsymbol{\lambda}_2$ which is in $\mathbb{R}^{m_{\ell-1}}$ and depends on $n_{\ell j}$, $j = 1, 2, \dots, m_l$. We have

$$\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n} = a \begin{pmatrix} H\lambda_{21} \\ H\lambda_{22} \\ \vdots \\ n_{\ell 1}\lambda_{m+1} \\ 2n_{\ell 1}\lambda_{\ell 1}\boldsymbol{\mu}_{\ell 1} \\ n_{\ell 2}\lambda_{m+2} \\ 2n_{\ell 2}\lambda_{\ell 2}\boldsymbol{\mu}_{\ell 2} \\ \vdots \\ n_{\ell m_{\ell-1}}\lambda_{m+m_{\ell-1}} \\ 2n_{\ell m_{\ell-1}}\lambda_{\ell m_{\ell-1}}\boldsymbol{\mu}_{\ell m_{\ell-1}} \\ -\sum_{i=1}^{m_{\ell-1}} n_{\ell i}\lambda_{m+i} \\ 2n_{\ell m_\ell}\lambda_{\ell m_\ell}\boldsymbol{\mu}_{\ell m_\ell} \\ H\lambda_{(\ell+1)1} \\ \vdots \\ H\lambda_{dm_d} \end{pmatrix}_{k\times 1}$$

where for $i = 2, \dots, d$, $i \neq \ell$, and $j = 1, \dots, m_i$,

$$H\lambda_{ij} = \begin{pmatrix} 0 \\ 2n_{ij}\lambda_{ij}\boldsymbol{\mu}_{ij} \end{pmatrix}_{(i+1)\times 1}.$$

Therefore

$$\mathbf{S}_n^{\lambda}(\boldsymbol{\theta}) = \mathbf{S}_n(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n} = \begin{pmatrix} \mathbf{S}_{21}^{\lambda} \\ \vdots \\ \mathbf{S}_{dm_d}^{\lambda} \end{pmatrix}.$$

For $i \neq \ell$, $\mathbf{S}_{ij}^{\lambda}$ is as in (3.11). For $i = l$ and $j = 1, \ldots, m_{\ell-1}$, we have

$$\mathbf{S}_{\ell j}^{\lambda} = n_{\ell j} \begin{pmatrix} \boldsymbol{\mu}_{\ell j}(\overline{\mathbf{X}}_{\ell j} - \mathrm{E}\mathbf{X}_{\ell j}) + a\lambda_{m+j} \\ \kappa_{\ell j}\overline{\mathbf{X}}_{\ell j} + 2a\lambda_{\ell j}\boldsymbol{\mu}_{\ell j} \end{pmatrix}. \tag{3.33}$$

For $i = \ell$, and $j = m_\ell$,

$$\mathbf{S}_{\ell m_\ell}^{\lambda} = n_{\ell m_\ell} \begin{pmatrix} \boldsymbol{\mu}_{\ell m_\ell}(\overline{\mathbf{X}}_{\ell m_\ell} - \mathrm{E}\mathbf{X}_{\ell m_\ell}) - a\frac{\sum_{i=1}^{m_{\ell-1}} n_{\ell i}\lambda_{m+i}}{n_{\ell m_\ell}} \\ \kappa_{\ell m_\ell}\overline{\mathbf{X}}_{\ell m_\ell} + 2a\lambda_{\ell m_\ell}\boldsymbol{\mu}_{\ell m_\ell} \end{pmatrix}_{(\ell+1)\times(\ell+1)}. \tag{3.34}$$

The negative derivative of $\mathbf{S}_n^{\lambda}(\boldsymbol{\theta})$, where $\mathbf{H}(\boldsymbol{\theta})$ and $\mathbf{H}_2(\boldsymbol{\theta})$ are respectively defined in (3.9) and (3.32), is

$$\begin{aligned} \mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) &= -\frac{\partial \mathbf{S}_n^{\lambda}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{F}_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}}\left(\mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n}\right) \\ &= \begin{pmatrix} \mathbf{F}_{21n}^{\lambda} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22n}^{\lambda} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{F}_{dm_d n}^{\lambda} \end{pmatrix}_{k\times k}, \end{aligned}$$

where

$$\mathbf{F}_{ijn}^{\lambda} = n_{ij} \begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & -2a\lambda_{ij}\mathbf{I}_{i\times i} \end{pmatrix}_{(i+1)\times(i+1)},$$

is the same as the matrix derived in (3.12), so the theory following this formula remains valid here including the formula (3.15).

To calculate $\boldsymbol{\lambda}_0$ from the equations (3.16) and

$$\mathbf{L}_0 + \mathbf{H}(\boldsymbol{\theta}_0)\mathbf{C}\boldsymbol{\lambda} + \mathbf{H}_2(\boldsymbol{\theta}_0)\mathbf{C}_2\boldsymbol{\lambda}_2 = 0,$$

where

$$\mathbf{C}_2 = \lim_{n\to\infty} \frac{\mathbf{C}_{2n}}{n} = a \begin{pmatrix} p_{l1} & 0 & \cdots & 0 \\ 0 & p_{l2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{lm_{\ell-1}} \end{pmatrix}_{m_{\ell-1}\times m_{\ell-1}},$$

we have

$$-2a\lambda_{ij} = \kappa_{0ij}A(\kappa_{0ij}) = \kappa_{0ij}||\mathrm{E}_0\mathbf{X}_{ij}||; \quad i = 2, \ldots, d \quad j = 1, \ldots, m_i$$

and

$$\lambda_{m+i} = 0; \quad i = 1, \ldots, m_{\ell-1}.$$

Hence, the formulae (3.19), (3.20) and (3.21) remain valid here.

We go through the formulae (3.11), (3.33) and (3.34) to calculate the MEs. From the second parts of these formulae which are all in the form of (3.23), we have

$$\widehat{\boldsymbol{\mu}}_{ij} = \frac{\overline{\mathbf{X}}_{ij}}{||\overline{\mathbf{X}}_{ij}||}; \quad i = 2, \ldots, d, \quad j = 1, \ldots, m_i.$$

We can use the same method as in (3.25) to verify

$$A^{(i)}(\kappa_{ij}) = ||\overline{\mathbf{X}}_{ij}||; \quad i = 2, \ldots, d \text{ and } i \neq \ell, j = 1, \ldots, m_i.$$

This formula for the ME of κ_{ij} is only valid when $i \neq \ell$. For $i = \ell$, we use formulae (3.33) and (3.34) and write

$$\boldsymbol{\mu}_{\ell j}(\overline{\mathbf{X}}_{\ell j} - A^{(\ell)}(\kappa_{\ell j})\boldsymbol{\mu}_{\ell j}) + a\lambda_{m+j} = 0; \quad j = 1, \ldots, m_{\ell-1},$$

which gives

$$a\widehat{\lambda}_{m+j} = A^{(\ell)}(\widehat{\kappa}_{\ell j}) - ||\overline{\mathbf{X}}_{\ell j}||; \quad j = 1, \ldots, m_{\ell-1}. \tag{3.35}$$

From (3.34), we have

$$-a\frac{\sum_{j=1}^{m_{\ell-1}} n_{\ell j}\widehat{\lambda}_{m+j}}{n_{\ell m_\ell}} = A^{(\ell)}(\widehat{\kappa}_{\ell m_\ell}) - ||\overline{\mathbf{X}}_{\ell m_\ell}||. \tag{3.36}$$

If we multiply both sides of (3.35) by $n_{\ell j}$ then sum from $j = 1$ to $m_{\ell-1}$, we have

$$a\sum_{j=1}^{m_{\ell-1}} n_{\ell j}\widehat{\lambda}_{m+j} = \sum_{j=1}^{m_{\ell-1}} n_{\ell j}A^{(\ell)}(\kappa) - \sum_{j=1}^{m_{\ell-1}} n_{\ell j}||\overline{\mathbf{X}}_{\ell j}||. \tag{3.37}$$

Combining (3.36) and (3.37) and setting $\kappa_{\ell 1} = \ldots = \kappa_{\ell m_\ell} = \kappa$ gives

$$A^{(\ell)}(\widehat{\kappa}) = \frac{\sum_{j=1}^{m_\ell} n_{\ell j}||\overline{\mathbf{X}}_{\ell j}||}{\sum_{j=1}^{m_\ell} n_{\ell j}}.$$

Define

$$\mathbf{U}_n^{\lambda}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{F}_n^{\lambda *}(\boldsymbol{\theta}) & -\mathbf{H}(\boldsymbol{\theta})\mathbf{C}_n & -\mathbf{H}_2(\boldsymbol{\theta})\mathbf{C}_{2n} \\ -\mathbf{C}_n^T\mathbf{H}^T(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ -\mathbf{C}_{2n}^T\mathbf{H}_2^T(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \end{pmatrix}_{(k+m+m_{\ell-1})\times(k+m+m_{\ell-1})}$$

and

$$\mathbf{J}_n = \sqrt{n}\mathbf{I}_{(k+m+m_{\ell-1})}.$$

As $n \to \infty$, assume

$$\mathbf{J}_n^{-1}\mathbf{U}_n^{\lambda_0}(\boldsymbol{\theta}_0)\mathbf{J}_n^{-T} \xrightarrow{\text{P}} \mathbf{U}_0.$$

Let

$$\mathbf{U}_0^{-1} = \begin{pmatrix} \mathbf{P}_{k\times k} & \mathbf{Q}_{k\times(m+m_{\ell-1})} \\ \mathbf{Q}^T_{(m+m_{\ell-1})\times k} & \mathbf{R}_{(m+m_{\ell-1})\times(m+m_{\ell-1})} \end{pmatrix}. \tag{3.38}$$

To test the hypothesis (3.29), we refer to Theorem A.4 in Appendix A which gives

$$d_n \xrightarrow{\text{D}} \inf_{\boldsymbol{\theta}\in\dot{C}_{\Omega}} ||\mathbf{N}-\boldsymbol{\theta}||^2 - \inf_{\boldsymbol{\theta}\in\dot{C}_{\Omega\cup\tau}} ||\mathbf{N}-\boldsymbol{\theta}||^2, \quad \text{as } n \to \infty,$$

where Ω and τ are respectively the sets defined by the null and alternative hypotheses in (3.29), and the random variable $\mathbf{N}$ is defined in (3.28).

The point $\mathbf{N}$ is in $(\sum_{i=1}^{d} im_i)$-dimensional cartesian coordinates. We project this point into the plane constructed by $\kappa_{\ell 1}, \ldots, \kappa_{\ell m_\ell}$ in $\mathbb{R}^{m_\ell}$. Therefore, this projection has dimension m_ℓ.

The cones $\dot{C}_\Omega$ and $\dot{C}_{\Omega\cup\tau}$ are defined as follows. Here, $\dot{C}_{\Omega\cup\tau} = \mathbb{R}^{m_\ell}$ and $\inf_{\boldsymbol{\theta}\in\dot{C}_{\Omega\cup\tau}} ||\mathbf{N} - \boldsymbol{\theta}||^2 = 0$. Thus

$$\Omega = \{(\kappa_{\ell 1}, \kappa_{\ell 2}, \ldots, \kappa_{\ell m_\ell}) : \kappa_{\ell 1} = \kappa_{\ell 2} = \ldots = \kappa_{\ell m_\ell} > 0\}$$

and

$$\widetilde{C}_\Omega = \{(\kappa_{\ell 1}, \kappa_{\ell 2}, \ldots, \kappa_{\ell m_\ell}) \in \mathbb{R}^{m_\ell} : \kappa_{\ell 1} = \kappa_{\ell 2} = \ldots = \kappa_{\ell m_\ell}\}, \tag{3.39}$$

which is a line passing through the origin with an equal parametric factor. For $\dot{C}_\Omega$ we have

$$\dot{C}_\Omega = \{\mathbf{P}^{1/2}\mathbf{F}_0^*\boldsymbol{\theta} : \boldsymbol{\theta} \in \widetilde{C}_\Omega\}.$$

Because in $\mathbf{P}^{1/2}$ and $\mathbf{F}_0^*$ all the elements in the plane $\kappa_{\ell 1} \dots \kappa_{\ell m_\ell}$ are equal, the equation of the line presented in (3.39) remains the same and we have

$$\dot{C}_\Omega = \{(\kappa_{\ell 1}, \kappa_{\ell 2}, \dots, \kappa_{\ell m_\ell}) \in \mathbb{R}^{m_\ell} : \kappa_{\ell 1} = \kappa_{\ell 2} = \dots = \kappa_{\ell m_\ell}\}.$$

Therefore to calculate the distance of the point $\mathbf{N}$ in an m_ℓ-dimensional space from the cone $\dot{C}_\Omega$, it is enough to find the Euclidean distance of $\mathbf{N}$ from the line $\kappa_1 = \kappa_2 = \dots = \kappa_{m_\ell}$. From the formula (3.40), this distance is

$$distance^2 = \sum_{i=1}^{m_\ell} N_i^2 - \frac{(\sum_{i=1}^{m_\ell}(-N_i))^2}{m_\ell},$$

which has a chi square distribution with $m_\ell - 1$ degrees of freedom. This follows from the next proposition by setting $\kappa_1 = \kappa_2 = \dots = \kappa_{m_\ell}$, so that $a_i = 1$ and $b_i = 0$ for all $i = 1, 2, \dots, m_\ell$.

Theorem 3.9. *Assume line L has the parametric equation*

$$\begin{cases} \frac{x_1 - b_1}{a_1} & = t \\ \frac{x_2 - b_2}{a_2} & = t \\ & \vdots \\ \frac{x_k - b_k}{a_k} & = t; \end{cases}$$

thus, as t varies over $\mathbb{R}$, we obtain points on the line L. Now, we want to find a value of t which minimises the distance of the point $\mathbf{N} = (N_1, \dots, N_k)$ from this line. Assume a point $\mathbf{M} = (a_1 t + b_1, \dots, a_k t + b_k) \in L$. Then, the Euclidean distance of $\mathbf{M}$ from $\mathbf{N}$ is

$$distance^2 = \sum_{i=1}^{k}(N_i - a_i t - b_i)^2.$$

The derivative of this distance with respect to t is

$$2distance\frac{\partial distance}{\partial t} = -2\sum_{i=1}^{k} a_i(N_i - a_i t - b_i),$$

so the minimum occurs at

$$t = t_{min} = \frac{\sum_{i=1}^{k} a_i(N_i - b_i)}{\sum_{i=1}^{k} a_i^2},$$

therefore the minimum distance of **N** *from* L *is*

$$distance^2 = \sum_{i=1}^{k}(N_i - b_i)^2 - \frac{\left(\sum_{i=1}^{k} a_i(N_i - b_i)\right)^2}{\sum_{i=1}^{k} a_i^2}. \tag{3.40}$$

□

3.3.2 Proof of Theorem 3.8

To prove this theorem, we need to find $\mathbf{P}$ and $\mathbf{F}_0^*$. However, it is not necessary to do this explicitly; noticing that the rank of $\mathbf{P}^{1/2}\mathbf{F}_0^*$ is no longer 1 suggests proving this theorem as follows.

To test the hypothesis (3.31), we need to define the cones $\widetilde{C}_\Omega$ and $\dot{C}_\Omega$. $\widetilde{C}_\Omega$ is

$$\widetilde{C}_\Omega = \{(\kappa^{(2)}, \kappa^{(3)}, \ldots, \kappa^{(d)}) \in \mathbb{R}^{d-1} : \kappa^{(2)} = \kappa^{(3)} = \ldots = \kappa^{(d)}\}.$$

But this cone after the transformation $\dot{C}_\Omega = \{\mathbf{P}^{1/2}\mathbf{F}_0^*\boldsymbol{\theta} : \boldsymbol{\theta} \in \widetilde{C}_\Omega\}$ changes to

$$\dot{C}_\Omega = \{(\kappa^{(2)}, \kappa^{(3)}, \ldots, \kappa^{(d)}) \in \mathbb{R}^{d-1} : a_2\kappa^{(2)} = a_3\kappa^{(3)} = \ldots = a_d\kappa^{(d)}\}.$$

The set $\dot{C}_\Omega$ is a line in the form of $a_2x_2 = a_3x_3 = \ldots = a_dx_d$. To find the distance of the point $\mathbf{N}$, described in (3.28) and Theorem 3.5, from the line $a_2x_2 = a_3x_3 = \ldots = a_dx_d$, we use the formula (3.40) which gives

$$distance^2 = \sum_{i=2}^{d} N_i^2 - \frac{\left(\sum_{i=2}^{d} \frac{N_i}{a_i}\right)^2}{\sum_{i=2}^{d} \frac{1}{a_i^2}},$$

and has a chi square distribution with $d-2$ degrees of freedom.

□

3.3.3 Proofs via Silvapulle and Sen Strategy

An alternative proof for Theorem 3.7 can be found in Silvapulle and Sen (2005). We use this to check our working. Consider

$$\begin{cases} H_0 & : \mathbf{R}\boldsymbol{\theta} = \mathbf{0} \\ H_A & : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0}, \end{cases}$$

where

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_{m_\ell - 1} & \mathbf{a} \end{pmatrix}_{(m_\ell - 1)\times m_\ell}, \quad \boldsymbol{\theta}^T = \begin{pmatrix} \kappa_{\ell 1} & \kappa_{\ell 2} & \cdots & \kappa_{\ell m_\ell} \end{pmatrix},$$

and

$$\mathbf{a}^T = \begin{pmatrix} -1 & -1 & \dots & -1 \end{pmatrix}_{1\times(m_\ell - 1)}.$$

This test is the one presented in (3.29). Since $\mathbf{R}$ is a linear space, we refer to Silvapulle and Sen, page 86 and the note before Proposition 3.7.3 where they explain the situation when the alternative hypothesis does not have any inequality constraints. In this case the distribution of the deviance statistic is chi square with degrees of freedom equal to the rank of $\mathbf{R}$. Here the distribution of d_n is chi square with $m_\ell - 1$ degrees of freedom.

For Theorem 3.8 and the hypothesis testing (3.31), it is enough to define $\mathbf{R}$ as

$$\mathbf{R} = \begin{pmatrix} Diagonal(a_2, a_3, \cdots, a_{d-1}) & \mathbf{a} \end{pmatrix}_{(m_\ell - 1)\times m_\ell}, \quad \mathbf{a}^T = \begin{pmatrix} -a_d, -a_d, \cdots, -a_d \end{pmatrix},$$

in the foregoing argument.

3.4 Example 1: A Full Analysis in 3 Dimensions

3.4.1 Introducing the Model

Consider the density function

$$f_{\mathbf{X}}(\mathbf{x}) = (f_{21}(\mathbf{x}))^{I_{\mathbf{x}}(S_{21})}(f_{22}(\mathbf{x}))^{I_{\mathbf{x}}(S_{22})}(f_{23}(\mathbf{x}))^{I_{\mathbf{x}}(S_{23})}(f_{31}(\mathbf{x}))^{I_{\mathbf{x}}(S_{31})}, \tag{3.41}$$

when $m_2 = 3$ and $m_3 = 1$. The model contains three 2 dimensional and one 3 dimensional spherical distributions. With $\mathbb{S}^d$ for the d dimensional sphere in $\mathbb{R}^d$, so that $\mathbf{x} \in \mathbb{S}^d$

implies $x_1^2 + x_2^2 + \cdots + x_d^2 = 1$, we denote

$$\begin{aligned} S_{21} &= \{\mathbf{x} \in \mathbb{S}^3 : \mathbf{x} = (x_1, x_2, 0), \quad x_1 \neq 0, x_2 \neq 0\}, \\ S_{22} &= \{\mathbf{x} \in \mathbb{S}^3 : \mathbf{x} = (x_1, 0, x_3), \quad x_1 \neq 0, x_3 \neq 0\}, \\ S_{23} &= \{\mathbf{x} \in \mathbb{S}^3 : \mathbf{x} = (0, x_2, x_3), \quad x_2 \neq 0, x_3 \neq 0\}, \\ S_{31} &= \{\mathbf{x} \in \mathbb{S}^3 : \mathbf{x} = (x_1, x_2, x_3), \quad x_1 \neq 0, x_2 \neq 0, x_3 \neq 0\}, \end{aligned} \tag{3.42}$$

and consider f_{21}, f_{22}, f_{23} to be 2 dimensional von Mises Fisher distributions with parameters $\boldsymbol{\mu}_{2j}$ and κ_{2j} and the density functions

$$f_{2j}(\mathbf{x}) = \frac{\exp\{\kappa_{2j}\boldsymbol{\mu}_{2j}^T\mathbf{x}_{2j}\}}{c^{(2)}(\kappa_{2j})}, \quad \boldsymbol{\mu}_{2j} \in \mathbb{S}^2, \mathbf{x}_{2j} \in \mathbb{S}^2, \kappa_{2j} > 0; j = 1, 2, 3,$$

and f_{31} to be a 3 dimensional von Mises Fisher distribution with parameters $\boldsymbol{\mu}_{31}$ and κ_{31} and the density function

$$f_{31}(\mathbf{x}) = \frac{\exp\{\kappa_{31}\boldsymbol{\mu}_{31}^T\mathbf{x}_{31}\}}{c^{(3)}(\kappa_{31})}, \quad \boldsymbol{\mu}_{31} \in \mathbb{S}^3, \mathbf{x}_{31} \in \mathbb{S}^3, \kappa_{31} > 0.$$

In these formulae, $c^{(2)}(\kappa)$ and $c^{(3)}(\kappa)$ are normalizing constants which are

$$c^{(2)}(\kappa) = 2\pi I_0(\kappa), \quad c^{(3)}(\kappa) = \frac{4\pi\sinh(\kappa)}{\kappa},$$

and $I_0(\kappa)$ is the modified Bessel function of the first kind.

Let n observations on identically independent random variables $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ be selected from the density function (3.41) and define z_{ijk} similar to (3.4). Then the logarithmic function corresponding to (3.41) is

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\mu}_{21}, \boldsymbol{\mu}_{22}, \boldsymbol{\mu}_{23}, \boldsymbol{\mu}_{31}, \kappa_{21}, \kappa_{22}, \kappa_{23}, \kappa_{31}) = \sum_{k=1}^{n}\{& z_{21k}\log(f_{21}(\mathbf{y}_k)) + z_{22k}\log(f_{22}(\mathbf{y}_k)) \\ &+ z_{23k}\log(f_{23}(\mathbf{y}_k)) + z_{31k}\log(f_{31}(\mathbf{y}_k))\}. \end{aligned} \tag{3.43}$$

Assume $k = 1, 2, \ldots, n$ and $\mathbf{X}_{2jk}$ be an element of $\mathbf{Y}_k$ which falls in region S_{2j} in (3.42), for $j = 1, 2, 3$, when the zero element is removed, so we have $\mathbf{X}_{2j} \in \mathbb{S}^2$. Therefore

$$\overline{\mathbf{X}}_{2j} = \frac{\sum_{k=1}^{n_{2j}} \mathbf{X}_{2jk}}{n_{2j}}, \quad j = 1, 2, 3 \tag{3.44}$$

and as a result for $\mathbf{X}_{31}$ in S_{31} we have

$$\overline{\mathbf{X}}_{31} = \frac{\sum_{k=1}^{n_{31}} \mathbf{X}_{31k}}{n_{31}}. \tag{3.45}$$

We write the log likelihood function (3.43) in the form

$$\mathcal{L}_n(\kappa_{21}, \boldsymbol{\mu}_{21}, \kappa_{22}, \boldsymbol{\mu}_{22}, \kappa_{23}, \boldsymbol{\mu}_{23}, \kappa_{31}, \boldsymbol{\mu}_{31}) = -\sum_{j=1}^{3} n_{2j} \log c^{(2)}(\kappa_{2j}) - n_{31} \log c^{(3)}(\kappa_{31}) + \sum_{i=2}^{3}\sum_{j=1}^{m_i} n_{ij}\kappa_{ij}\boldsymbol{\mu}_{ij}^T \overline{\mathbf{X}}_{ij}, \tag{3.46}$$

where $\sum_{i=2}^{3}\sum_{j=1}^{m_i} n_{ij} = n$, $m_2 = 3$, $m_3 = 1$ and $\overline{\mathbf{X}}_{ij}$ for $i = 2, 3$, $j = 1, \ldots, m_i$ are the sample means in (3.44) and (3.45).

We take $\boldsymbol{\theta} = (\kappa_{21}, \boldsymbol{\mu}_{21}, \kappa_{22}, \boldsymbol{\mu}_{22}, \kappa_{23}, \boldsymbol{\mu}_{23}, \kappa_{31}, \boldsymbol{\mu}_{31})$ and the true parameter to be $\boldsymbol{\theta}_0 = (\kappa_{021}, \boldsymbol{\mu}_{021}, \kappa_{022}, \boldsymbol{\mu}_{022}, \kappa_{023}, \boldsymbol{\mu}_{023}, \kappa_{031}, \boldsymbol{\mu}_{031})$. Then the parameter space is

$$\Theta := (0,\infty) \times \mathbb{S}^2 \times (0,\infty) \times \mathbb{S}^2 \times (0,\infty) \times \mathbb{S}^2 \times (0,\infty) \times \mathbb{S}^3 = \mathbb{S}^9 \times (0,\infty)^4. \tag{3.47}$$

We collect all the restrictions on the parameters into an $s \times 1$ vector $\mathbf{h}(\boldsymbol{\theta})$ such that we can write them as $\mathbf{h}(\boldsymbol{\theta}) = \mathbf{0}$. Restrictions can be defined by the parameter space or by the null or alternative hypothesis. When they are under the hypotheses, we call them *model* constraints as introduced by Neuenschwander and Flury (1997). Assume $\mathbf{h}(\boldsymbol{\theta})$ is twice differentiable.

In this section we would like to find the MEs corresponding to the function (3.46), discuss their properties and test the hypothesis

$$\begin{cases} H_0 & : \kappa_{21} = \kappa_{22} = \kappa_{23} \\ H_A & : \textit{otherwise}, \end{cases} \tag{3.48}$$

We do this by finding the asymptotic distribution of $d_n = -2(\sup_{\boldsymbol{\theta}\in\Omega} \mathcal{L}_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\mathcal{T}} \mathcal{L}_n(\boldsymbol{\theta}))$. The parameter space under the null hypothesis, Ω, for the hypothesis (3.48) is

$$\Omega = \mathbb{S}^9 \times (0,\infty)^2. \tag{3.49}$$

The first and second derivatives of $\mathcal{L}_n(\boldsymbol{\theta})$ exist on a neighborhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$. We define $\mathbf{S}_n(\boldsymbol{\theta})$ to be the first derivative of the likelihood function (3.46) and $\mathbf{F}_n(\boldsymbol{\theta})$ to be the negative of the second derivative, so we have

$$\mathbf{S}_n(\boldsymbol{\theta}) := \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -n_{21}\dfrac{c^{(2)}_{\kappa_{21}}}{c^{(2)}(\kappa_{21})} + n_{21}\boldsymbol{\mu}_{21}^T\overline{\mathbf{X}}_{21} \\ n_{21}\kappa_{21}\overline{\mathbf{X}}_{21} \\ -n_{22}\dfrac{c^{(2)}_{\kappa_{22}}}{c^{(2)}(\kappa_{22})} + n_{22}\boldsymbol{\mu}_{22}^T\overline{\mathbf{X}}_{22} \\ n_{22}\kappa_{22}\overline{\mathbf{X}}_{22} \\ -n_{23}\dfrac{c^{(2)}_{\kappa_{23}}}{c^{(2)}(\kappa_{23})} + n_{23}\boldsymbol{\mu}_{23}^T\overline{\mathbf{X}}_{23} \\ n_{23}\kappa_{23}\overline{\mathbf{X}}_{23} \\ -n_{31}\dfrac{c^{(3)}_{\kappa_{31}}}{c^{(3)}(\kappa_{31})} + n_{31}\boldsymbol{\mu}_{31}^T\overline{\mathbf{X}}_{31} \\ n_{31}\kappa_{31}\overline{\mathbf{X}}_{31} \end{pmatrix}_{13\times 1} \tag{3.50}$$

and

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\theta})}{\partial \theta \partial \boldsymbol{\theta}^T} = \begin{pmatrix} n_{21}\mathbf{F_{21}} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & n_{22}\mathbf{F}_{22} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & n_{23}\mathbf{F}_{23} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & n_{31}\mathbf{F}_{31} \end{pmatrix}_{13\times 13}, \tag{3.51}$$

where

$$\mathbf{F}_{ij} = \begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & \mathbf{0} \end{pmatrix}_{(i+1)\times(i+1)} ; \quad i = 2, 3, \quad j = 1, \ldots, m_i,$$

$m_2 = 3$ and $m_3 = 1$. We have

$$Det(\mathbf{F}_n(\boldsymbol{\theta})) = (n_{21}n_{22}n_{23})^3 n_{31}^4 Det(\mathbf{F}_{21})Det(\mathbf{F}_{22})Det(\mathbf{F}_{23})Det(\mathbf{F}_{31}),$$

where $Det(\mathbf{F}_{ij}) = 0$; therefore, $\mathbf{F}_n(\boldsymbol{\theta})$ is singular. The eigenvalues of $\mathbf{F}_{ij}$ are

$$\begin{cases} 0 \quad ; \text{d-2 times, d=3,4} \\ \frac{1}{2}\left(a(\kappa_{ij}) - \sqrt{a(\kappa_{ij})^2 + 4||\overline{\mathbf{X}}_{ij}||^2}\right) \\ \frac{1}{2}\left(a(\kappa_{ij}) + \sqrt{a(\kappa_{ij})^2 + 4||\overline{\mathbf{X}}_{ij}||^2}\right) \end{cases} \tag{3.52}$$

for $i = 2, 3$ and $j = 1, \ldots, m_i$, $m_2 = 1, 2, 3$ and $m_3 = 1$.

As we see from (3.50)

$$E\mathbf{S}_n(\boldsymbol{\theta}) = \begin{pmatrix} 0 \\ n_{21}\kappa_{21}\dfrac{c_{\kappa_{21}}}{c(\kappa_{21})}\boldsymbol{\mu}_{21} \\ 0 \\ n_{22}\kappa_{22}\dfrac{c_{\kappa_{22}}}{c(\kappa_{22})}\boldsymbol{\mu}_{22} \\ 0 \\ n_{23}\kappa_{23}\dfrac{c_{\kappa_{23}}}{c(\kappa_{23})}\boldsymbol{\mu}_{23} \\ 0 \\ n_{31}\kappa_{31}\dfrac{c_{\kappa_{31}}}{c(\kappa_{31})}\boldsymbol{\mu}_{31} \end{pmatrix}_{13\times 1}.$$

This is not zero for any $\boldsymbol{\theta} \in \Theta$, and $\mathbf{F}_n(\boldsymbol{\theta})$ is singular for all $\boldsymbol{\theta} \in \Theta$. So the "standard" asymptotic theory for MEs does not apply here.

3.4.2 Setup for the Hypothesis $H_0 : \kappa_{21} = \kappa_{22} = \kappa_{23}$

For the hypothesis (3.48) we have

$$\mathbf{h}_1(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}_{21}^T\boldsymbol{\mu}_{21} - 1 \\ \boldsymbol{\mu}_{22}^T\boldsymbol{\mu}_{22} - 1 \\ \boldsymbol{\mu}_{23}^T\boldsymbol{\mu}_{23} - 1 \\ \boldsymbol{\mu}_{31}^T\boldsymbol{\mu}_{31} - 1 \end{pmatrix} = \mathbf{0}_{4\times 1} \quad \text{and} \quad \mathbf{h}_2(\boldsymbol{\theta}) = \begin{pmatrix} \kappa_{21} - \kappa_{22} \\ \kappa_{22} - \kappa_{23} \end{pmatrix} = \mathbf{0}_{2\times 1}.$$

The derivative of these vectors with respect to $\boldsymbol{\theta}$ is

$$\mathbf{H}_1(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 2\boldsymbol{\mu}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & 0 \\ \mathbf{0} & 2\boldsymbol{\mu}_{22} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & 2\boldsymbol{\mu}_{23} & \mathbf{0} \\ 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 2\boldsymbol{\mu}_{31} \end{pmatrix}_{13\times 4}, \mathbf{H}_2(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 0 \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ -1 & 1 \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ 0 & -1 \\ \mathbf{0}_{6\times 6} & \mathbf{0}_{6\times 6} \end{pmatrix}_{13\times 2}.$$

We define the vectors of Lagrange Multipliers to be

$$\boldsymbol{\lambda}_{21} = \begin{pmatrix} \lambda_{21} \\ \lambda_{22} \\ \lambda_{23} \\ \lambda_{31} \end{pmatrix}_{4\times 1} \quad \text{and} \quad \boldsymbol{\lambda}_2 = \begin{pmatrix} \lambda_5 \\ \lambda_6 \end{pmatrix}_{2\times 1},$$

where $\boldsymbol{\lambda}_1 \in \mathbb{R}^4$, and $\boldsymbol{\lambda}_2 \in \mathbb{R}^2$. There exist constant vectors $\mathbf{C}_{1n}$ and $\mathbf{C}_{2n}$

$$\mathbf{C}_{1n} = a \begin{pmatrix} n_{21} & 0 & 0 & 0 \\ 0 & n_{22} & 0 & 0 \\ 0 & 0 & n_{23} & 0 \\ 0 & 0 & 0 & n_{31} \end{pmatrix}_{4\times 4} \quad \text{and} \quad \mathbf{C}_{2n} = a \begin{pmatrix} n_{21} & 0 \\ 0 & n_{22} \end{pmatrix}_{2\times 2}$$

when $a > 0$, such that

$$\boldsymbol{\lambda}_{1n} = \mathbf{C}_{1n}\boldsymbol{\lambda}_1, \quad \text{and} \quad \boldsymbol{\lambda}_{2n} = \mathbf{C}_{2n}\boldsymbol{\lambda}_2, \tag{3.53}$$

so we have

$$\mathbf{H}_1(\boldsymbol{\theta})\boldsymbol{\lambda}_{1n} + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n} = a \begin{pmatrix} n_{21}\lambda_5 \\ 2n_{21}\lambda_{21}\boldsymbol{\mu}_{21} \\ n_{22}\lambda_6 - n_{21}\lambda_5 \\ 2n_{22}\lambda_{22}\boldsymbol{\mu}_{22} \\ -n_{22}\lambda_6 \\ 2n_{23}\lambda_{23}\boldsymbol{\mu}_{23} \\ 0 \\ 2n_{31}\lambda_{31}\boldsymbol{\mu}_{31} \end{pmatrix}_{13\times 1}.$$

To find the MEs with their distributions and prove their consistency and also to be able to do the hypothesis test (3.48), we apply Maller (2015) methodology and define

$$\mathbf{S}_n^{\lambda}(\boldsymbol{\theta}) = \mathbf{S}_n(\boldsymbol{\theta}) + \mathbf{H}_1(\boldsymbol{\theta})\boldsymbol{\lambda}_{1n} + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n} = \begin{pmatrix} \mathbf{S}_{21}^{\lambda} \\ \mathbf{S}_{22}^{\lambda} \\ \mathbf{S}_{23}^{\lambda} \\ \mathbf{S}_{31}^{\lambda} \end{pmatrix}, \tag{3.54}$$

where

$$\mathbf{S}_{21}^{\lambda} = n_{21}\begin{pmatrix} \boldsymbol{\mu}_{21}^T(\overline{\mathbf{X}}_{21} - E\mathbf{X}_{21}) + a\lambda_5 \\ \kappa_{21}\overline{\mathbf{X}}_{21} + 2a\lambda_{21}\boldsymbol{\mu}_{21} \end{pmatrix}_{3\times 3},$$

$$\mathbf{S}_{22}^{\lambda} = n_{22}\begin{pmatrix} \boldsymbol{\mu}_{22}^T(\overline{\mathbf{X}}_{22} - E\mathbf{X}_{22}) + a\lambda_6 - a\dfrac{n_{21}}{n_{22}}\lambda_5 \\ \kappa_{22}\overline{\mathbf{X}}_{22} + 2a\lambda_{22}\boldsymbol{\mu}_{22} \end{pmatrix}_{3\times 3},$$

$$\mathbf{S}_{23}^{\lambda} = n_{23}\begin{pmatrix} \boldsymbol{\mu}_{23}^T(\overline{\mathbf{X}}_{23} - E\mathbf{X}_{23}) - a\dfrac{n_{22}}{n_{23}}\lambda_6 \\ \kappa_{23}\overline{\mathbf{X}}_{23} + 2a\lambda_{23}\boldsymbol{\mu}_{23} \end{pmatrix}_{3\times 3}$$

and

$$\mathbf{S}_{31}^{\lambda} = n_{31}\begin{pmatrix} \boldsymbol{\mu}_{31}^T(\overline{\mathbf{X}}_{31} - E\mathbf{X}_{31}) \\ \kappa_{31}\overline{\mathbf{X}}_{31} + 2a\lambda_{31}\boldsymbol{\mu}_{31} \end{pmatrix}_{4\times 4}.$$

Also, the derivative of $\mathbf{S}_n^{\lambda}(\theta)$ is

$$\mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) = \frac{\partial \mathbf{S}_n^{\lambda}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{F}_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}}\left(\mathbf{H}_1(\boldsymbol{\theta})\boldsymbol{\lambda}_{1n} + \mathbf{H}_2(\boldsymbol{\theta})\boldsymbol{\lambda}_{2n}\right) \tag{3.55}$$

$$= \begin{pmatrix} \mathbf{F}_{1n}^{\lambda} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & \mathbf{F}_{2n}^{\lambda} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{F}_{3n}^{\lambda} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{F}_{4n}^{\lambda} \end{pmatrix}, \tag{3.56}$$

where for $i = 2, 3$ and $j = 1, \ldots, m_i$, we have

$$\mathbf{F}_{ijn}^{\lambda} = n_{ij}\begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & -2\lambda_{ij}a\mathbf{I}_i \end{pmatrix}_{(i+1)\times(i+1)} \tag{3.57}$$

such that $m_2 = 3$ and $m_3 = 1$. Let $n \to \infty$, then by the weak law of large numbers

$$\frac{1}{n}\mathbf{F}_{ijn}^{\lambda} \xrightarrow{\text{P}} \mathbf{F}_{ij}^{\lambda},$$

where

$$\mathbf{F}_{ij}^{\lambda} = p_{ij}\begin{pmatrix} a(\kappa_{ij}) & -\boldsymbol{\mu}^T||E\mathbf{X}_{ij}|| \\ -\boldsymbol{\mu}||E\mathbf{X}_{ij}|| & -2a\lambda_{ij}\mathbf{I}_i \end{pmatrix}.$$

As discussed in the Appendix, Langevin Example section, we need to define $\mathbf{F}_n^{\lambda*}(\boldsymbol{\theta})$ described in (A.9). We choose $\mathbf{G}_n(\boldsymbol{\theta})$ to be

$$\mathbf{G}_n(\boldsymbol{\theta}) = \frac{\sqrt{b}}{2}\begin{pmatrix} \sqrt{n_{21}} & 0 & 0 & 0 \\ 0 & \sqrt{n_{22}} & 0 & 0 \\ 0 & 0 & \sqrt{n_{23}} & 0 \\ 0 & 0 & 0 & \sqrt{n_{31}} \end{pmatrix}_{4\times 4}.$$

Then we have

$$\begin{aligned}\mathbf{F}_n^{\lambda*}(\boldsymbol{\theta}) &= \mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) + \mathbf{H}_1(\boldsymbol{\theta})\mathbf{G}_n\mathbf{G}_n^T\mathbf{H}_1^T(\boldsymbol{\theta}) \\ &= \begin{pmatrix} \mathbf{F}_{21n}^{\lambda*} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{3\times3} & \mathbf{F}_{22n}^{\lambda*} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{F}_{23n}^{\lambda*} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{F}_{31n}^{\lambda*} \end{pmatrix}_{13\times13},\end{aligned}$$

where

$$\mathbf{F}_{ijn}^{\lambda*} = n_{ij}\begin{pmatrix} a(\kappa_{ij}) & -\overline{\mathbf{X}}_{ij}^T \\ -\overline{\mathbf{X}}_{ij} & -2\lambda_{ij}a\mathbf{I} + b\boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^T \end{pmatrix}_{(i+1)\times(i+1)}.$$

By the weak law of large numbers

$$\frac{1}{n}\mathbf{F}_n^{\lambda*}(\boldsymbol{\theta}) \xrightarrow{\text{P}} \mathbf{F}^{\lambda*}(\boldsymbol{\theta}) := \begin{pmatrix} p_{21}\mathbf{E}\mathbf{F}_{21}^{\lambda*} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{3\times3} & p_{22}\mathbf{E}\mathbf{F}_{22}^{\lambda*} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & p_{23}\mathbf{E}\mathbf{F}_{23}^{\lambda*} & \mathbf{0}_{3\times4} \\ \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & \mathbf{0}_{4\times3} & p_{31}\mathbf{E}\mathbf{F}_{31}^{\lambda*} \end{pmatrix},$$

as $\sum_{i=2}^{3}\sum_{j=1}^{m_i} n_{ij} = n \to \infty$, where

$$\mathbf{E}\mathbf{F}_{ij}^{\lambda*} = \begin{pmatrix} a(\kappa_{ij}) & -\boldsymbol{\mu}_{ij}^T||E\mathbf{X}_{ij}|| \\ -\boldsymbol{\mu}_{ij}||E\mathbf{X}_{ij}|| & -2\lambda_{ij}a\mathbf{I}_i + b\boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^T \end{pmatrix}. \tag{3.58}$$

If we choose b to be bigger than all the $||E\mathbf{X}_{ij}||^2/a(\kappa_{ij})$ elements, then (A.13) remains valid.

3.4.3 The True Value of λ

In order to calculate $\boldsymbol{\lambda}_0$, it is sufficient to solve the system of equations

$$\mathbf{L}_0 + \mathbf{H}_1(\boldsymbol{\theta}_0)\mathbf{C}_1\boldsymbol{\lambda}_{21} + \mathbf{H}_2(\boldsymbol{\theta}_0)\mathbf{C}_2\boldsymbol{\lambda}_2 = 0, \tag{3.59}$$

where

$$\mathbf{D}_n^{-1}\mathbf{S}_n(\boldsymbol{\theta}) \xrightarrow{\text{P}} \mathbf{L} := \begin{pmatrix} 0 \\ p_{21}\kappa_{21}\dfrac{c_{\kappa 1}}{c(\kappa_{21})}\boldsymbol{\mu}_{21} \\ 0 \\ p_{22}\kappa_{22}\dfrac{c_{\kappa 2}}{c(\kappa_{22})}\boldsymbol{\mu}_{22} \\ 0 \\ p_{23}\kappa_{23}\dfrac{c_{\kappa 3}}{c(\kappa_{23})}\boldsymbol{\mu}_{23} \\ 0 \\ p_{31}\kappa_{31}\dfrac{c_{\kappa 4}}{c(\kappa_{31})}\boldsymbol{\mu}_{31} \end{pmatrix}_{13\times 1}, \quad \text{as} \quad \text{n} \to \infty \tag{3.60}$$

and

$$\mathbf{C}_1 = \lim_{n\to\infty}\frac{\mathbf{C}_{1n}}{n} = a\begin{pmatrix} p_{21} & 0 & 0 & 0 \\ 0 & p_{22} & 0 & 0 \\ 0 & 0 & p_{23} & 0 \\ 0 & 0 & 0 & p_{31} \end{pmatrix}_{4\times 4}, \quad \mathbf{C}_2 = \lim_{n\to\infty}\frac{\mathbf{C}_{2n}}{n} = a\begin{pmatrix} p_{21} & 0 \\ 0 & p_{22} \end{pmatrix}_{2\times 2} \tag{3.61}$$

and we have

$$\begin{aligned} -2a\lambda_{0ij} &= \frac{\kappa_{0ij}c_{\kappa_{0ij}}}{c(\kappa_{0ij})} = \kappa_{0ij}||E_0\mathbf{X}_{ij}||; \quad i = 2,3 \quad j = 1,\dots,m_i, \\ \lambda_{05} &= \lambda_{06} = 0. \end{aligned} \tag{3.62}$$

Hence

$$\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{S}_{21}^{\lambda_0} \\ \mathbf{S}_{22}^{\lambda_0} \\ \mathbf{S}_{23}^{\lambda_0} \\ \mathbf{S}_{31}^{\lambda_0} \end{pmatrix},$$

where

$$\mathbf{S}_{ij}^{\lambda_0} = n_{ij}\begin{pmatrix} \boldsymbol{\mu}_{0ij}^T(\overline{\mathbf{X}}_{ij} - E_0\mathbf{X}_{ij}) \\ \kappa_{0ij}(\overline{\mathbf{X}}_{ij} - E_0\mathbf{X}_{ij}) \end{pmatrix}; \quad i = 2,3, \quad j = 1,\dots,m_i. \tag{3.63}$$

We have

$$\mathrm{Var}_0(\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)) = \mathrm{E}_0\{(\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0))(\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0))^T\} = n\begin{pmatrix} p_{21}\mathbf{v}_{21}^{\lambda} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & p_{22}\mathbf{v}_{22}^{\lambda} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & p_{23}\mathbf{v}_{23}^{\lambda} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & \mathbf{0}_{4\times 3} & p_{31}\mathbf{v}_{31}^{\lambda} \end{pmatrix} \tag{3.64}$$

$$= n\mathbf{V}(\boldsymbol{\theta}_0) = n\mathbf{V}_0$$

where

$$\mathbf{v}_{ij}^{\lambda} = \begin{pmatrix} a(\kappa_{0ij}) & \kappa_{0ij}a(\kappa_{0ij})\boldsymbol{\mu}_{0ij}^T \\ \kappa_{0ij}a(\kappa_{0ij})\boldsymbol{\mu}_{0ij} & \kappa_{0ij}^2\mathrm{Var}_0(\mathbf{X}_{ij}) \end{pmatrix}; \quad i = 2,3, \quad j = 1,\ldots,m_i.$$

Letting E_0 and Var_0 be the expectation and variance when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we have

$$\mathrm{Var}_0(\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0)) = \mathbf{V}_0.$$

This gives us

$$\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) = O_p(1), \quad \text{as} \quad n \to \infty$$

and (A.12) is satisfied. Assumption (A.13) of the paper comes from (3.58) and the discussion following that.

3.4.4 MEs of the Parameters

Now we can calculate MEs through solving the system of equations $\mathbf{S}_n^{\lambda}(\widehat{\boldsymbol{\theta}}_n) = 0$ and $\mathbf{h}(\widehat{\boldsymbol{\theta}}_n) = 0$. From $\mathbf{h}(\widehat{\boldsymbol{\theta}}) = 0$, we have $\widehat{\kappa}_{21} = \widehat{\kappa}_{22} = \widehat{\kappa}_{23}$ which are all assumed to be equal to $\widehat{\kappa}$. Focusing on $\mathbf{S}_n^{\lambda}(\widehat{\boldsymbol{\theta}}_n) = 0$ and from $\widehat{\kappa}\overline{\mathbf{X}}_{ij} + 2a\lambda_{ij}\boldsymbol{\mu}_{ij} = 0$ for $i = 2,3$ and $j = 1,\ldots,m_i$ we have

$$\begin{aligned} \widehat{\kappa}\overline{\mathbf{X}}_{ij} &= -2a\widehat{\lambda}_{ij}\boldsymbol{\mu}_{ij} \\ \widehat{\kappa}||\overline{\mathbf{X}}_{ij}|| &= -2a\widehat{\lambda}_{ij}||\boldsymbol{\mu}_{ij}|| \\ \widehat{\kappa}||\overline{\mathbf{X}}_{ij}|| &= -2a\widehat{\lambda}_{ij}; \quad i = 2,3 \quad j = 1,\ldots,m_i. \end{aligned} \tag{3.65}$$

Also we can write

$$\widehat{\kappa}\overline{\mathbf{X}}_{ij} + 2a\widehat{\lambda}_i \boldsymbol{\mu}_{ij} = 0$$
$$\widehat{\boldsymbol{\mu}_{ij}} = \frac{\widehat{\kappa}\overline{\mathbf{X}}_{ij}}{-2a\widehat{\lambda}_{ij}}. \tag{3.66}$$

From (3.65) and (3.66) we have

$$\widehat{\boldsymbol{\mu}_{ij}} = \frac{\widehat{\kappa}\overline{\mathbf{X}}_{ij}}{\widehat{\kappa}||\overline{\mathbf{X}}_{ij}||}$$
$$\widehat{\boldsymbol{\mu}_{ij}} = \frac{\overline{\mathbf{X}}_{ij}}{||\overline{\mathbf{X}}_{ij}||}; \quad i = 2,3 \quad j = 1,\ldots,m_i.$$

From other equations in $\mathbf{S}_n^{\lambda}(\widehat{\boldsymbol{\theta}}_n) = 0$, we have

$$\boldsymbol{\mu}_{21}^T(\overline{\mathbf{X}}_1 - E\mathbf{X}_{21}) + a\lambda_5 = 0$$
$$\frac{\overline{\mathbf{X}}_{21}^T}{||\overline{\mathbf{X}}_{21}||}(\overline{\mathbf{X}}_{21} - \frac{c_{\widehat{\kappa}}}{c(\widehat{\kappa})}\frac{\overline{\mathbf{X}}_{21}}{||\overline{\mathbf{X}}_{21}||}) + a\lambda_5 = 0$$
$$a\widehat{\lambda}_5 = \frac{c_{\widehat{\kappa}}}{c(\widehat{\kappa})} - ||\overline{\mathbf{X}}_{21}|| \tag{3.67}$$

and

$$-a\widehat{\lambda}_6 = \frac{n_{23}}{n_{22}}(\frac{c_{\widehat{\kappa}}}{c(\widehat{\kappa})} - ||\overline{\mathbf{X}}_{23}||).$$

If we add three system equations in the first row of $\mathbf{S}_{ij}$'s, we obtain the formula

$$\frac{c_{\widehat{\kappa}}^{(2)}}{c^{(2)}(\widehat{\kappa})} = \frac{\sum_{i=1}^{3} n_{ij}||\overline{\mathbf{X}}_i||}{\sum_{i=1}^{3} n_{ij}}. \tag{3.68}$$

Finally, from $\boldsymbol{\mu}_{31}^T(\overline{\mathbf{X}}_{31} - E\mathbf{X}_{31}) = 0$, we have

$$\frac{c_{\widehat{\kappa}_{31}}^{(3)}}{c^{(3)}(\widehat{\kappa}_{31})} = ||\overline{\mathbf{X}}_{31}||.$$

3.4.5 Asymptotic Distributions

Theorem A.3 discusses the distribution of $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$. Suppose
1) $\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) \xrightarrow{\text{D}} \mathbf{Z}$, as n $\to \infty$,

2) define

$$\mathbf{U}_n^{\lambda}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{F}_n^{\lambda *}(\boldsymbol{\theta}) & -\mathbf{H}_1(\theta)\mathbf{C}_{1n} & -\mathbf{H}_2(\boldsymbol{\theta})\mathbf{C}_{2n} \\ -\mathbf{C}_{1n}^T\mathbf{H}_1^T(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \\ -\mathbf{C}_{2n}^T\mathbf{H}_2^T(\boldsymbol{\theta}) & \mathbf{0} & \mathbf{0} \end{pmatrix}_{(13+4+2)\times(13+4+2)},$$

in such a way that for a nonsingular matrix $\mathbf{J}_n$, we have $\mathbf{J}_n^{-1}\mathbf{U}_n^{\lambda_n}(\boldsymbol{\theta}_n)\mathbf{J}_n^{-T} \xrightarrow{\text{P}} \mathbf{U}_0$. Then as $n \to \infty$

$$\mathbf{J}_n \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\text{D}} -\mathbf{U}_0^{-1} \begin{pmatrix} \mathbf{Z} \\ \mathbf{0} \end{pmatrix}.$$

For our example, we define $\mathbf{J}_n$ to be

$$\mathbf{J}_n = \sqrt{n}\mathbf{I}_{13+4+2},$$

then by using a well known theorem in probability which says: If $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$ and f is a continuous function then $f(X_n, Y_n) \xrightarrow{\text{P}} f(X, Y)$, we have

$$\mathbf{J}_n^{-T}(\mathbf{U}_n^{\lambda}(\boldsymbol{\theta}_n))^{-1}\mathbf{J}_n^{-1} =: \mathbf{U}_n \xrightarrow{\text{P}} \mathbf{U_0} = \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}^T & \mathbf{R} \end{pmatrix}_{19\times 19}.$$

Matrix $\mathbf{P}$ has a format similar to Figure 3.1. In this figure the white regions represent zeros and the red ones represent non zero elements. If we denote its elements as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{21} & \mathbf{P}_5 & \mathbf{P}_5 & \mathbf{0} \\ \mathbf{P}_5 & \mathbf{P}_{22} & \mathbf{P}_5 & \mathbf{0} \\ \mathbf{P}_5 & \mathbf{P}_5 & \mathbf{P}_{23} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{P}_{31} \end{pmatrix}_{13\times 13},$$

then with $\overline{a(\kappa_0)} = p_{21}a(\kappa_{021}) + p_{22}a(\kappa_{022}) + p_{23}a(\kappa_{023})$, we get

$$\mathbf{P}_{ij} = \begin{pmatrix} \frac{1}{\overline{\overline{a(\kappa_0)}}} & \mathbf{0}^T \\ \mathbf{0} & \frac{1}{p_{ij}\kappa_{0ij}A(\kappa_{0ij})}(\mathbf{I}_i - \boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}^T) \end{pmatrix}_{(i+1)\times(i+1)} ; \quad i = 2, j = 1, 2, 3. \tag{3.69}$$

For $i = 3$ and $j = 1$, we have

$$\mathbf{P}_{31} = \begin{pmatrix} \frac{1}{p_{31}a(\kappa_{031})} & \mathbf{0}^T \\ \mathbf{0} & \frac{1}{p_{31}\kappa_{031}A(\kappa_{031})}(\mathbf{I}_i - \boldsymbol{\mu}_{31}\boldsymbol{\mu}_{31}^T) \end{pmatrix}_{(3+1)\times(3+1)}$$

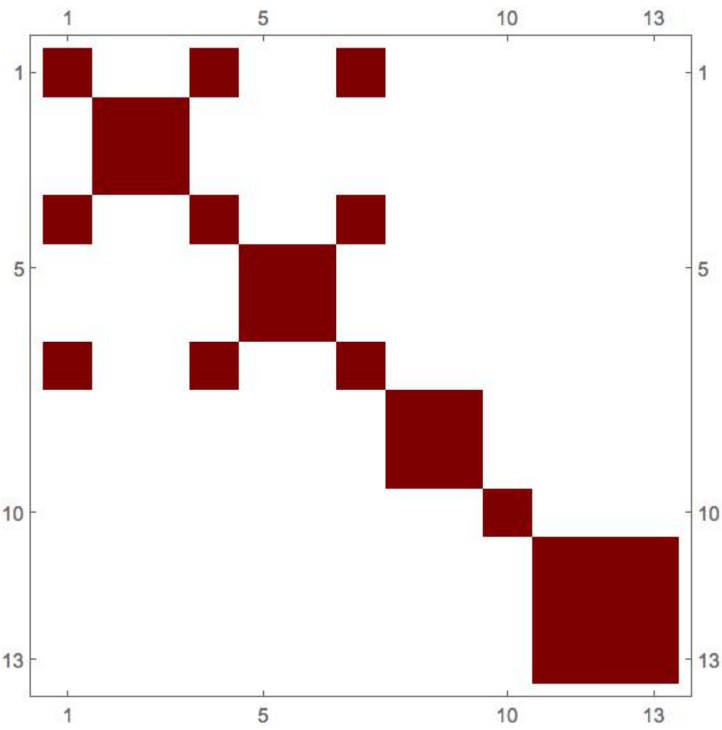

FIGURE 3.1: Schematic matrix plot for $\mathbf{P}$ under $H_0 : \kappa_{21} = \kappa_{22} = \kappa_{23}$

and $\mathbf{P}_5$ is

$$\mathbf{P}_5 = \begin{pmatrix} \frac{1}{\overline{a(\kappa_0)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{3\times 3}.$$

The matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 & \mathbf{Q}_3 & \mathbf{0} & \mathbf{Q}_5 & \mathbf{Q}_6 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}_4 & \mathbf{0} & \mathbf{0} \end{pmatrix}_{13\times 6},$$

where

$$\mathbf{Q}_5 = \begin{pmatrix} \frac{-1}{p_{21}} + \frac{a(\kappa_{021})}{\overline{a(\kappa_0)}} \\ \mathbf{0}_{2\times 1} \\ \frac{a(\kappa_{021})}{\overline{a(\kappa_0)}} \\ \mathbf{0}_{2\times 1} \\ \frac{a(\kappa_{021})}{\overline{a(\kappa_0)}} \\ \mathbf{0}_{2\times 1} \end{pmatrix}_{9\times 1}, \quad \mathbf{Q}_6 = \frac{1}{p_{22}\overline{a(\kappa_0)}} \begin{pmatrix} -p_{23}a(\kappa_{023}) \\ \mathbf{0}_{2\times 1} \\ -p_{23}a(\kappa_{023}) \\ \mathbf{0}_{2\times 1} \\ p_{21}a(\kappa_{021}) + p_{22}a(\kappa_{022}) \\ \mathbf{0}_{2\times 1} \end{pmatrix}_{9\times 1},$$

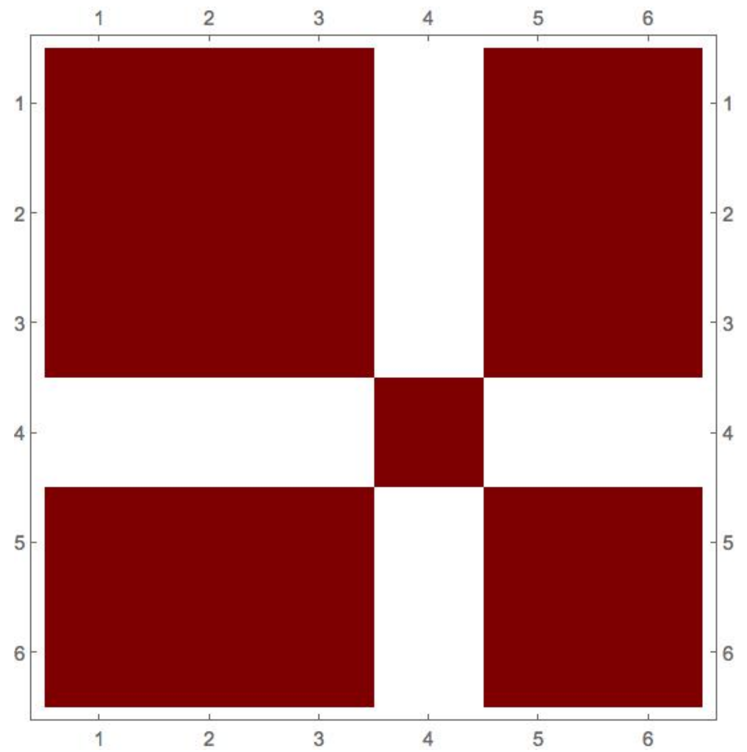

FIGURE 3.2: Schematic matrix plot for **R** under $H_0 : \kappa_{21} = \kappa_{22} = \kappa_{23}$

$$\mathbf{Q}_1 = \frac{-1}{2p_{21}} \begin{pmatrix} \frac{p_{21}A(\kappa_{021})}{a(\kappa_0)} \\ \boldsymbol{\mu}_{21} \\ \frac{p_{21}A(\kappa_{021})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \\ \frac{p_{21}A(\kappa_{021})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \end{pmatrix}_{9\times 1}, \quad \mathbf{Q}_2 = \frac{-1}{2p_{22}} \begin{pmatrix} \frac{p_{22}A(\kappa_{022})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \\ \frac{p_{22}A(\kappa_{022})}{a(\kappa_0)} \\ \boldsymbol{\mu}_{22} \\ \frac{p_{22}A(\kappa_{022})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \end{pmatrix}_{9\times 1}, \quad \mathbf{Q}_3 = \frac{-1}{2p_{23}} \begin{pmatrix} \frac{p_{23}A(\kappa_{023})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \\ \frac{p_{23}A(\kappa_{023})}{a(\kappa_0)} \\ \mathbf{0}_{2\times 1} \\ \frac{p_{23}A(\kappa_{023})}{a(\kappa_0)} \\ \boldsymbol{\mu}_{23} \end{pmatrix}_{9\times 1}$$

and

$$\mathbf{Q}_4 = \frac{-1}{2p_{31}} \begin{pmatrix} \frac{A(\kappa_{031})}{a(\kappa_{031})} \\ \boldsymbol{\mu}_{31} \end{pmatrix}_{4\times 1}.$$

The matrix **R** is a 6×6 symmetric matrix as shown in Figure 3.2. The first 3×3 red square in this plot has the diagonal elements

$$r_{ii} = \frac{\overline{a(\kappa_0)}(b + A(\kappa_{02j})k_{0i})}{-p_{2j}} + A(\kappa_{02j})^2, \quad i = 1, 2, 3.$$

The other elements in this square are, for $i \neq j = 1, 2, 3$,

$$r_{ij} = r_{ji} = A(\kappa_{02i})A(\kappa_{02j}).$$

The little 1×1 square in the middle is

$$r_{44} = \frac{\overline{a(\kappa_0)}(b + A(\kappa_{031})\kappa_{031})}{-p_{31}} + \frac{\overline{a(\kappa_{031})}A(\kappa_{031}^2)}{p_{31}a(\kappa_{031})}.$$

The diagonal elements of the last 2×2 red matrix are

$$\begin{aligned} r_{55} &= \frac{4a(\kappa_{021})\overline{a(\kappa_0)}}{-p_{21}} + 4a(\kappa_{021})^2, \\ r_{66} &= \frac{4a(\kappa_{023})\overline{a(\kappa_0)}}{-p_{22}^2} + \frac{4a(\kappa_{023})^2 p_{23}^2}{p_{22}^2}, \end{aligned}$$

with

$$r_{65} = r_{56} = -\frac{4p_{23}a(\kappa_{021})a(\kappa_{023})}{p_{22}}.$$

The 2×3 or 3×2 red rectangles in Figure 3.2 have the elements

$$r_{i5} = r_{5i} = -2A(\kappa_{02i})a(\kappa_{021}), \quad i = 2, 3$$

and

$$r_{i6} = r_{6i} = \frac{2p_{23}A(\kappa_{02i})a(\kappa_{023})}{p_{22}}, \quad i = 1, 2$$

and

$$r_{51} = r_{15} = \frac{2A(\kappa_{021})\left(\overline{a(\kappa_0)} - p_{21}a(\kappa_{021})\right)}{p_{21}}$$

and

$$r_{63} = r_{36} = \frac{2A(\kappa_{023})\left(\overline{a(\kappa_0)} - p_{23}a(\kappa_{023})\right)}{p_{22}}.$$

Now, we can verify the covariance matrix of the MEs as

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\mathrm{D}} \mathbf{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{P}\mathbf{V}_0\mathbf{P} & \mathbf{P}\mathbf{V}_0\mathbf{Q} \\ \mathbf{Q}^T\mathbf{V}_0\mathbf{P} & \mathbf{Q}^T\mathbf{V}_0\mathbf{Q} \end{pmatrix}\right).$$

Applying the calculated matrix for $\mathbf{P}$ and $\mathbf{Q}$, from (3.64), we obtain

$$\mathbf{P}\mathbf{V}_0\mathbf{P} = \mathbf{P}$$

and

$$\mathbf{PV}_0\mathbf{Q} = \begin{pmatrix} pvq_1 & pvq_2 & pvq_3 & 0 & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \\ pvq_1 & pvq_2 & pvq_3 & 0 & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \\ pvq_1 & pvq_2 & pvq_3 & 0 & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \\ 0 & 0 & 0 & pvq_4 & \mathbf{0} \\ \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 2} \end{pmatrix}_{13\times 6},$$

where for $i = 1, 2, 3$ the elements are

$$pvq_i = -\frac{A(\kappa_{02i}) + \kappa_{02i}a(\kappa_{02i})}{2\overline{a(\kappa_0)}}$$

and for $i = 4$, we have

$$pvq_4 = \frac{A(\kappa_{031}) + \kappa_{031}a(\kappa_{031})}{2p_{31}a(\kappa_{031})}.$$

The matrix $\mathbf{Q}^T\mathbf{V}_0\mathbf{Q}$ is a 6×6 matrix similar to Figure 3.3 with non zero elements in the red regions. There are 5 matrices in total. The middle one has only one element as

$$q_{44} = \frac{(\kappa_{031}a(\kappa_{031}) + A(\kappa_{031}))^2}{4p_{31}a(\kappa_{031})}.$$

The first square matrix is a 3×3 matrix with elements, for $i = 1, 2, 3$ and $j = 1, 2, 3$,

$$q_{ii} = \{(A(\kappa_{02i}) + \kappa_{02i}a(\kappa_{02i}))^2 + \frac{\kappa_{02i}^2 a(\kappa_{02i})\left(\overline{a(\kappa_0)} - p_{2i}\kappa_{02i}\right)}{p_{2i}}\}/(4\overline{a(\kappa_0)}),$$
$$q_{ji} = q_{ij} = \{A(\kappa_{02i})A(\kappa_{02j}) + A(\kappa_{02i})\kappa_{02j}a(\kappa_{02j}) + A(\kappa_{02j})\kappa_{02i}a(\kappa_{02i})\}/(4\overline{a(\kappa_0)}).$$

The elements in the last 2×2 square are

$$q_{55} = \frac{a(\kappa_{021})\left(p_{22}a(\kappa_{022}) + p_{23}a(\kappa_{023})\right)}{p_{21}\overline{a(\kappa_0)}},$$
$$q_{65} = q_{56} = \frac{p_{23}a(\kappa_{021})a(\kappa_{023})}{p_{22}\overline{a(\kappa_0)}},$$
$$q_{66} = \frac{p_{23}a(\kappa_{023})\left(p_{21}a(\kappa_{021}) + p_{22}a(\kappa_{022})\right)}{p_{22}^2\overline{a(\kappa_0)}}.$$

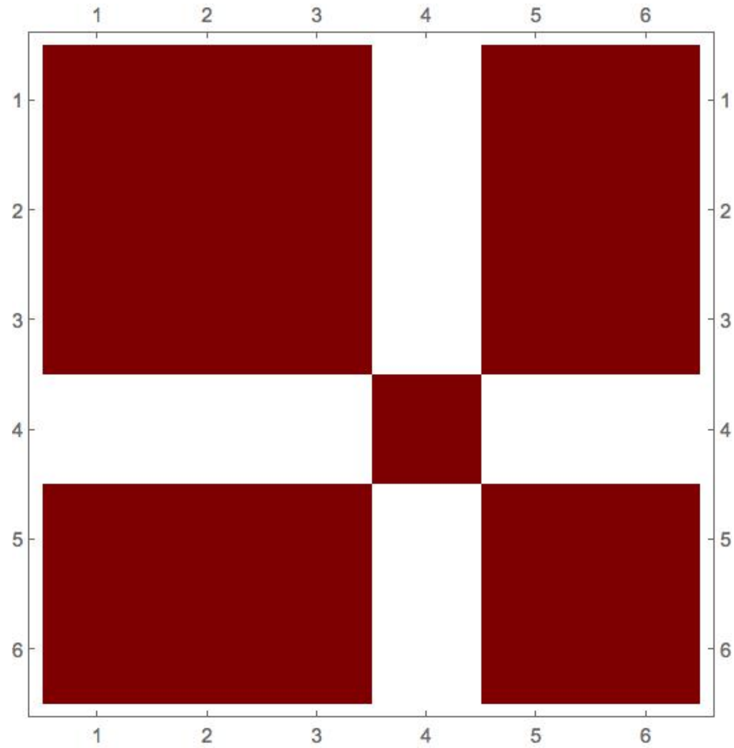

FIGURE 3.3: Schematic matrix plot for $\mathbf{Q}^T\mathbf{V}_0\mathbf{Q}$

The elements of the 3×2 and 2×3 rectangles are

$$
\begin{aligned}
q_{51} = q_{15} &= \frac{\kappa_{021} a(\kappa_{021}) \left(p_{22} a(\kappa_{022}) + p_{23} a(\kappa_{023})\right)}{2 p_{21} \overline{a(\kappa_0)}}, \\
q_{25} = q_{52} &= -\frac{\kappa_{022} a(\kappa_{021}) a(\kappa_{022})}{2 \overline{a(\kappa_0)}}, \\
q_{53} = q_{35} &= -\frac{\kappa_{023} a(\kappa_{021}) a(\kappa_{023})}{2 \overline{a(\kappa_0)}}, \\
q_{61} = q_{16} &= \frac{p_{23} \kappa_{021} a(\kappa_{021}) a(\kappa_{023})}{2 p_{22} a(\kappa_0)}, \\
q_{62} = q_{26} &= \frac{p_{23} \kappa_{022} a(\kappa_{022}) a(\kappa_{023})}{2 p_{22} a(\kappa_0)}, \\
q_{63} = q_{36} &= -\frac{\kappa_{023} a(\kappa_{023}) \left(p_{22} a(\kappa_{022}) + p_{21} a(\kappa_{021})\right)}{2 p_{21} \overline{a(\kappa_0)}}.
\end{aligned}
$$

Theorem 3.10. *The asymptotic distribution of d_n for testing the hypothesis $H_0 : \kappa_{21} = \kappa_{22} = \kappa_{23}$ in a model containing three 2 dimensional and one 3 dimensional von Mises Fisher distribution is chi square with 2 degrees of freedom.*

Proof of Theorem 3.10. The result follows from Theorem 3.7. □

3.5 Example 2: A Spherical Model in 10 Dimensions

Consider

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=2}^{10} (f_i(\mathbf{x}))^{I_{\mathbf{x}}(S_i)} \tag{3.70}$$

where for $i = 2, \ldots, 10$,

$$S_i = \{\mathbf{x} : \mathbf{x} \in \mathbb{S}^d \quad \text{with } d - i \text{ elements zero}\}$$

and

$$I_{\mathbf{x}}(S_i) = \begin{cases} 1 & ; \mathbf{x} \in S_i \\ 0 & ; \mathbf{x} \notin S_i. \end{cases}$$

The function $f_i(\mathbf{x})$ for $i = 2, \ldots, 10$ is an i dimensional von Mises Fisher distribution with the density

$$f_i(\mathbf{x}) = \{c^{(i)}(\kappa^{(i)})\}^{-1} \exp\{\kappa^{(i)} \boldsymbol{\mu}_i^T \mathbf{x}\}, \quad \mathbf{x} \in \mathbb{S}^i, ||\boldsymbol{\mu}_i|| = 1, \kappa^{(i)} > 0.$$

We are going to test the hypotheses

$$\begin{cases} H_0 & : \kappa^{(2)} = \kappa^{(3)} = \ldots = \kappa^{(10)} \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{3.71}$$

where $\kappa^{(i)}$ is the concentration parameter of the i dimensional von Mises Fisher distribution.

Theorem 3.11. *The asymptotic distribution of the deviance statistic for the hypothesis test* (3.71) *is chi square with 8 degrees of freedom.*

Proof of the Theorem 3.11. The proof of this theorem follows from Theorem 3.8.

□

Chapter 4

Data Analysis

4.1 Introduction

We analyze asset allocations from different indices in 3 dimensions, considering S&P500, NIKKEI225 and AllOrdinaries and model them as outlined in Chapter 2. From the Yahoo Finance website, we selected daily historical returns from the three indices in Table 4.1 and provided some Mathematica code in order to match the selected returns against given dates. Then we were able to calculate the covariance matrices of monthly returns and ultimately calculate the portfolios via maximizing the Sharpe ratio. Tables B.1 to B.9 contain the square roots (to fit them on a sphere) of the monthly asset allocations related to the three indices.

We also extended the method to higher dimensions by considering the 10 different indices listed in Table 4.2; we matched the dates and calculated the monthly covariance matrices. Tables B.10 to B.16 contain the square roots of the monthly asset allocations calculated for the ten indices introduced in 4.2 based on the daily returns available on the Yahoo Finance website.

As discussed in Chapter 2, we calculated portfolios with no short sales (NSS) by maximizing the Sharpe ratio

$$\frac{\mathbf{x}^T\boldsymbol{\mu}}{\sqrt{\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x}}}, \tag{4.1}$$

under the two constraints

$$\sum_{i=1}^{d} x_i = 1, \quad and \quad x_i \geq 0; \quad i = 1, 2, \cdots, d.$$

TABLE 4.1: Three different indices to form a portfolio based on daily returns in Yahoo Finance website

Orders In Portfolios	**Name**	**Country**	**Abb.**	**Start Date**	**Number of Data**
1	SP500	U.S.	$^\wedge GSPC$	3/Jan/1950	16202
2	NIKKEI225	Japan	$^\wedge N225$	4/Jun/1984	7479
3	AllOrdinaries	Australia	$^\wedge AORD$	3/Aug/1984	7543

TABLE 4.2: Ten different indices to form a portfolio based on daily returns in Yahoo Finance website

Orders In Portfolios	**Name**	**Country**	**Abb.**	**Start Date**	**Number of Data**
1	DAX	Germany	$^\wedge GDAXI$	26/Nov/1990	5951
2	FTSE100	U.K.	$^\wedge FTSE$	3/Jan/1984	7927
3	HSI	China	$^\wedge HSI$	31/Dec/1986	6816
4	BOND 20^+ Years	U.S.	TLT	1/Aug/2002	3001
5	NIKKEI225	Japan	$^\wedge N225$	4/Jun/1984	7479
6	SP/ASX 200	Australia	$^\wedge AXJO$	23/Nov/1992	5448
7	SP500	U.S.	$^\wedge GSPC$	3/Jan/1950	16202
8	Volatility SP500	China	$^\wedge VIX$	2/Jan/1990	6149
9	AllOrdinaries	Australia	$^\wedge AORD$	3/Aug/1984	7543
10	CAC40	France	$^\wedge FCHI$	1/Mar/1990	6139

There are different methods to estimate the monthly covariance matrix in (4.1). The most common method is by calculating the general covariance matrix through its usual formula

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})^T, \tag{4.2}$$

where n is the number of available reported daily returns in a specific month. With high dimensional data the covariance matrix can be poorly estimated, so we considered some ways of reducing the number of parameters. In particular, we investigated the use of Equicorrelation, Dynamic Equicorrelation and Block Equicorrelation models to estimate the covariance matrices. These methods are described in Section 4.2.

Calculating a monthly portfolio yields a vector of nonnegative proportions which sum to one. Taking the square root of each element of the vector, transforms the vector to lie on a sphere and we can use the von Mises Fisher distribution to analyse the data. We analyze the data by means of hypothesis tests and study the behaviour of the MEs by plotting them as time series. The following code, written in Mathematica, produces monthly portfolios by means of the sample covariance matrix in (4.2):

```
vs = {}; vs1 = {}; vs2 = {}; sDate = {}; spbondall = {}; vs3 = {}; \
Sigmas = {}; mus = {};
Do[
  Do[
   mSP = {}; dSP = {};
   mni = {}; dni = {};
   mal = {}; dal = {};
   Do[
    If[SP[[i, 1]] == y && SP[[i, 2]] == m,
     AppendTo[mSP, {SP[[i, 3]], SP[[i, 5]]}]]
    , {i, 1, 15000}];
   Do[
    If[NIK[[k, 1]] == y && NIK[[k, 2]] == m,
     AppendTo[mni, {NIK[[k, 3]], NIK[[k, 5]]}]]
    , {k, 1, 7479}];
   Do[
    If[ALLO[[k, 1]] == y && ALLO[[k, 2]] == m,
     AppendTo[mal, {ALLO[[k, 3]], ALLO[[k, 5]]}]]
    , {k, 1, 7543}];
   mRet = {};
   Do[
    Do[
     If[mSP[[i, 1]] == j, dSP = mSP[[i, 2]]], {i, 1, Length[mSP]}];
    Do[
     If[mni[[i, 1]] == j, dni = mni[[i, 2]]], {i, 1, Length[mni]}];
    Do[
     If[mal[[i, 1]] == j, dal = mal[[i, 2]]], {i, 1, Length[mal]}];
    If[NumberQ[dSP] && NumberQ[dni] && NumberQ[dal],
     AppendTo[mRet, {dSP, dni, dal}]]; dSP = {}; dni = {}; dal = {}
    , {j, 1, 31}];
   If[mRet != {},
    AppendTo[spbondall, mRet];
    mus = Mean[mRet];
    Sigmas = Covariance[mRet];
    v = {x1, x2, x3};
    vs = FindArgMax[{v.mus/Sqrt[v.Sigmas.v],
```

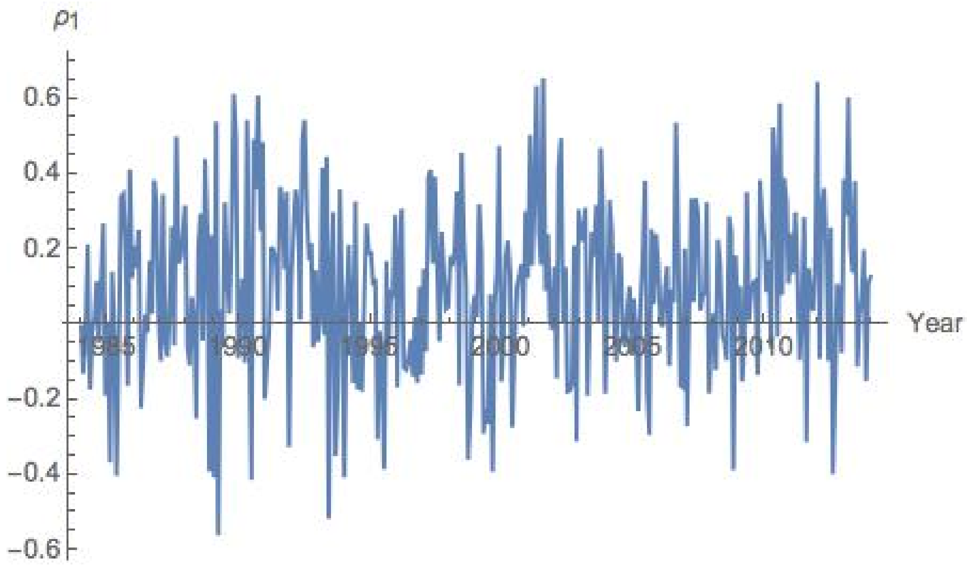

FIGURE 4.1: Monthly correlations between SP500 and NIKKEI225

```
      x1 + x2 + x3 == 1 && x1 > 0 && x2 > 0 && x3 > 0}, {x1, x2, x3}];
   AppendTo[vs3, vs];
   AppendTo[vs2, Point[Sqrt[vs]]];
   AppendTo[vs1, Sqrt[vs]];

   If[vs != {}, AppendTo[sDate, Flatten[{y, m, Sqrt[vs]}]]];
   vs = {}; Sigmas = {}; mus = {}]
  , {m, 1, 12}]
 , {y, 1984, 2014}];
```

4.2 The Study of Correlation

Figures 4.1, 4.2 and 4.3 show the monthly correlations between the indices in Table 4.1. Define ρ_1 to be the monthly return correlation between SP500 and NIKKEI225, ρ_2 to be the monthly return correlation between SP500 and AllOrdinaries and ρ_3 to be the monthly return correlation between NIKKEI225 and AllOrdinaries. In order to simplify the analysis we pose some questions:

1) Does an equicorrelation model fit this data?

2) Does a dynamic equicorrelation model fit the data?

3) Does a block dynamic equicorrelation model fit the data?

We will answer these questions in the following sections.

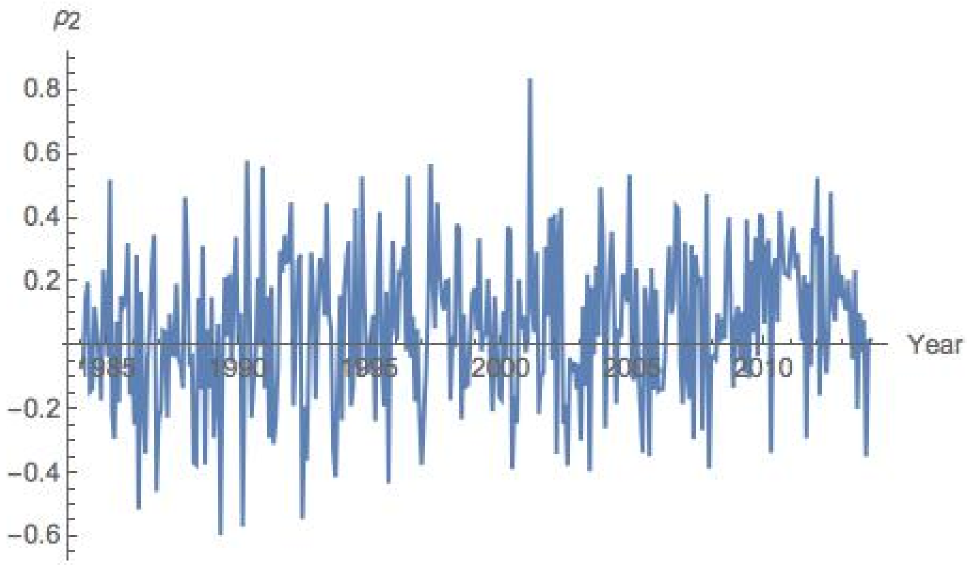

FIGURE 4.2: Monthly correlations between SP500 and AllOrdinaries

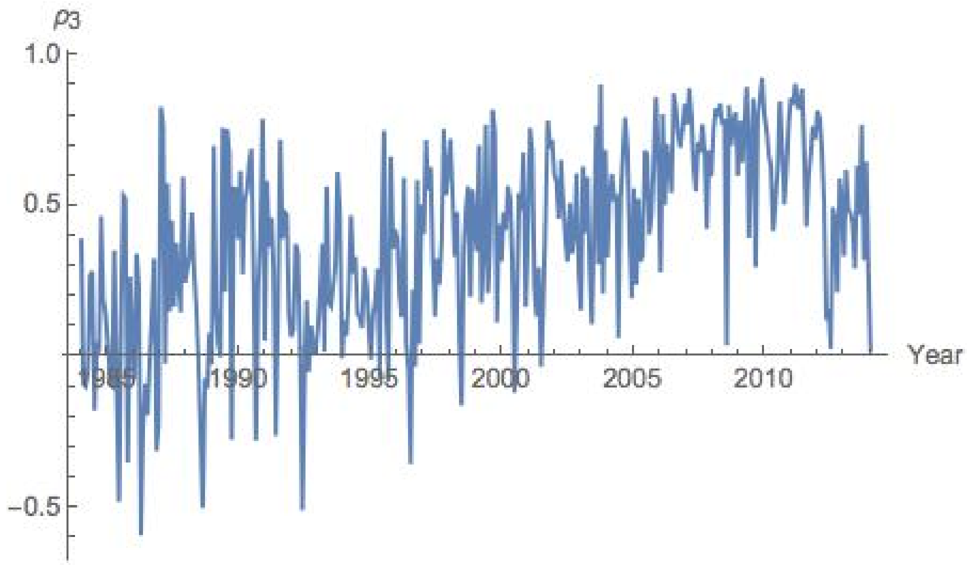

FIGURE 4.3: Monthly correlations between NIKKEI225 and AllOrdinaries

4.2.1 Equicorrelation

The equicorrelation method was introduced by Engle (2002) and has had substantial impact on the modelling of financial time series. In this method, we replace the off-diagonal elements of the correlation matrix with the average of the correlations, that replaces ρ_i, for $i = 1, \ldots, d$, in a correlation matrix by $\frac{\rho_1+\rho_2+\ldots+\rho_d}{d}$, where d is the number of assets. The process is repeated each month. Figure 4.4 shows the average $\frac{\rho_1+\rho_2+\rho_3}{3}$ over time for the three mentioned indices in Table 4.1. There appears to be an upward trend in these values which suggests a lack of stationarity in the original series.

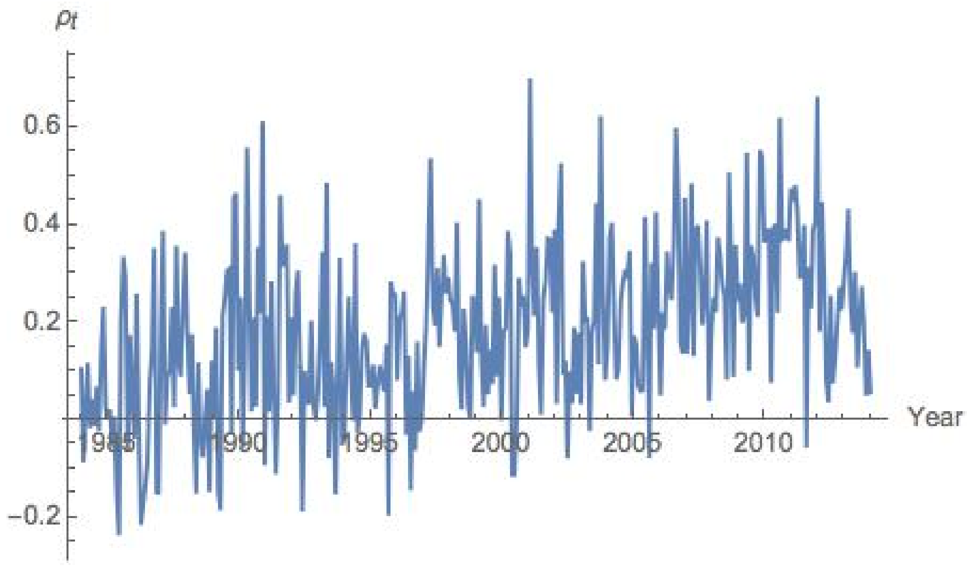

FIGURE 4.4: Monthly $\rho_t = \frac{\rho_1+\rho_2+\rho_3}{3}$ for the data in Tables B.1 to B.9

4.2.2 Dynamic Equicorrelation

The dynamic equicorrelation method was introduced by Engle and Kelly (2012) as a refinement of the equicorrelation method. They assumed the correlation matrix to be of the form

$$\mathbf{R}_t = (1-\rho_t)\mathbf{I}_d + \rho_t \mathbf{J}_d,$$

where ρ_t is the equicorrelation. The model is based on the assumption that all pairwise correlations are equal in a given month; d is the number of assets in the analysis, $\mathbf{I}_d$ is the $d \times d$ identity matrix and $\mathbf{J}_d$ is the $d \times d$ constant matrix of ones. In order to have a positive definite covariance matrix, we assume

$$\frac{-1}{d-1} < \rho_t < 1. \tag{4.3}$$

Then all the eigenvalues of the correlation matrix are positive.

In their dynamic equicorrelation model, Engle and Kelly assume the monthly returns of the assets have a multivariate normal distribution with a mean vector $\boldsymbol{\mu}$ and a covariance matrix with the structure

$$\mathbf{\Sigma}_t^* = Diagonal(\sigma_1, \sigma_2, \cdots, \sigma_d)\left((1-\rho_t)\mathbf{I}_d + \rho_t \mathbf{J}_d\right) Diagonal(\sigma_1, \sigma_2, \cdots, \sigma_d).$$

The log likelihood function is

$$\mathcal{L}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}^*| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{*-1}(\mathbf{x}_i - \boldsymbol{\mu}), \tag{4.4}$$

where

$$|\boldsymbol{\Sigma}^*| = \sigma_1^2\sigma_2^2\cdots\sigma_d^2(1-\rho)^{d-1}(1+(d-1)\rho).$$

The matrix $\boldsymbol{\Sigma}^{*-1}$ exists under constraint (4.3) and is calculated as

$$\boldsymbol{\Sigma}^{*-1} = Diagonal\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \cdots, \frac{1}{\sigma_d}\right)\left(\frac{1}{(1-\rho_t)}\mathbf{I}_d + \frac{\rho_t}{(1-\rho)(1+(d-1)\rho)}\mathbf{J}_d\right) \\ \times Diagonal\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \cdots, \frac{1}{\sigma_d}\right).$$

Maximizing (4.4) under (4.3) gives us the MLEs of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}^*$ which can be computed numerically in Mathematica. We use these MLEs to calculate NSS portfolios comprised of the three indices in Table 4.1 using the formula (4.1). We do not provide the spherical data calculated by this method of evaluating the covariance matrix in here and the data presented in Tables B.1 to B.9 are calculated from the sample covariance matrix. The following code is related to this part:

```
mu3 = {mu1, mu2, mu3};
Sigma3 =
  DiagonalMatrix[{a1, a2,
     a3}].((1 - rho)*IdentityMatrix[3] + rho*
      ConstantArray[1, {3, 3}]).DiagonalMatrix[{a1, a2, a3}];
mle = {}; ml3 = {}; vmle = {}; vmle1 = {}; DateEqui = {}; spnikalleqi \
= {}; vmle2 = {}; Sigmamle = {}; mRet = {}; mus = {};
Do[
  Do[
   mSP = {}; dSP = {};
   mni = {}; dni = {};
   mal = {}; dal = {};
   Do[
    If[SP[[i, 1]] == y && SP[[i, 2]] == m,
     AppendTo[mSP, {SP[[i, 3]], SP[[i, 5]]}]]
    , {i, 1, 15000}];
   Do[
    If[NIK[[k, 1]] == y && NIK[[k, 2]] == m,
```

```
    AppendTo[mni, {NIK[[k, 3]], NIK[[k, 5]]}]]
   , {k, 1, 7479}];
  Do[
   If[ALLO[[k, 1]] == y && ALLO[[k, 2]] == m,
    AppendTo[mal, {ALLO[[k, 3]], ALLO[[k, 5]]}]]
   , {k, 1, 7543}];
  mRet = {};
  Do[
   Do[
    If[mSP[[i, 1]] == j, dSP = mSP[[i, 2]]], {i, 1, Length[mSP]}];
   Do[
    If[mni[[i, 1]] == j, dni = mni[[i, 2]]], {i, 1, Length[mni]}];
   Do[
    If[mal[[i, 1]] == j, dal = mal[[i, 2]]], {i, 1, Length[mal]}];
   If[NumberQ[dSP] && NumberQ[dni] && NumberQ[dal],
    AppendTo[mRet, {dSP, dni, dal}]]; dSP = {}; dni = {}; dal = {}
   , {j, 1, 31}];
  If[mRet != {},
   AppendTo[spnikalleqi, mRet];

   lmle = -Length[mRet]/2*Log[Det[Sigma3]] -
     1/2 Sum[(mRet[[i]] -
          Mean[mRet]).Inverse[Sigma3].(mRet[[i]] -
          Mean[mRet]), {i, 1, Length[mRet]}];
   mle =
    FindArgMax[{lmle, rho > -1/2 && rho < 1 && a1 > 0 &&
       a2 > 0 && a3 > 0}, {{rho, 0.1}, a1, a2, a3}];
   AppendTo[ml3, mle];
   Sigmamle =
    DiagonalMatrix[{mle[[2]], mle[[3]],
       mle[[4]]}].((1 - mle[[1]])*IdentityMatrix[3] +
       mle[[1]]*ConstantArray[1, {3, 3}]).DiagonalMatrix[{mle[[2]],
       mle[[3]], mle[[4]]}];
   v = {x1, x2, x3};
   mus = Mean[mRet];
   vmle =
    FindArgMax[{v.mus/Sqrt[v.Sigmamle.v],
      Total[v] == 1 && x1 > 0 && x2 > 0 && x3 > 0}, {x1, x2, x3}];
   AppendTo[vmle1, Sqrt[vmle]];
```

```
    AppendTo[vmle2, Point[Sqrt[vmle]]];


    If[vmle != {}, AppendTo[DateEqui, Flatten[{y, m, Sqrt[vmle]}]]];
    vmle = {}; Sigmamle = {}; mle = {}; mus = {}]
   , {m, 1, 12}]
  , {y, 1984, 2014}];
```

4.2.3 Dynamic Equicorrelation vs General Model in 3 Dimensions

With ρ_1 the correlation between SP500 and NIKKEI225, ρ_2 the correlation between SP500 and AllOrdinaries and ρ_3 the correlation between NIKKEI225 and AllOrdinaries in a specific month, we test the hypotheses

$$\begin{cases} H_0 & : \rho_1 = \rho_2 = \rho_3 \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{4.5}$$

in each month. Under the dynamic equicorrelation assumption, the correlation matrix is considered to be of the form

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \tag{4.6}$$

where ρ satisfies (4.3). In this case, $\mathcal{L}_n(\theta)$ is

$$\mathcal{L}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}), \tag{4.7}$$

which has the supremum

$$\sup_{\theta\in\Theta}\mathcal{L}_n(\theta) = \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log|S| - \frac{n}{2}. \tag{4.8}$$

Here

$$\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in (-\infty, +\infty)^d, \text{and } \boldsymbol{\Sigma} \text{ is a positive definite matrix in } d \times d\}, \tag{4.9}$$

and Ω, the parameter space under the null hypothesis (4.5), is

$$\Omega = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in (-\infty, +\infty)^d, \boldsymbol{\Sigma} \text{ is a dynamic equicorrelation } d \times d \text{ matrix satisfying (4.3)}\}.$$

Therefore,

$$\sup_{\theta\in\Omega}\mathcal{L}_n(\theta) = \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log|\widehat{\boldsymbol{\Sigma}}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^T\widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}),$$

where $\overline{\mathbf{x}}$ is the sample mean of the daily returns and $\widehat{\boldsymbol{\Sigma}}$ maximizes (4.7) under (4.3). Figure 4.5 shows the values of

$$d_n = -2(\sup_{\boldsymbol{\theta}\in\Omega}\mathcal{L}_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\Theta}\mathcal{L}_n(\boldsymbol{\theta}))$$

over time (monthly values from August 1984 to July 2014) for the hypothesis (4.5) and the data are the daily returns of the indices in Table 4.1. The maximum value of d_n is 64.796 for August 2007 and the minimum is 22.98 for July 2014. The asymptotic distribution of d_n under the null hypothesis (4.5) is chi square with 2 degrees of freedom for which the calculated values in Figure 4.5 have the minimum 22.98. This leads to the rejection of H_0 at the 0.05 level. This result confirms that there exists adequate evidence to reject a dynamic equicorrelation model for our data. The following program calculated the relevant d_n in Figure 4.5:

```
dnrho3 = {};
Do[
 Sigmamle =
  DiagonalMatrix[{ml3[[j, 2]], ml3[[j, 3]],
     ml3[[j, 4]]}].((1 - ml3[[j, 1]])*IdentityMatrix[3] +
     ml3[[j, 1]]*ConstantArray[1, {3, 3}]).DiagonalMatrix[{ml3[[j,
       2]], ml3[[j, 3]], ml3[[j, 4]]}];
 dn3 = -2*(((-3*Length[spnikalleqi[[j]]])/2*Log[2*Pi] -
       Length[spnikalleqi[[j]]]/2*Log[Det[Sigmamle]] -
       1/2 Sum[(spnikalleqi[[j, i]] -
            Mean[spnikalleqi[[
              j]]]).Inverse[Sigmamle].(spnikalleqi[[j, i]] -
             Mean[spnikalleqi[[j]]]), {i, 1,
          Length[spnikalleqi[[j]]]}]) -
     ((-3*Length[spnikalleqi[[j]]])/2*Log[2*Pi] -
       Length[spnikalleqi[[j]]]/2*
        Log[Det[Covariance[spnikalleqi[[j]]]]] -
       Length[spnikalleqi[[j]]]/2));
 AppendTo[dnrho3, {j, dn3}]
 , {j, 1, 358}]
 ListLinePlot[dnrho3[[All, 2]], DataRange -> {1984, 2014},
```

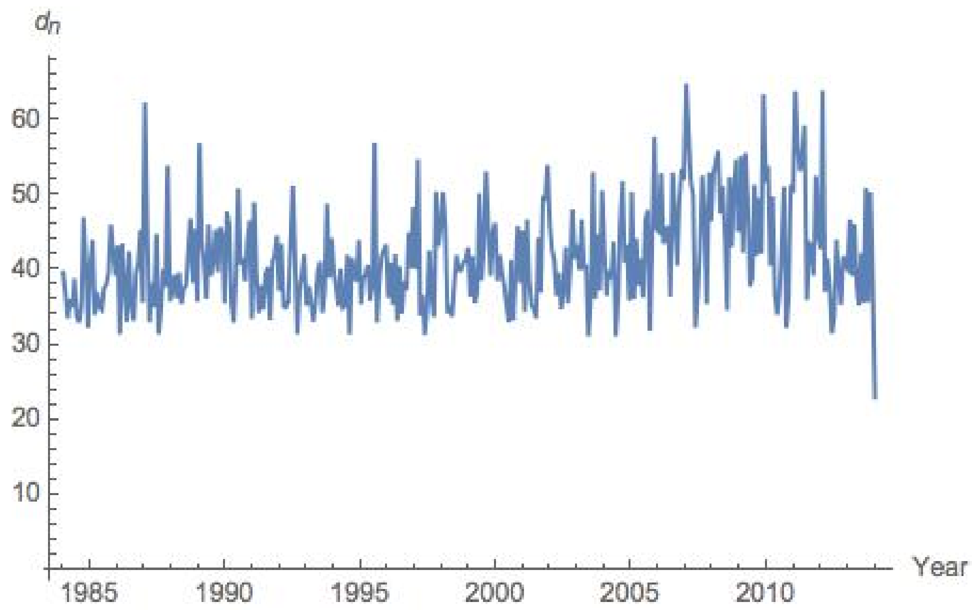

FIGURE 4.5: d_n for the hypothesis test $H_0 : \rho_1 = \rho_2 = \rho_3$ and the data are the daily returns of the indices in Table 4.1

```
AxesLabel -> {Year, Subscript[d, n]}]
```

4.2.4 Block Equicorrelation Model in 3 Dimensions

In the previous section, the null hypothesis of equality of the pairwise correlations between the three indices was rejected. In this section, we investigate further the pairwise correlations individually. We begin by comparing ρ_1 which is the correlation between SP500 and NIKKEI225 and ρ_2 which is the correlation between SP500 and AllOrdinaries. If we consider the correlation matrix as

$$\mathbf{R}_t = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}, \tag{4.10}$$

we can test the hypothesis

$$\begin{cases} H_0 & : \rho_1 = \rho_2 \\ H_A & : \rho_1 \neq \rho_2 \end{cases} \quad \text{or} \quad \begin{cases} H_0 & : \rho_{(SP500,NIKKEI225)} = \rho_{(SP500,AllOrdinaries)} \\ H_A & : \rho_{(SP500,NIKKEI225)} \neq \rho_{(SP500,AllOrdinaries)} \end{cases} \tag{4.11}$$

to decide whether or not the correlation between (SP500, NIKKEI225) is the same as the correlation between (SP500, AllOrdinaries). To test the hypotheses (4.11) the log likelihood function is considered to be (4.7) and Θ to be (4.9), so the supremum of the

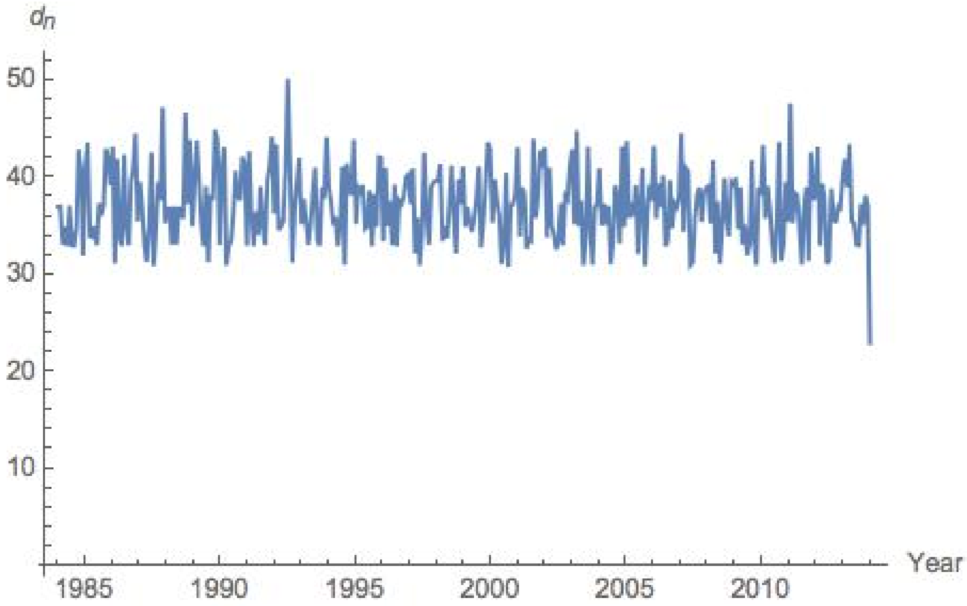

FIGURE 4.6: d_n for the hypotheses (4.11) and the data are the daily returns of the indices in Table 4.1

log likelihood function on Θ is (4.8). Under the null hypothesis, $\boldsymbol{\mu}_\Omega$ is in $(-\infty, +\infty)^d$ and $\boldsymbol{\Sigma}_\Omega$ is a positive definite matrix such that its correlation matrix is as follows:

$$\mathbf{R}_{BE} = \begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_3 \\ \rho_1 & \rho_3 & 1 \end{pmatrix}. \tag{4.12}$$

The eigenvalues of this correlation matrix are $1 - \rho_3$, $\frac{2+\rho_3-\sqrt{8\rho_1^2+\rho_3^2}}{2}$ and $\frac{2+\rho_3+\sqrt{8\rho_1^2+\rho_3^2}}{2}$. We maximize (4.7) under the constraint that these eigenvalues are positive. Figure 4.6 shows the monthly calculated d_n for these hypotheses. The maximum of d_n is 50.1518 for March 1993 and the minimum is 22.9563 for July 2014. Since the distribution of d_n under the null hypothesis (4.11) is chi square with 3 degrees of freedom, we conclude that the data furnishes strong evidence that this type of block equicorrelation model does not fit the data.

4.2.4.1 Block Equicorrelation - Second Case

Assume the correlation matrix for the data in Table 4.1 to be in the form of (4.10). To test the hypotheses

$$\begin{cases} H_0 & : \rho_{(NIKKEI225,SP500)} = \rho_{(NIKKEI225,AllOrdinaries)} \\ H_A & : \rho_{(NIKKEI225,SP500)} \neq \rho_{(NIKKEI225,AllOrdinaries)}, \end{cases} \tag{4.13}$$

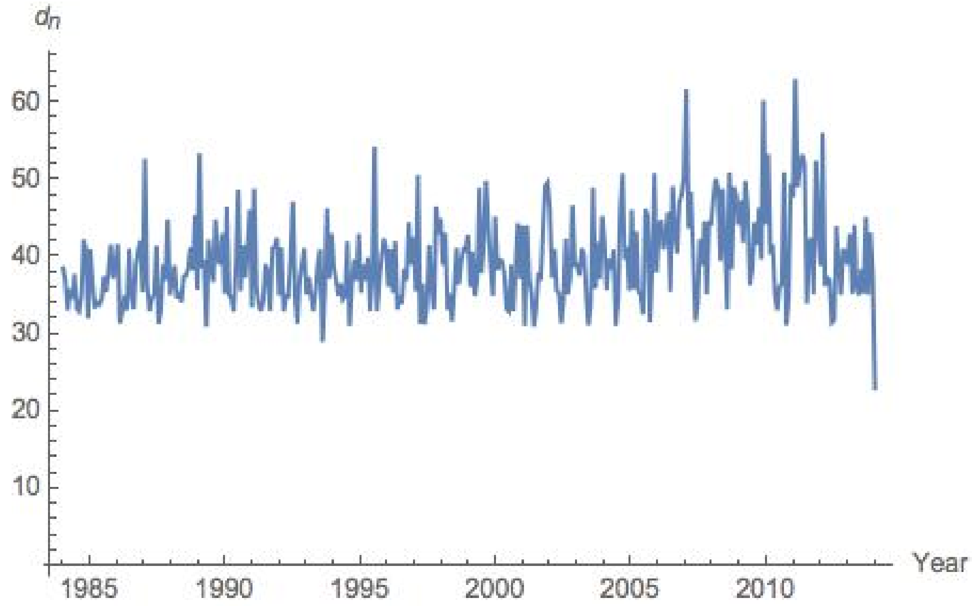

FIGURE 4.7: d_n for the hypotheses (4.13) and the data are the daily returns of the indices in Table 4.1

which is

$$\begin{cases} H_0 & : \rho_1 = \rho_3 \\ H_A & : \rho_1 \neq \rho_3, \end{cases}$$

the parameter space is (4.9). Under the null hypothesis, the correlation matrix is of the form

$$\mathbf{R}_{BE2} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}. \tag{4.14}$$

The eigenvalues of this matrix are

$$\begin{aligned} 1 - \rho_2&, \\ \frac{2 + \rho_2 + \sqrt{\rho_2^2 + 8\rho_1^2}}{2}&, \\ \frac{2 + \rho_2 - \sqrt{\rho_2^2 + 8\rho_1^2}}{2}&. \end{aligned} \tag{4.15}$$

To calculate the supremum of $\mathcal{L}_n(\boldsymbol{\theta})$ under the null hypothesis, we find the supremum of $\mathcal{L}_n(\boldsymbol{\theta})$ under the constraint that the eigenvalues in (4.15) are all positive.

Figure 4.7 shows the monthly values of d_n. The maximum of d_n is 63.0165 for August 2011 and the minimum of d_n is 22.951 for July 2014. So, it is obvious that the null hypothesis of (4.13) is rejected at the level of 0.05 and the block equicorrelation model does not fit the data in Tables B.1 to B.9.

4.2.4.2 Block Equicorrelation - Third Case

To test the hypotheses

$$\begin{cases} H_0 & : \rho_{(AllOrdinaries,SP)} = \rho_{(AllOrdinaries,NOKKEI225)} \\ H_A & : \rho_{(AllOrdinaries,SP)} \neq \rho_{(AllOrdinaries,NOKKEI225)}, \end{cases} \tag{4.16}$$

which is

$$\begin{cases} H_0 & : \rho_2 = \rho_3 \\ H_A & : \rho_2 \neq \rho_3, \end{cases}$$

the correlation matrix is

$$\mathbf{R}_{BE3} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_2 & 1 \end{pmatrix}. \tag{4.17}$$

The eigenvalues of this matrix are

$$\begin{gathered} 1-\rho_1, \\ \frac{2+\rho_1-\sqrt{\rho_1^2+8\rho_2^2}}{2}, \\ \frac{2+\rho_1+\sqrt{\rho_1^2+8\rho_2^2}}{2}. \end{gathered} \tag{4.18}$$

To find $\sup_{\boldsymbol{\theta}\in\Omega} \mathcal{L}_n(\boldsymbol{\theta})$, we maximize (4.7) under the constraint that the eigenvalues of (4.17) given in (4.18) are positive. Figure 4.8 shows different values of d_n over the period. The maximum is 58.9938 for August 2012 and the minimum is 22.8784 for July 2014. So, we reject the hypothesis that the block equicorrelation model fits our data.

4.2.5 Summary for 3 Dimensions

As we see from Figures 4.5 to 4.8, the dynamic and block equicorrelation models do not fit the data in Table 4.1. Therefore, we are unable to use these models to estimate the covariance matrix and thereby to calculate the monthly portfolios. So, we fall back on using the sample covariance matrix or the equicorrelation model to estimate the covariance matrix.

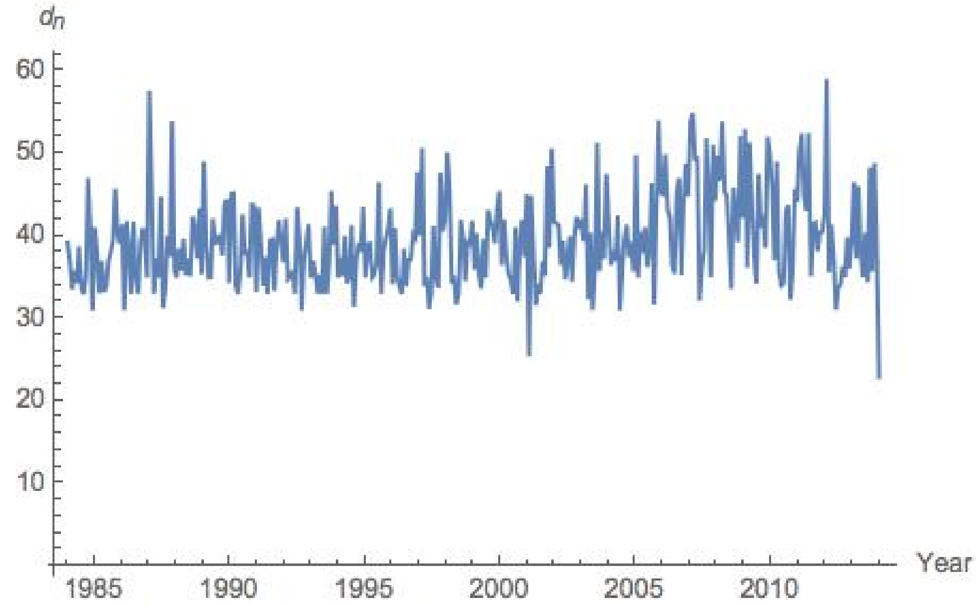

FIGURE 4.8: d_n for the hypotheses (4.16) and the data are the daily returns of the indices in Table 4.1

4.3 Investigating the Data in 3 Dimensions

We calculated NSS portfolios for the indices in Table 4.1 based on their daily returns and saw that they can be classified into seven categories. The square root of the monthly asset allocations of these three indices are in Tables B.1 to B.9. The first category is when all the elements in the allocation vector are greater than zero and the capital will be allocated to all three indices. In categories where one element equals zero, investment on this index is not indicated and the whole investment will be on the remaining two assets. In cases where two elements are zero, the relevant indices will be eliminated and the whole investment will be on the remaining asset.

The algorithm in Chapter 3 can be organised as follows:

$$
\begin{aligned}
S_1 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_3 = 0, x_1, x_2 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
S_2 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_2 = 0, x_1, x_3 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
S_3 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_1 = 0, x_2, x_3 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
S_4 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_1, x_2, x_3 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
Corner1 : S_5 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_3 = 0, x_2 = 0, x_1 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
Corner2 : S_6 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_3 = 0, x_1 = 0, x_2 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}, \\
Corner3 : S_7 &= \{\mathbf{x} = (x_1, x_2, x_3) : x_1 = 0, x_2 = 0, x_3 \neq 0, x_1^2 + x_2^2 + x_3^2 = 1\}.
\end{aligned}
\tag{4.19}
$$

Consequently, we specify the data for each category, split them and analyse them individually. We eliminate the corner points from our analysis, since in this case the whole investment would be only on one asset and there is no diversification effect. Such portfolios would not be used in practice.

We consider a 3 dimensional von Mises Fisher distribution for each category and test the hypothesis whether or not a particular component of $\mathbf{x}$ is significantly different from zero. This tells us whether that component is present in the optimum portfolio to a significant degree.

Consider

$$f_i(\mathbf{x}) = \frac{\exp\{\kappa_i(\mu_{i1}x_{i1} + \mu_{i2}x_{i2} + \mu_{i3}x_{i3})\}}{c^{(3)}(\kappa_i)}; \quad i = 1, 2, \cdots, 7,$$

where

$$c^{(3)}(\kappa) = \frac{4\pi \sinh(\kappa)}{\kappa}; \quad \kappa > 0$$

is the normalizing constant of the von Mises Fisher distribution, $x_{i1}^2 + x_{i2}^2 + x_{i3}^2 = 1$ and $\mu_{i1}^2 + \mu_{i2}^2 + \mu_{i3}^2 = 1$. The log likelihood function when we draw n_i independent samples individually from each category is

$$\mathcal{L}_n(\theta) = -n_i \log(c^{(3)}(\kappa_i)) + n_i\kappa_i(\mu_{i1}\overline{x}_{i1} + \mu_{i2}\overline{x}_{i2} + \mu_{i3}\overline{x}_{i3}); \quad i = 1, 2, \cdots, 7.$$

Table 4.3 shows the results of the separate hypothesis tests

$$\begin{cases} H_0 & : \mu_{13} = 0 \\ H_A & : \mu_{13} \neq 0 \end{cases}, \quad \begin{cases} H_0 & : \mu_{22} = 0 \\ H_A & : \mu_{22} \neq 0 \end{cases}, \quad \begin{cases} H_0 & : \mu_{31} = 0 \\ H_A & : \mu_{31} \neq 0, \end{cases} \tag{4.20}$$

which are related to the categories S_1, S_2 and S_3. The asymptotic distributions of d_n for testing (4.20) are chi square with one degree of freedom and since all the values of d_n in Table 4.3 are very small, the data does support the null hypotheses (the theory) in (4.20) and they are not rejected at the 0.05 level.

Graph 4.9 illustrates the data, showing the seven mentioned categories. Our analysis shows that 185 portfolios fall in the middle (region S_4), 39 in S_1, 71 in S_2, 17 in S_3, 20 in S_5, 11 in S_6 and 15 in S_7. This gives us the idea of using the model proposed in Chapter 3 to analyze the data which we will do in Section 4.6.

Figure 4.10 shows the monthly asset allocations calculated from the three indices in Table 4.1 when the equicorrelation method is used to calculate the covariance matrix. Using the equicorrelation model, 30 portfolios fall in region S_1, 52 in region S_2, 19 in

TABLE 4.3: Results of the hypothesis tests in (4.20)

Region	Null Hypothesis	Model of Covariance	n	d_n Value
S_1	$\mu_{13} = 0$	Sample	39	9.56×10^{-5}
S_2	$\mu_{22} = 0$	Sample	71	1.75×10^{-4}
S_3	$\mu_{31} = 0$	Sample	17	4.94×10^{-5}
S_1	$\mu_{13} = 0$	Equicorrelation	30	9.564×10^{-5}
S_2	$\mu_{22} = 0$	Equicorrelation	52	1.747×10^{-4}
S_3	$\mu_{31} = 0$	Equicorrelation	19	4.940×10^{-5}

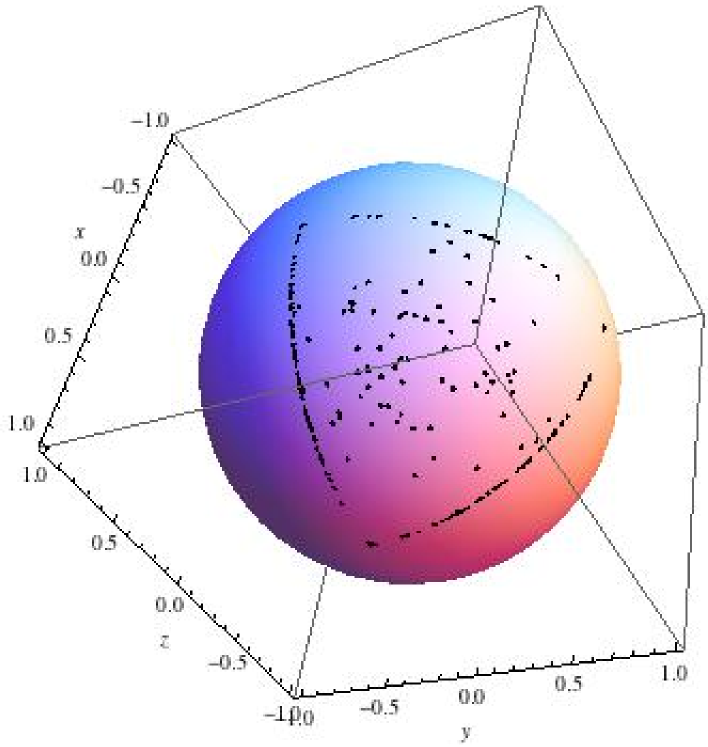

FIGURE 4.9: Monthly asset allocations transformed into spherical data when sample covariance matrix is used to estimate the covariance matrix for the data set presented in Tables B.1 to B.9

region S_3, 198 in the middle S_4, 17 in S_5, 18 in S_6 and 24 in region S_7. Table 4.3 shows the results of hypothesis tests based on the equicorrelation model.

4.4 Goodness of Fit Tests

To validate our analyses we want to test whether the data come from a von Mises Fisher distribution or not. For this reason, we use two different methodologies, one given by Watson (1983) and the other one by Mardia, Holmes and Kent (1984).

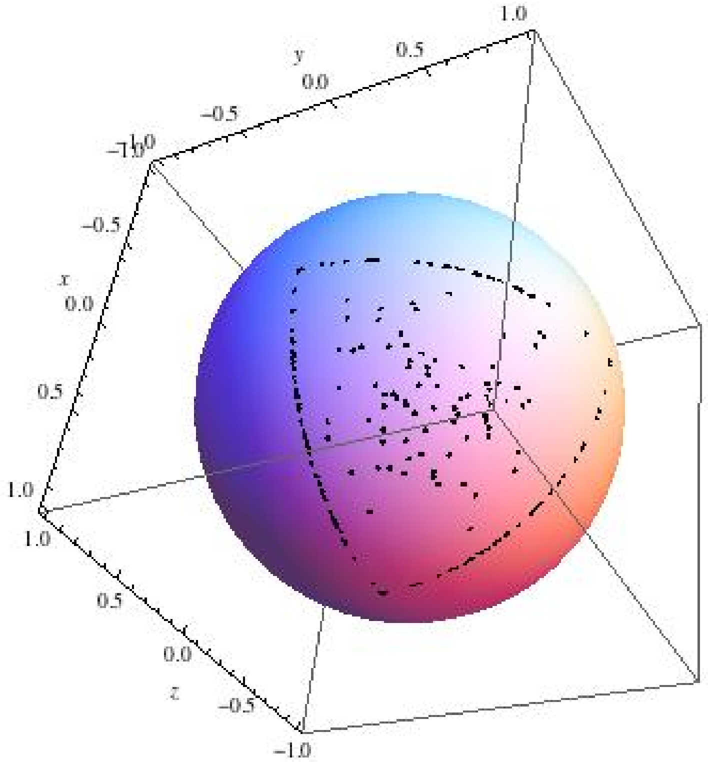

FIGURE 4.10: Monthly asset allocations transformed into spherical data when equicorrelation model is used to estimate the covariance matrix of the daily returns for the indices in Table 4.1

4.4.1 Watson's Method for Goodness of Fit Test for the von Mises Fisher Distribution

Watson (1983) introduced some methods to test the goodness of fit of von Mises Fisher distributions for the 3 and higher dimensional cases. These methods are presented in Section 4.4.1.2. The U^2 test introduced by Watson in his earlier paper in 1961 is applicable to data on a circle, a 2 dimensional case and is explained in Section 4.4.1.1 of this chapter.

4.4.1.1 Goodness of Fit Test in 2 dimensions

To test whether the random sample $X_1, X_2, \ldots, X_n$ is drawn from a specified continuous distribution function $F(x)$, Watson (1961) introduced the following statistic which measures the difference between $F(x)$ and the empirical distribution function $F_n(x)$

$$U_n^2 = n \int_{-\infty}^{\infty} \left(F_n(x) - F(x) - \int_{-\infty}^{\infty} (F_n(y) - F(y))\, dF(y) \right)^2 dF(x).$$

$F_n(x)$ is the empirical distribution function and it is

$$F_n(x) = \frac{\sum_{i=1}^{n} I_{(X_i \leq x)}(x)}{n},$$

where $I_A(x)$ is the indicator function and it is one when $x \in A$ and zero when $x \notin A$. He proved that the limiting distribution of U_n^2 is

$$P(U_n^2 > u) = \sum_{m=1}^{\infty} 2(-1)^{m-1} \exp\{-2m^2\pi^2 u\}$$

and provided simpler formulae for the U_n^2 based on $\nu_i = F(x_{(i)})$ ($x_{(i)}$ is the order statistic and the subscript (i) shows the i-th order statistic of the sample) which are

$$U_n^2 = \sum_{i=1}^{n} \left(\nu_i - \frac{2i-1}{2n} - \overline{\nu} + \frac{1}{2}\right)^2 + \frac{1}{12n}$$

and

$$U_n^2 = \sum_{i=1}^{n} \nu_i^2 - 2\sum_{i=1}^{n} \frac{2i-1}{2n}\nu_i + \frac{n}{3} + n\left(\overline{\nu} - \frac{1}{2}\right)^2.$$

We use these formulae in our calculations.

The 2 dimensional goodness of fit test for the von Mises Fisher distribution is relevant when we categorise the observations in Tables B.1 to B.9 into seven regions and test whether or not the 2 dimensional data in regions S_1, S_2 and S_3 have von Mises Fisher distributions. The results of the goodness of fit tests for these regions are shown in Table 4.4. Based on this result, the data in region S_3 has the 2 dimensional von Mises Fisher distribution at 0.01 level however, for the other two regions S_1 and S_2 the hypothesis that the data come from a 2 dimensional von Mises Fisher distribution is rejected.

TABLE 4.4: Results of the goodness of fit tests for 2 dimensional data in Tables B.4, B.5 and B.6 belong to the regions (4.19)

Region	n	$\widehat{\kappa}$	U_n^2	p-value
S_1	39	3.92458	1.12804	4.27×10^{-10}
S_2	71	3.51717	1.29227	1.67×10^{-11}
S_3	16	3.07728	0.237211	0.0185142

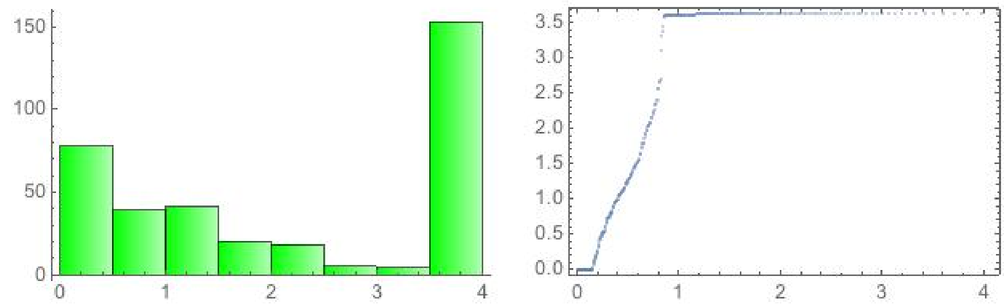

FIGURE 4.11: Histogram and quantile plot for a 3 dimensional goodness of fit test based on $\kappa(1-\cos\theta) \sim \chi^2(2)/2$ for the 358 data described in Tables B.1 to B.9; $r_q = 0.9095$

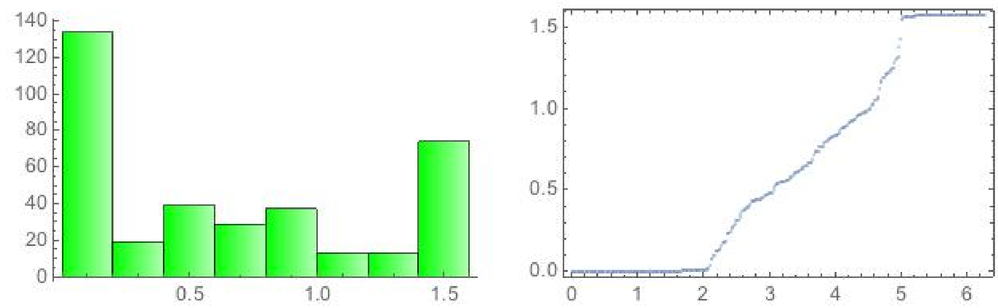

FIGURE 4.12: Histogram and quantile plot for a 3 dimensional goodness of fit test based on $\phi \sim U[0, 2\pi]$ for the 358 data described in Tables B.1 to B.9; $r_q = 0.97$

4.4.1.2 Goodness of Fit Test in 3 Dimensions

Watson (1983) in his paper suggests using $\kappa(1-\cos\theta)$ and ϕ, where θ and ϕ are the spherical polar coordinates of the data in 3 dimensions and shows that if some data actually come from a von Mises Fisher distribution, then $\kappa(1-\cos\theta)$ has an exponential distribution, even when we replace κ with its estimator $\widehat{\kappa}_n$ for the purpose of calculation and ϕ has a uniform distribution in $[0, 2\pi]$. Here we have $\theta = \arccos[Z]$ and $\phi = \arctan(\frac{Y}{X})$ as the random variables based on the random vector (X, Y, Z) from a 3 dimensional von Mises Fisher distribution.

For the observations in Tables B.1 to B.9, the results of the above goodness of fit tests for the entire 358 observations and the partial 185 interior observations falling in the region S_4, are shown in Figures 4.11 to 4.14. For the quantile plots in these figures, the correlation between quantile values of the data and the fitted distributions are presented to judge the behaviour of the quantile plots better. These correlations are shown by r_q in the caption of figures.

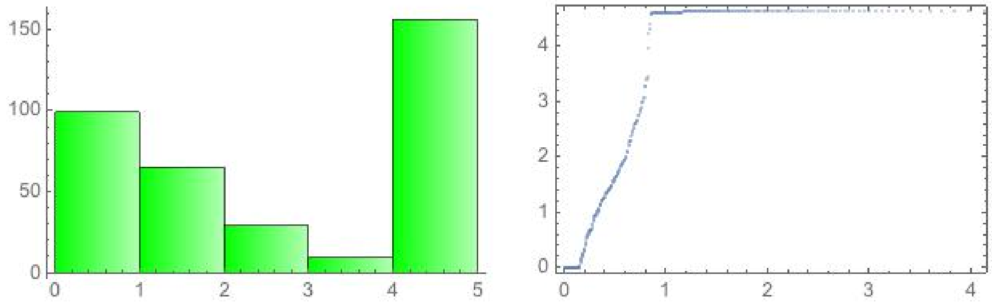

FIGURE 4.13: Histogram and quantile plot for a 3 dimensional goodness of fit test based on $\kappa(1-\cos\theta) \sim \chi^2(2)/2$ for the 185 data falling in the region S_4 described in Tables B.1 to B.3; $r_q = 0.9485$

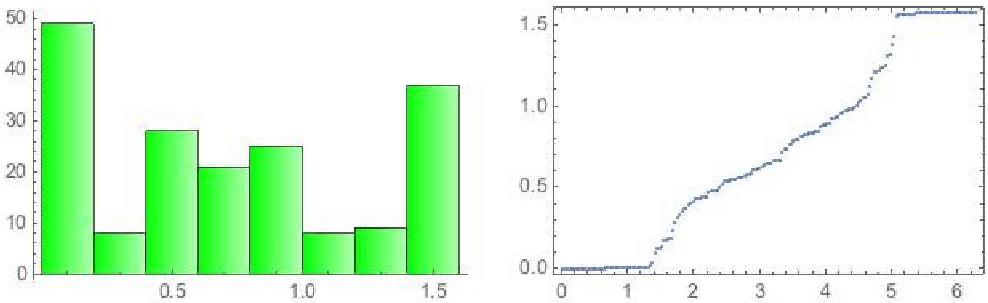

FIGURE 4.14: Histogram and quantile plot for a 3 dimensional goodness of fit test based on $\phi \sim U[0, 2\pi]$ for the 185 data falling in the region S_4 described in Tables B.1 to B.3; $r_q = 0.9857$

4.4.1.3 Goodness of Fit Test in Dimensions greater than 3

In higher dimensions, where $d > 3$, Watson (1983) considered the goodness of fit statistic

$$2\left(\kappa + \frac{d-3}{4}\right)(1 - \mathbf{x}^T\boldsymbol{\mu}) \sim \chi^2(d-1). \tag{4.21}$$

To fit the data in Tables B.10 to B.16 a 10 dimensional von Mises Fisher distribution, we have

$$\begin{aligned} \widehat{\kappa} &= 10.6071, \\ \widehat{\boldsymbol{\mu}} &= (0.20091, 0.124089, 0.241181, 0.668905, 0.177297, 0.0884374, 0.527516, \\ &\quad 0.25553, 0.217584, 0.092023)^T. \end{aligned} \tag{4.22}$$

Figure 4.15 shows the histogram and quantile plot of the statistic (4.21) which has a chi square distribution with 9 degrees of freedom for the data in Tables B.10 to B.16.

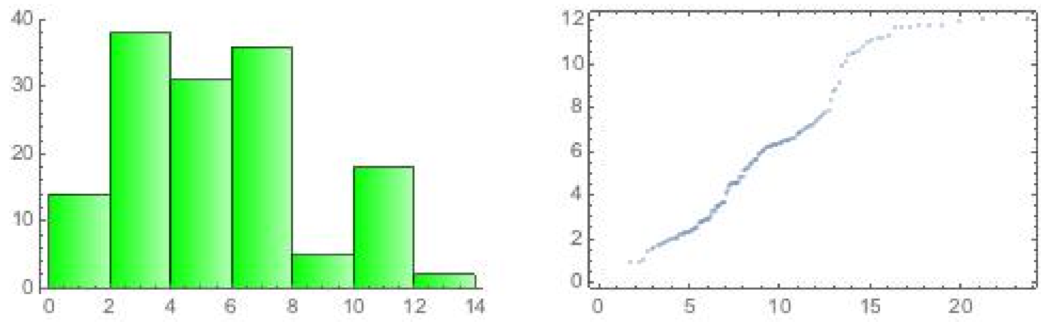

FIGURE 4.15: Histogram and quantile plot for the goodness of fit test for the 10 dimensional indices data described in Tables B.10 to B.16 based on (4.21); $r_q = 0.9857$

As can be seen from the analysis in this section, the von Mises Fisher distribution does not fit the data too well in some cases, and applying other spherical distributions is suggested in order to find an appropriate model. However, this contemplation goes beyond the scope of this thesis and we will postpone it until future work.

4.4.2 Mardia et al's Method for the Goodness of Fit Test for the von Mises Fisher Distribution

Mardia, Holmes and Kent (1984) consider Fisher Bingham distribution as an alternative to the von Mises Fisher distribution. The density function of the Fisher Bingham is

$$f_{\mathbf{X}}(\mathbf{x}) = C\exp\{\kappa\boldsymbol{\mu}^T\mathbf{x} + \sum_{i=2}^{d}\lambda_i\left(\boldsymbol{\gamma}_i^T\mathbf{x}\right)^2\}, \quad \mathbf{x}\in\mathbb{S}^d,$$

where C is the normalizing constant and

$$\kappa > 0, \quad \boldsymbol{\mu}\in\mathbb{S}^d,$$

are the parameters related to the von Mises Fisher distribution and the additional parameters $\lambda_2,\ldots,\lambda_d \in \mathbb{R}$ and a set of orthogonal vectors of parameters $\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_d$ are related to the Bingham distribution.

The Fisher Bingham distribution forms a canonical exponential family and can be written as

$$f_{\mathbf{X}}(\mathbf{x}) = C\exp\{\boldsymbol{\pi}_1^T\mathbf{u}^{(1)} + \boldsymbol{\pi}_2^T\mathbf{u}^{(2)}\}, \tag{4.23}$$

by utilizing the sufficient statistics $\mathbf{u}^T = (\mathbf{u}^{(1)T}, \mathbf{u}^{(2)T})$ and the parameters $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ where C is considered to be the normalizing constant. The statistic $\mathbf{u}^{(1)}$ is

$$\mathbf{u}^{(1)} = \mathbf{x} = (x_1, x_2, \ldots, x_d)^T,$$

and all the terms $x_i x_j$, for $i \leq j$, are considered to be in $\mathbf{u}^{(2)}$. We divide $\mathbf{u}^{(2)}$ in three parts as $\mathbf{u}^{(2)^T} = (\mathbf{v}^{(1)T}, \mathbf{v}^{(2)T}, \mathbf{v}^{(3)T})$, where

$$\begin{aligned}
\mathbf{v}^{(1)} &= (x_2^2, \ldots, x_d^2)^T, \\
\mathbf{v}^{(2)} &= (x_1 x_2, \ldots, x_1 x_d)^T, \\
\mathbf{v}^{(3)} &= (x_2 x_3, \ldots, x_{d-1} x_d)^T.
\end{aligned}$$

For the goodness of fit test

$$\begin{cases} H_0 & : \text{data come from the von Mises Fisher distribution} \\ H_A & : \text{data come from the Fisher Bingham distribution,} \end{cases}$$

it is suggested to test

$$\begin{cases} H_0 & : \boldsymbol{\pi}_2 = \mathbf{0} \\ H_A & : \boldsymbol{\pi}_2 \neq \mathbf{0}, \end{cases}$$

in the density function of the form (4.23).

Assume we rotate the coordinate system by an orthogonal matrix $\mathbf{H}$,

$$\mathbf{x}^* = \mathbf{H}^T \mathbf{x}, \tag{4.24}$$

to have the sample mean $\overline{\mathbf{x}}$ lie along the positive x_1 axis. We introduced the matrix $\mathbf{Q}$ in Chapter 2, equation (2.7) and in a special case of 3 dimensions in (2.9) in order to rotate $\boldsymbol{\mu}$ to $(1, 0, \ldots, 0)^T$. We can consider matrix $\mathbf{H}$ as $\mathbf{Q}$ when $\boldsymbol{\mu}$ is replaced by $\frac{\mathbf{x}}{||\mathbf{x}||}$.

Mardia, Holmes and Kent (1984) verified all of the non-zero moments of orders less than or equal to 4 from the von Mises Fisher distribution when the mean direction $\boldsymbol{\mu}$ lies along the positive x_1 axis. Let $I_\nu(z)$ be the modified Bessel function of the first kind and $\nu = \frac{1}{2}d - 1$. For simplicity, we consider $I_\nu = I_\nu(\kappa)$. Assume $2 \leq j < k \leq d$ and

$$f_{\mathbf{X}}(\mathbf{x}) = c(\kappa)^{-1} \exp\{\kappa x_1\}, \quad \kappa > 0, \quad \mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{S}^d,$$

where $c(\kappa)$ is obtained via (2.16) verified in Chapter 2. Under this distribution the moments are

$$\begin{aligned}
a_1 &= \mathrm{EX}_1 = \mathrm{I}_{\nu+1}/\mathrm{I}_\nu, \\
a_2 &= \mathrm{EX}_1^2 = \{\frac{\mathrm{d}-1}{\mathrm{d}}\mathrm{I}_{\nu+2} + \frac{1}{\mathrm{d}}\mathrm{I}_\nu\}/\mathrm{I}_\nu, \\
b_0 &= \mathrm{EX}_\mathrm{j}^2 = \kappa^{-1}\mathrm{I}_{\nu+1}/\mathrm{I}_\nu, \\
b_1 &= \mathrm{EX}_1\mathrm{X}_\mathrm{j}^2 = \kappa^{-1}\mathrm{I}_{\nu+2}/\mathrm{I}_\nu, \\
b_2 &= \mathrm{EX}_1^2\mathrm{X}_\mathrm{j}^2 = \kappa^{-1}\{\frac{\mathrm{d}+1}{\mathrm{d}+2}\mathrm{I}_{\nu+3} + \frac{1}{\mathrm{d}+2}\mathrm{I}_{\nu+1}\}/\mathrm{I}_\nu, \\
d_1 &= \mathrm{EX}_\mathrm{j}^2\mathrm{X}_\mathrm{k}^2 = \kappa^{-2}\mathrm{I}_{\nu+2}/\mathrm{I}_\nu, \\
d_2 &= \mathrm{EX}_\mathrm{j}^4 = 3\kappa^2\mathrm{I}_{\nu+2}/\mathrm{I}_\nu.
\end{aligned} \tag{4.25}$$

Therefore, the non-zero variances and covariances are

$$\begin{aligned}
\mathrm{Var}(\mathrm{X}_1) &= a_2 - a_1^2 = a_2^*, \\
\mathrm{Var}(\mathrm{X}_\mathrm{j}) &= b_0, \\
\mathrm{Var}(\mathrm{X}_\mathrm{j}^2) &= d_2 - b_0^2 = d_2^*, \\
\mathrm{Cov}(\mathrm{X}_\mathrm{j}^2, \mathrm{X}_\mathrm{k}^2) &= d_1 - b_0^2 = d_1^*, \\
\mathrm{Var}(\mathrm{X}_1\mathrm{X}_\mathrm{k}) &= b_2, \\
\mathrm{Var}(\mathrm{X}_\mathrm{j}\mathrm{X}_\mathrm{k}) &= d_1, \\
\mathrm{Cov}(\mathrm{X}_\mathrm{j}, \mathrm{X}_1\mathrm{X}_\mathrm{j}) &= b_1, \\
\mathrm{Cov}(\mathrm{X}_1, \mathrm{X}_\mathrm{j}^2) &= b_1 - a_1 b_0 = b_1^*.
\end{aligned} \tag{4.26}$$

For the null hypothesis $\boldsymbol{\pi}_2 = \mathbf{0}$ in (4.23), Mardia, Holmes and Kent (1984) consider score statistics. They assume a sample of n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from $\mathbf{X} = (X_1, \ldots, X_d)^T$. We consider the j-th sample, $1 \le j \le n$, to be $\mathbf{x}_j = (x_{1j}, \ldots, x_{dj})^T$ and denote the sample mean vector by

$$\overline{\mathbf{x}} = (\overline{x}_1, \ldots, \overline{x}_d)^T, \quad \overline{x}_i = \frac{\sum_{j=1}^n x_{ij}}{n},$$

for $i = 1, \ldots, d$ and the sample sum of squares and product matrix $\overline{\mathbf{T}} = (\bar{t}_{ij})$ by

$$\bar{t}_{ij} = \sum_{k=1}^n x_{ki} x_{kj}, \quad 1 \le i, j \le d.$$

The score statistics under the null hypothesis $H_0 : \boldsymbol{\pi}_2 = \mathbf{0}$ is

$$W_n = n\left(\overline{\mathbf{u}}^{(2)} - \widehat{\boldsymbol{\delta}}_{2.1}\right)^T \widehat{\boldsymbol{\Sigma}}_{22.1}^{-1}\left(\overline{\mathbf{u}}^{(2)} - \widehat{\boldsymbol{\delta}}_{2.1}\right)^T$$

which, under the null, has an asymptotic chi square distribution with $\frac{d(d+1)}{2} - 1$ degrees of freedom as $n \to \infty$. In W_n, $\boldsymbol{\delta}_{2.1}$ and $\boldsymbol{\Sigma}_{22.1}$ are the conditional mean vector and covariance matrix of $\overline{\mathbf{u}}^{(2)}$ given $\overline{\mathbf{u}}^{(1)}$ which depend on κ and $\boldsymbol{\mu}$ and can be computed via the moments in (4.26). The vector $\widehat{\boldsymbol{\delta}}_{2.1}$ and the matrix $\widehat{\boldsymbol{\Sigma}}_{22.1}$ are used to show that we estimate these quantities using the maximum likelihoods $\widehat{\kappa}$ and $\widehat{\boldsymbol{\mu}}$.

We can simplify W_n for the calculation purposes by using the formulae in (4.25). Assume

$$\overline{\mathbf{T}}^* = \mathbf{H}^T\overline{\mathbf{T}}\mathbf{H}, \tag{4.27}$$

where the matrix $\mathbf{H}$ is introduced in (4.24) and

$$\begin{aligned}
f_{11} &= d_2^* - \frac{b_1^{*2}}{a_2^*},\\
f_{12} &= d_1^* - \frac{b_1^{*2}}{a_2^*},\\
g_{11} &= \frac{f_{11} + (d-3)f_{12}}{2d_1(f_{11} + (d-2)f_{12})},\\
g_{12} &= \frac{-f_{12}}{2d_1(f_{11} + (d-2)f_{12})},\\
g_{21} &= (b_2 - \frac{b_1^2}{b_0})^{-1},\\
g_{31} &= d_1^{-1},
\end{aligned}$$

then we have

$$W_n = n\left(\widehat{g}_{11}\sum_{j=2}^{d}(\bar{t}_{jj}^* - \widehat{b}_0)^2 + 2\widehat{g}_{12}\sum_{2\le j<k}(\bar{t}_{jj}^* - \widehat{b}_0)(\bar{t}_{kk}^* - \widehat{b}_0) + \widehat{g}_{21}\sum_{j=2}^{d}(\bar{t}_{1j}^*)^2 + \widehat{g}_{31}\sum_{2\le j<k}(\bar{t}_{jk}^*)^2\right). \tag{4.28}$$

To see if the square roots of asset allocations from different dimensions have a von Mises Fisher distribution, we calculate W_n in (4.28) through the following coding:

```
tij = Transpose[vonMises].vonMises/Length[vonMises];
H = H0
ts = Transpose[H].tij.H;
n = n0;
```

```
p = d;
nu = p/2 - 1;
k = k0;
a1 = BesselI[nu + 1, k]/BesselI[nu, k];
a2 = (((p - 1)/p)*BesselI[nu + 2, k] + BesselI[nu, k]/p)/
  BesselI[nu, k];
b0 = BesselI[nu + 1, k]/(k*BesselI[nu, k]);
b1 = BesselI[nu + 2, k]/(k*BesselI[nu, k]);
b2 = (((p + 1)/(p + 2))*BesselI[nu + 3, k] +
   BesselI[nu + 1, k]/(p + 2))/(k*BesselI[nu, k]);
d1 = BesselI[nu + 2, k]/(k^2*BesselI[nu, k]);
d2 = (BesselI[nu + 2, k]*3)/(k^2*BesselI[nu, k]);
a2s = a2 - a1^2;
d2s = d2 - b0^2;
d1s = d1 - b0^2;
b1s = b1 - a1*b0;
f11 = d2s - b1s^2/a2s;
f12 = d1s - b1s^2/a2s;
g11 = (f11 + (p - 3)*f12)/(2*d1*(f11 + (p - 2)*f12));
g12 = -f12/(2*d1*(f11 + (p - 2)*f12));
g21 = (b2 - b1^2/b0)^(-1);
g31 = (d1)^(-1);
Wu = n*(g11*((ts[[2, 2]] - b0)^2 + (ts[[3, 3]] - b0)^2) +
    2*g12*(ts[[2, 2]] - b0)*(ts[[3, 3]] - b0) +
    g21*(ts[[1, 2]]^2 + ts[[1, 3]]^2) + g31*(ts[[2, 3]]^2))
```

In the 3 dimensions, the matrix **H** in the second line of the above coding can be chosen to be

```
mu1 = Mean[vonMises][[1]]/Norm[Mean[vonMises]];
mu2 = Mean[vonMises][[2]]/Norm[Mean[vonMises]];
mu3 = Mean[vonMises][[3]]/Norm[Mean[vonMises]];
H = {{mu1, mu3/Sqrt[mu1^2 + mu3^2], (-mu1*mu2)/
     Sqrt[mu1^2 + mu3^2]}, {mu2, 0,
     Sqrt[mu1^2 + mu3^2]}, {mu3, -mu1/
     Sqrt[mu1^2 + mu3^2], (-mu2*mu3)/
     Sqrt[mu1^2 + mu3^2]}};
```

and in the 2 dimensions, the matrix **H** can be considered as

```
mu1 = Mean[vonMises][[1]]/Norm[Mean[vonMises]];
mu2 = Mean[vonMises][[2]]/Norm[Mean[vonMises]];
H = N[{{mu1, mu2}, {mu2, -mu1}}];
```

in order to calculate the matrix $\overline{\mathbf{T}}^*$ in (4.27).

4.4.2.1 Application to the real data

We applied the data in Table 1, column (b) of $n = 34$ directions of magnetism in the Great Whin Sill collected by Creer, Irving and Nairn (1959) and calculated the value of 7.52 for W_n. This value for W_n equals to the one presented by Mardia, Holmes and Kent (1984), page 77. Therefore, we are sure that the above program is correct and can be applied for the square roots of the asset allocation data.

Table 4.5 shows the values of W_n for the 3 indices in Table 4.1 and their correspondent spherical data in Tables B.1 to B.9, categorised in 4 regions. As this table shows, the null hypothesis that the data are from a von Mises Fisher distribution is not rejected for the data located in the middle, the regions S_1 and S_3 at 0.05 level. For the region S_2, this null hypothesis is not rejected at 0.01 level.

TABLE 4.5: Results for the goodness of fit test based on the method presented by Mardia et al (1984) for the 3 dimensional spherical data related to the 3 indices in Table 4.1

Region	n	Dimension	W_n	$\chi^2_{0.05}$(d.f.)
Whole Region	358	3	76.505	$\chi^2_{0.05}(5) = 11.07$
Middle	185	3	5.76419	$\chi^2_{0.05}(5) = 11.07$
S_1	39	2	2.68197	$\chi^2_{0.05}(2) = 5.99$
S_2	71	2	6.44142	$\chi^2_{0.05}(2) = 5.99$
S_3	16	2	5.64648	$\chi^2_{0.05}(2) = 5.99$

4.5 Hypothesis Tests in 10 dimensions

Referring to the indices in Table 4.2, we calculated a portfolio for each month from August 1 2002 to July 2014; the resulting number of portfolios is 144. This gives us 144 10 dimensional spherical observations to be analysed. Tables B.10 to B.16 show the square roots of the asset allocations calculated from the ten indices in Table 4.2.

We assume that the data have a 10 dimensional von Mises Fisher distribution and aim to test some hypotheses for the parameters. To test the individual hypotheses

$$\begin{cases} H_0 & : \mu_i = 0 \\ H_A & : \mu_i \neq 0, \end{cases} \tag{4.29}$$

for $i = 1, 2, \cdots, 10$, we calculate the 10 deviance statistics d_n. The log likelihood function is a 10 dimensional spherical distribution

$$\mathcal{L}_n(\boldsymbol{\theta}) = -n \log c^{(10)}(\kappa) + n\kappa(\mu_1 \overline{x}_1 + \ldots + \mu_{10}\overline{x}_{10}).$$

Following the results from chapter 3, we find the asymptotic distribution of d_n for the hypotheses (4.29) is chi square with one degree of freedom. The results are included in Table 4.6.

The values that I chose for any of the hypotheses in Table 4.6 are based on the MLEs in (4.22) on page 116. As we see the null hypothesis that checks whether the mean parameter for Index Number 4, which is Bond 20^+ from the U.S., has the value 0.7, is not rejected at a 0.05 level. Another hypothesis concerns Index Number 7, which is the SP500, and the null hypothesis that tests whether the mean parameter for this asset is 0.5 is not rejected at the 0.05 level. The null hypothesis that tests whether the two mean parameters related to Bond 20^+ and SP500 are equal is rejected. The null hypothesis that suggests investing all of our budget in Index Number 4 is rejected too. The null hypothesis which tests the equality of the mean parameters for the FTSE100 from the U.K., the ASX200 from Australia and the CAC40 from France is not rejected. Also the equality of the mean parameters for the DAX from Germany, the HSI from China, the Volatility SP500 from China and the AllOrdinaries from Australia is not rejected. The null hypothesis that the FTSE100 from U.K. and the NIKKEI225 from Japan have the same mean parameters is not rejected either. The hypotheses that each index has the mean parameter zero are all rejected.

4.6 Spherical Subcomponent Model for the Portfolios

4.6.1 3 Dimensional Portfolios

Following the theory from Chapter 3, we are able to carry out the hypothesis test

$$\begin{cases} H_0 & : \kappa_1 = \kappa_2 = \kappa_3 \\ H_A & : \text{at least one equality is not satisfied,} \end{cases}$$

Table 4.6: Testing some hypotheses for the 10 dimensional data in Tables B.10 to B.16 when assuming the data to come from a von Mises Fisher distribution

H_0	$\sup_{\theta\in\Omega}\mathcal{L}_n(\theta)$	d_n	dis. of d_n
$\mu_1 = 0$	-107.241	40.3158	$\chi^2(1)$
$\mu_2 = 0$	-95.1807	16.1949	$\chi^2(1)$
$\mu_3 = 0$	-116.266	58.3649	$\chi^2(1)$
$\mu_4 = 0$	-274.359	374.552	$\chi^2(1)$
$\mu_5 = 0$	-102.969	31.7717	$\chi^2(1)$
$\mu_6 = 0$	-91.1664	8.1662	$\chi^2(1)$
$\mu_7 = 0$	-213.463	252.759	$\chi^2(1)$
$\mu_8 = 0$	-119.223	64.2791	$\chi^2(1)$
$\mu_9 = 0$	-111.734	49.3019	$\chi^2(1)$
$\mu_{10} = 0$	-91.4196	8.67266	$\chi^2(1)$
$\mu_6 = \mu_{10} = 0$	-95.4785	16.7905	$\chi^2(1)$
$\mu_2 = \mu_6 = \mu_{10} = 0$	-103.484	32.8006	$\chi^2(2)$
$\mu_2 = \mu_6 = \mu_{10}$	-87.5107	0.854884	$\chi^2(3)$
$\mu_1 = \mu_3 = \mu_8 = \mu_9$	-87.9284	1.69034	$\chi^2(4)$
$\mu_2 = \mu_5$	-87.7488	1.33117	$\chi^2(2)$
$\mu_1 = \mu_2 = \mu_3 =$ $\mu_5 = \mu_8 = \mu_9$	-92.6511	11.1357	$\chi^2(6)$
$\mu_4 = 0.70$	-88.6712	3.17585	$\chi^2(1)$
$\mu_7 = 0.5$	-87.6661	1.16561	$\chi^2(1)$
$\mu_4 = \mu_7$	-91.1475	8.12847	$\chi^2(1)$
$\mu_4 = 1$	-323.825	473.483	$\chi^2(1)$

for the data in Tables B.1 to B.9. The log likelihood function is (3.46), the whole parameter space is (3.47) and the null parameter space is (3.49). We found the value of d_n for the above hypothesis test for the spherical observations of Tables B.1 to B.9 to be 0.214807. The asymptotic distribution of d_n from Theorem 3.10 is chi square with 2 degrees of freedom. Therefore, we conclude that the observations do not provide evidence to reject this null hypothesis at the level of 0.01. We accept that the kappas of the two dimensional von Mises Fisher distributions of the spherical subcomponent model for the three mentioned indices are equal at the level of 0.01. The ME of the κ is calculated from the formula (3.68) as $\widehat{\kappa} = 3.55957$ with the standard deviation calculated from matrix (3.69) for $\{\overline{a(\kappa_0)}\}^{-1} = \{p_{21}a(\kappa_{021}) + p_{22}a(\kappa_{022}) + p_{23}a(\kappa_{023})\}^{-1}$ as 2.48693. These values are calculated from

$$\begin{array}{ccc} n_{21} = 39, & n_{22} = 71, & n_{23} = 17 \\ ||\overline{\mathbf{x}}_{21}|| = 0.86056, & ||\overline{\mathbf{x}}_{22}|| = 0.841996, & ||\overline{\mathbf{x}}_{23}|| = 0.81554, \\ \widehat{\kappa}_{21} = 3.9246, & \widehat{\kappa}_{21} = 3.51718, & \widehat{\kappa}_{21} = 3.07728. \end{array}$$

If we use the equicorrelation model to calculate the covariance matrix in compiling the monthly portfolios for the indices in Table 4.1, the value for d_n calculated as described above becomes $d_n = 0.399292$. Again this value is small, there is no change in the conclusion and we can say that the κs in the 2 dimensional spherical subcomponent model are all equal.

4.6.2 Spherical Subcomponent Model for the 10 Dimensional Portfolios

Consider the 10 dimensional spherical data based on the indices in Table 4.2 shown in Tables B.10 to B.16 and the estimating function

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\sum_{i=1}^{10} n_i \log c^{(i)}(\kappa_i) + \sum_{i=1}^{10} n_i \kappa_i \boldsymbol{\mu}_i^T \overline{\mathbf{x}}_i,$$

for the data. For the hypotheses

$$\begin{cases} H_0 & : \kappa^{(1)} = \kappa^{(2)} = \ldots = \kappa^{(10)} \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{4.30}$$

the value of d_n is 17.1851. Theorem 3.11 proves that the asymptotic distribution of d_n for the hypotheses (4.30) is chi square with 8 degrees of freedom. As the quantile values of the chi square distribution with 8 degrees of freedom are 15.5073 and 20.0902 at the levels of 0.05 and 0.01 respectively ($\chi^2_{0.95}(8) = 15.5073 < 17.1851 < 20.0902 = \chi^2_{0.99}(8)$), the null hypothesis is not rejected at the level of 0.01, however it is rejected at the level of 0.05.

If we consider the equicorrelation model, the value of d_n becomes 8.47509 which shows that the null hypothesis in (4.30) is not rejected at the level of 0.05 in this case. Based on this evidence it seems reasonable to conclude that all 10 κs are the same.

4.7 Precision of Estimation of Parameters

For the spherical subcomponent model (3.46), the maximum estimators are obtained from

$$\widehat{\boldsymbol{\mu}}_i = \frac{\overline{\mathbf{x}}_i}{||\overline{\mathbf{x}}_i||} \quad \text{and} \quad \frac{c_{\widehat{\kappa}_i}}{c(\widehat{\kappa}_i)} = ||\overline{\mathbf{x}}_i||$$

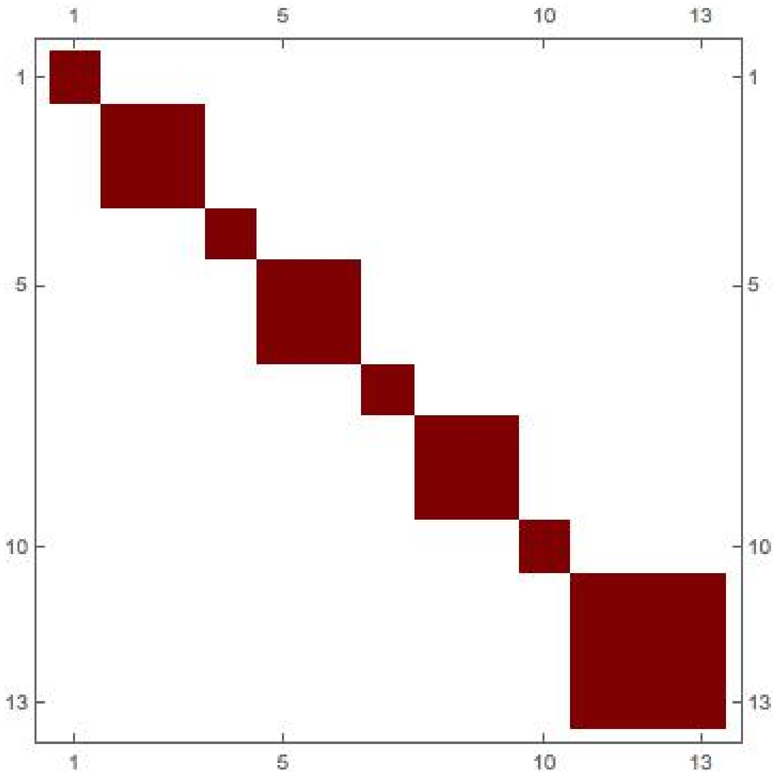

FIGURE 4.16: A schematic matrix plot for the variance matrix of the MEs for the spherical subcomponent model in (3.46)

for $i = 1, 2, 3, 4$. So, for the data in Table 4.1, we have

$$\begin{aligned}
&\widehat{\kappa}_4 = 4.6345, \quad \widehat{\mu}_{41} = 0.584492, \quad \widehat{\mu}_{42} = 0.545406, \quad \widehat{\mu}_{43} = 0.600751,\\
&\widehat{\kappa}_3 = 3.40978, \quad \widehat{\mu}_{31} = 0.890676, \quad \widehat{\mu}_{32} = 0.454638,\\
&\widehat{\kappa}_2 = 3.53859, \quad \widehat{\mu}_{21} = 0.753628, \quad \widehat{\mu}_{22} = 0.657302,\\
&\widehat{\kappa}_1 = 3.9245, \quad \widehat{\mu}_{11} = 0.608174, \quad \widehat{\mu}_{12} = 0.793804.
\end{aligned} \tag{4.31}$$

A schematic matrix plot for the variance matrix of the MEs is given by the matrix $\mathbf{P}$ in the format of Figure 4.16. In Figure 4.16 the white areas are considered to be all zeros. If we denote each pair of the small and large red squares by a matrix $\mathbf{P}_i$, for $i = 1, 2, 3, 4$, then they are

$$\mathbf{P}_i = \begin{pmatrix} \frac{1}{p_i a(\kappa_{0i})} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\kappa_{0i} A(\kappa_{0i})}(\mathbf{I} - \boldsymbol{\mu}_{0i}\boldsymbol{\mu}_{0i}^T) \end{pmatrix}.$$

For the calculated MEs in (4.31), the variance matrix is

$$
\mathrm{Var}(\widehat{\boldsymbol{\theta}}) = \frac{1}{39+71+16+185}
\begin{pmatrix}
198.53 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2.0647 & -1.5819 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1.5819 & 1.212 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 85.997 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.6343 & -0.7273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -0.7273 & 0.8338 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 208.568 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8381 & -1.642 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.642 & 3.2168 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 36.402 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3044 & -0.1474 & -0.1623 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1474 & 0.3248 & -0.1515 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.1623 & -0.1515 & 0.2955
\end{pmatrix}.
$$

Therefore, from the formula $\theta : \widehat{\theta} \pm 1.96\sqrt{\mathrm{Var}(\widehat{\theta})}$, we have

$$
\begin{aligned}
&\kappa_1 : 3.9245 \pm 1.5658, \quad \mu_{11} : 0.6081 \pm 0.1595, \quad \mu_{12} : 0.7938 \pm 0.1223 \\
&\kappa_2 : 3.5385 \pm 1.0305, \quad \mu_{21} : 0.7536 \pm 0.0886, \quad \mu_{22} : 0.6573 \pm 0.1013 \\
&\kappa_3 : 3.4097 \pm 1.6050, \quad \mu_{31} : 0.8906 \pm 0.1017, \quad \mu_{32} : 0.4546 \pm 0.1993 \\
&\kappa_4 : 4.6345 \pm 0.6705, \quad \mu_{41} : 0.5844 \pm 0.0611, \quad \mu_{42} : 0.5454 \pm 0.0633, \\
&\qquad\qquad \mu_{43} : 0.6007 \pm 0.0603.
\end{aligned}
$$

The 95 percent confidence intervals reported here can be used to estimate the allocations assigned to each index in the portfolio. The mean parameters $\widehat{\mu}_{ij}$ play this role. For example, from this result, we can conclude that the second 2 dimensional distribution has the smallest $\widehat{\kappa}$, so we choose this version. So, I choose to have $0.753628^2 \times 100 = 56.80$ percent of the capital allocated to SP500 and $0.657302^2 \times 100 = 43.20$ percent of the capital invested in AllOrdinaries to get the highest return for a given level of risk. Taking into account the confidence interval we can say:

If we invest a percentage between 66.5 and 70.93 of our capital in SP500 and the rest in AllOrdinaries, then we get the most profit from these 2 dimensional portfolios with 95 percent confidence.

Chapter 5

Simulation Results

5.1 Introduction

The aim of this chapter is to discuss both the simulation methods and the results of the asymptotic distribution of d_n obtained from the hypothesis testings conducted in the previous chapters.

5.2 Simulation from the von Mises Fisher Distribution

Ulrich (1984), inspired by Saw (1978), introduced a model in which many known spherical distributions were included. Ulrich argued that the model encompassed most known spherical distributions but since then recently introduced spherical distributions, such as Kent or Fisher Bingham distributions, have more parameters and this is no longer the case. Let $g(u)$ be a continuous non-negative and increasing function and for a positive d denote

$$c(\kappa) = \int_{-1}^{1} g(\kappa t) \frac{(1-t^2)^{\frac{d-3}{2}}}{B\left(\frac{1}{2}, \frac{1}{2}(d-1)\right)} dt,$$

where $\kappa \geq 0$ is considered to be the concentration parameter of the spherical distribution and $B\left(\frac{1}{2}, \frac{1}{2}(d-1)\right)$ is the beta function with parameters $1/2$ and $(d-1)/2$ that is equal to $\frac{\Gamma(1/2)\Gamma((d-1)/2)}{\Gamma(1/2+(d-1)/2)}$ for the gamma function as $\Gamma(\alpha)$. Let $\mathbf{X}$ be a random vector in $\mathbb{S}^d$. Saw (1978) defines the density of $\mathbf{X}$ to be

$$g(\mathbf{x}; \kappa, \boldsymbol{\mu}) = \frac{g(\kappa \boldsymbol{\mu}^T \mathbf{x})}{\alpha_d c(\kappa)}, \tag{5.1}$$

where $\boldsymbol{\mu}$ is the modal vector in $\mathbb{S}^d$ and $\alpha_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})}$. To get the density of the von Mises Fisher distribution, it is enough to substitute $g(u) = \exp\{u\}$ in (5.1). At this stage, Ulrich presents and proves the following theorem:

Ulrich Theorem: Let W be a random variable with the density

$$\frac{g(\kappa w)(1-w^2)^{\frac{d-3}{2}}}{c(\kappa)B\left(\frac{1}{2},\frac{d-1}{2}\right)}; \quad w \in (-1,1), \quad d \geq 2 \tag{5.2}$$

and let $\mathbf{V}$ be independent of W and have a $(d-1)$-dimensional uniform distribution on the sphere. Then the vector $\mathbf{X}$, where

$$\mathbf{X}^T = \left(\sqrt{1-W^2}\mathbf{V}^T, W\right), \tag{5.3}$$

has the density $g(\mathbf{X}; \kappa, \boldsymbol{\mu})$ from (5.1) with the modal vector $\boldsymbol{\mu}^T = (0,0,\ldots,0,1)$.

5.2.1 Rotational Matrix

Ulrich's theorem suggests a method of generating data with mean or modal vector $[0,0,\cdots,0,1]^T$. In Theorem 5.1 we define a rotational matrix which rotates the mean vector $[0,0,\cdots,0,1]^T$ to $\boldsymbol{\mu} = [\mu_1,\mu_2,\cdots,\mu_d]^T$ in the general case of d dimensions. We use this matrix in the simulations.

Theorem 5.1. *An orthogonal matrix based on*

$$\mathbf{B} = \begin{pmatrix} \mu_d\mathbf{J} & -\mathbf{b} \\ \mathbf{a}^T & \mu_d \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\mathbf{J}-\mathbf{b}\mathbf{a}^T}{\mu_d} & \mathbf{a} \\ -\mathbf{b}^T & \mu_d \end{pmatrix},$$

where

$$\mathbf{a} = [\mu_1,\cdots,\mu_{d-1}]^T,$$
$$\mathbf{b} = [\mu_{d-1},\cdots,\mu_1]^T,$$
$$\mathbf{J} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}_{(d-1)\times(d-1)}$$

rotates the mean vector $[0,0,\cdots,0,1]^T$ *to* $\boldsymbol{\mu} = [\mu_1,\mu_2,\cdots,\mu_d]^T$ *in a way that* $\boldsymbol{\mu} = \mathbf{B}[0,0,\cdots,0,1]^T$. *Matrices* $\mathbf{Q}^{(2)}$ *and* $\mathbf{Q}^{(3)}$ *presented in* (5.4) *and* (5.5) *show rotational matrices in 2 and 3 dimensions.*

For the 2 dimensional case, we use

$$\mathbf{Q}^{(2)} = \begin{pmatrix} \mu_2 & \mu_1 \\ -\mu_1 & \mu_2 \end{pmatrix} \tag{5.4}$$

to rotate a von Mises Fisher distribution with mean vector $[0,1]^T$ to $[\mu_1,\mu_2]^T$. The matrix

$$\mathbf{Q}^{(3)} = \begin{pmatrix} \frac{-\mu_3}{\sqrt{1-\mu_2^2}} & \frac{\mu_1\mu_2}{\sqrt{1-\mu_2^2}} & \mu_1 \\ 0 & -\sqrt{1-\mu_2^2} & \mu_2 \\ \frac{\mu_1}{\sqrt{1-\mu_2^2}} & \frac{\mu_2\mu_3}{\sqrt{1-\mu_2^2}} & \mu_3 \end{pmatrix} \tag{5.5}$$

rotates the mean vector $[0,0,1]^T$ to $[\mu_1,\mu_2,\mu_3]^T$ in 3 dimensions.

5.2.2 Simulation Method for Generating Data from the von Mises Fisher Distribution in 3 Dimensions

In the case $d = 3$, (5.2) becomes

$$\frac{2g(\kappa w)}{c(\kappa)}; \quad w \in (-1,1).$$

The cumulative distribution of W, when we consider a von Mises Fisher distribution, is

$$F_W(w) = \int_{-1}^{w} \frac{2g(\kappa t)}{c(\kappa)} dt = \int_{-1}^{w} \frac{2\exp\{\kappa t\}}{c(\kappa)} dt = \frac{2(\exp\{\kappa w\} - \exp\{-\kappa\})}{\kappa c(\kappa)}$$

and

$$c(\kappa) = \int_{-1}^{1} \frac{\exp\{\kappa t\}}{B\left(\frac{1}{2},1\right)} dt = \frac{2\left(\exp\{\kappa\} - \exp\{-\kappa\}\right)}{\kappa}$$

and hence

$$F_W(w) = \frac{\exp\{\kappa w\} - \exp\{-\kappa\}}{\exp\{\kappa\} - \exp\{-\kappa\}}.$$

To generate data from the distribution W, we first generate U from a uniform distribution and then make the following transformation

$$u = \frac{\exp\{\kappa w\} - \exp\{-\kappa\}}{\exp\{\kappa\} - \exp\{-\kappa\}},$$

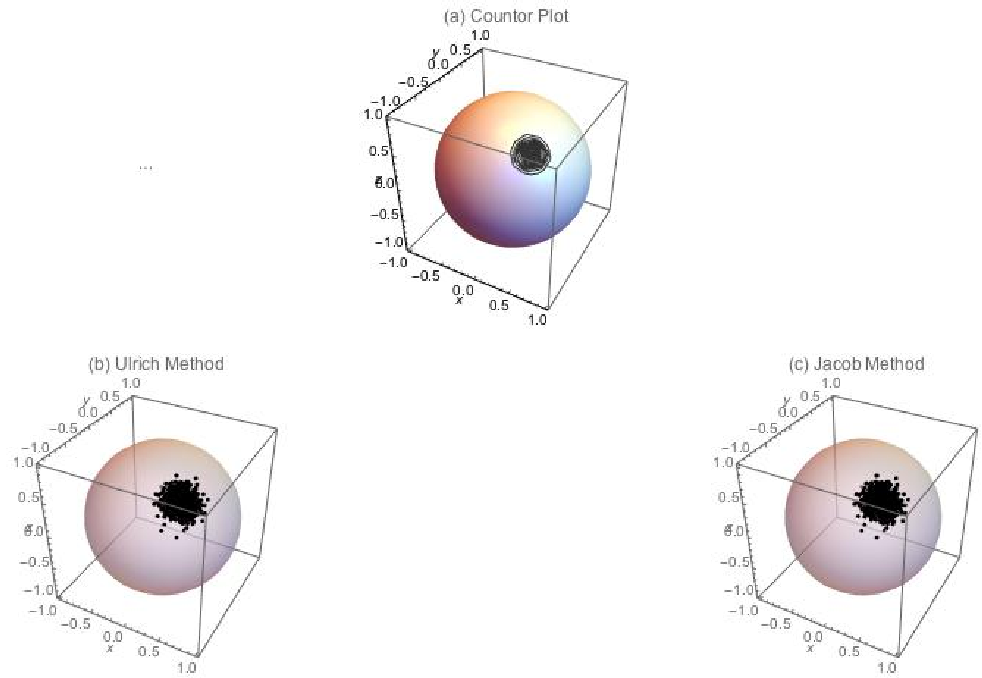

FIGURE 5.1: Simulating 1000 data from $VM(100, (0.5, -0.5, \sqrt{1 - 2 \times 0.5^2}))$ and comparing formulae (5.6) and (5.7) with the related contour plot: (a) the contour plot of the distribution, (b) the simulated data obtained from the formula (5.6) and (c) the simulated data obtained from the formula (5.7)

which gives

$$w = \frac{\log\left(\exp\{-\kappa\} + 2u \sinh \kappa\right)}{\kappa}. \tag{5.6}$$

Jakob (2012) suggests that it is advisable to use an equivalent form of (5.6) which is

$$w = 1 + \frac{\log u + \log\left(1 - \frac{u-1}{u} \exp\{-2\kappa\}\right)}{\kappa}. \tag{5.7}$$

This is a numerically well-behaved expression that prevents overflow for large values of κ. Figure 5.1 shows a von Mises Fisher distribution with mean vector $(0.5, -0.5, 0.7071)$ and the concentration parameter $\kappa = 100$ simulated in both ways. As the figure shows there are no significant differences between these two methods even for a large κ such as 100. For the simulation method, we use the code in the next section, while using the normal method to generate uniform data on the sphere.

5.2.3 Simulation Method for Generating Data from the von Mises Fisher Distribution in dimensions other than 3

For dimensions other than three, Ulrich uses an acceptance-rejection algorithm to generate data from the density $f(x)$. To implement this, we consider an envelope density, say $e(x)$, and calculate

$$M = \max_x \frac{f(x)}{e(x)} < \infty.$$

It is supposed that generating data from the envelope density $e(x)$ is routine and can be done easily. Then we generate X from $e(x)$ and independently U from a uniform distribution and if the condition $U \leq \frac{f(X)}{Me(X)}$ is satisfied, then we accept X as data from the density $f(x)$. Ulrich uses the envelope density

$$e(x; b) = \frac{2b^{(d-1)/2}}{B(\frac{d-1}{2}, \frac{d-1}{2})} \frac{(1-x^2)^{(d-3)/2}}{[(1+b)-(1-b)x]^{d-1}}; \quad x \in (-1, 1), b \in (0, 1), \tag{5.8}$$

to generate data from (5.2). It is easy to generate data from $e(x)$, because if $Y \sim Beta(\frac{d-1}{2}, \frac{d-1}{2})$, then $X = \frac{1-(1+b)Y}{1-(1-b)Y} \sim e(x; b)$. Ulrich in his Theorem 2 proves that the value of b which is suitable for this envelope is

$$b = \frac{-2\kappa + \sqrt{4\kappa^2 + (d-1)^2}}{d-1}.$$

Wood (1994) corrects the algorithm introduced by Ulrich which does not work correctly. His modified algorithm is coded in Mathematica as follows. This program generates data from a von Mises Fisher distribution with the concentration parameter *kappa* and the rotational matrix Q introduced in Theorem 5.1.

```
n = size;
v1 = {}; v = {}; s1 = {}; s = {}; vonMises = {};
dimension = d;
m = (d - 1)/2;
kappa = k;
b = (-2*k + Sqrt[4*k^2 + (d - 1)^2])/(d - 1);
x0 = (1 - b)/(1 + b);
c = k*x0 + (d - 1)*Log[1 - x0^2];
Do[z = RandomVariate[BetaDistribution[m, m]];
 u = RandomReal[{0, 1}];
 w = (1 - (1 + b) z)/(1 - (1 - b) z);
 v1 = RandomVariate[NormalDistribution[0, 1], d - 1];
```

```
v = v1/Norm[v1];
s1 = Sqrt[1 - w^2] v;
s = Q.Flatten[{s1, w}];
If[Log[u] < k*w + (d - 1)*Log[1 - x0*w] - c, AppendTo[vonMises, s]];
v1 = {}; v = {}; s1 = {}; s = {}
, {n}]
```

5.3 Methods of Uniform Simulation on the Sphere

To use the forgoing program, we need to generate data from a uniform distribution on the sphere. In this section, we describe different methods doing this.

The simplest and most efficient method for generating data from a uniform distribution on the $(d-1)$ dimensional sphere is to generate $d-1$ independent observations from a standard normal distribution and then transform them by

$$\frac{V}{\sqrt{V^T V}} = \frac{V}{||V||} \tag{5.9}$$

to locate them on the sphere. Alternatively, we can use a uniform distribution on the interval $(-1, 1)$ instead of a normal distribution and then apply (5.9) to simulate uniform data on the sphere.

For a 3 dimensional case, to use the Ulrich Theorem to generate data from a 3 dimensional von Mises Fisher distribution, we first need to generate data from a 2 dimensional uniform distribution on the circle. In this case, we can also present a method of uniform simulation on the sphere by considering the angle θ. We generate θ from $U(0, 2\pi)$, and then let $x = \cos\theta$ and $y = \sin\theta$, so (x, y) has the desired uniform distribution on the circle. The following code generates data from a 3 dimensional von Mises Fisher distribution with the concentration parameter kappa and modal vector $\boldsymbol{\mu} = (\text{mu1}, \text{mu2}, \text{mu3})^T$, which we call $VM(\text{kappa}, \boldsymbol{\mu})$ for the sake of simplicity, while considering three different methods of uniform simulation on the sphere. After running this program, we have three different sets of data from $VM(\text{kappa}, \boldsymbol{\mu})$ using the three different methods of uniform simulation on the sphere. Each sample has size n. The output of this program includes three sets of data, namely 'vonMises3u' containing n 3 dimensional $VM(\text{kappa}, \boldsymbol{\mu})$ observations obtained from the uniform method to simulate uniform data on the sphere, 'vonMises3n' using the normal method of uniform simulation, and finally 'vonMises3t' using the angle method for uniform simulation on the sphere. Figure 5.2 compares these three different methods.

```
n = size;
vu1 = {}; vu = {}; su1 = {}; su = {}; vonMises3u = {};
vn1 = {}; vn = {}; sn1 = {}; sn = {}; vonMises3n = {};
vt1 = {}; vt = {}; st1 = {}; st = {}; vonMises3t = {};
kappa =k;
mu1 = mu1_0;
mu2 = mu2_0;
mu3 = Sqrt[1 - mu1_0^2 - mu2_0^2];
Q = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
   mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
   Sqrt[1 - mu2^2], mu3}};
Do[
 vu1 = RandomVariate[UniformDistribution[{-1, 1}], 2];
 vu = vu1/Norm[vu1];
 vn1 = RandomVariate[NormalDistribution[0, 1], 2];
 vn = vn1/Norm[vn1];
 theta = RandomVariate[UniformDistribution[{0, 2*Pi}]];
 vt = {{Cos[theta], Sin[theta]}};
 u = RandomVariate[UniformDistribution[{0, 1}]];
 w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*kappa)])/kappa;
 su1 = Sqrt[1 - w^2]*vu;
 su = Q.Flatten[{su1, w}];
 sn1 = Sqrt[1 - w^2]*vn;
 sn = Q.Flatten[{sn1, w}];
 st1 = Sqrt[1 - w^2]*vt;
 st = Q.Flatten[{st1, w}];
 AppendTo[vonMises3u, su];
 AppendTo[vonMises3n, sn];
 AppendTo[vonMises3t, st];
 vu1 = {}; vu = {}; su1 = {}; su = {};
 vn1 = {}; vn = {}; sn1 = {}; sn = {};
 vt1 = {}; vt = {}; st1 = {}; st = {}
 , {n}];
```

5.3.1 Power Function for Testing $H_0 : \kappa = 30$

We confirmed with some initial exploration that all 3 methods work well and there is little to choose between them in terms of computational time. As an example of a simulation, Figure 5.2 compares three power functions related to the hypothesis test (5.11) while using the three methods of generating uniform data on the sphere. This figure also confirms that there are no significant differences among these three methods. The following is the code to calculate the power functions.

```
a = 0; n = size; r = replication;
po95u = 0; power95u = {};
po95n = 0; power95n = {};
po95t = 0; power95t = {};
For[po95u = 0, po95u < r, a = a + 1,
  po95u = 0; po95n = 0; po95t = 0;
  Do[
   vu1 = {}; vu = {}; su1 = {}; su = {}; vonMises3u = {};
   vn1 = {}; vn = {}; sn1 = {}; sn = {}; vonMises3n = {};
   vt1 = {}; vt = {}; st1 = {}; st = {}; vonMises3t = {};
   k = 30 + a;
   mu1 = 0.5;
   mu2 = 0.5;
   mu3 = Sqrt[1 - mu1^2 - mu2^2];
   Q = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
   mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
   Sqrt[1 - mu2^2], mu3}};
   Do[
    vu1 = RandomVariate[UniformDistribution[{-1, 1}], 2];
    vu = vu1/Norm[vu1];
    vn1 = RandomVariate[NormalDistribution[0, 1], 2];
    vn = vn1/Norm[vn1];
    theta = RandomVariate[UniformDistribution[{0, 2*Pi}]];
    vt = {{Cos[theta], Sin[theta]}};
    u = RandomVariate[UniformDistribution[{0, 1}]];
    w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*k)])/k;
    su1 = Sqrt[1 - w^2]*vu;
    su = Q.Flatten[{su1, w}];
    sn1 = Sqrt[1 - w^2]*vn;
    sn = Q.Flatten[{sn1, w}];
```

```
    st1 = Sqrt[1 - w^2]*vt;
    st = Q.Flatten[{st1, w}];
    AppendTo[vonMises3u, su];
    AppendTo[vonMises3n, sn];
    AppendTo[vonMises3t, st];
    vu1 = {}; vu = {}; su1 = {}; su = {};
    vn1 = {}; vn = {}; sn1 = {}; sn = {};
    vt1 = {}; vt = {}; st1 = {}; st = {}
    , {n}];
   LAu = n*(-Log[c3[kk]] + kk*(Norm[Mean[vonMises3n]]));
   L0u = n*(-Log[c3[30]] + 30*Norm[Mean[vonMises3n]]);
   dLAu = FindMaxValue[{LAu, kk > 0}, {{kk, 30 + a}}];
   dn2u = 2*(dLAu - L0u);
   If[dn2u > Quantile[ChiSquareDistribution[1], 0.95],
    po95u = po95u + 1];
   LAn = n*(-Log[c3[kk]] + kk*(Norm[Mean[vonMises3n]]));
   L0n = n*(-Log[c3[30]] + 30*Norm[Mean[vonMises3n]]);
   dLAn = FindMaxValue[{LAn, kk > 0}, {{kk, 30 + a}}];
   dn2n = 2*(dLAn - L0n);
   If[dn2n > Quantile[ChiSquareDistribution[1], 0.95],
    po95n = po95n + 1];
   LAt = n*(-Log[c3[kk]] + kk*Norm[Mean[vonMises3t]]);
   L0t = n*(-Log[c3[30]] + 30*Norm[Mean[vonMises3t]]);
   dLAt = FindMaxValue[{LAt, kk > 0}, {{kk, 30 + a}}];
   dn2t = 2*(dLAt - L0t);
   If[dn2t > Quantile[ChiSquareDistribution[1], 0.95],
    po95t = po95t + 1];
   vonMise3n = {}; vonMises3u = {}; vonMises3t = {}
   , {r}];
  AppendTo[power95u, {k, po95u/ek}];
  AppendTo[power95n, {k, po95n/ek}];
  AppendTo[power95t, {k, po95t/ek}];
  ];
ListLinePlot[{power95u, power95n, power95t},
 AxesLabel -> {\[Kappa], PowerFunction}, PlotStyle -> Automatic,
 PlotLegends -> Placed[{"Uniform", "Normal", "Theta"}, After],
 PlotMarkers -> {Automatic, 16}]
```

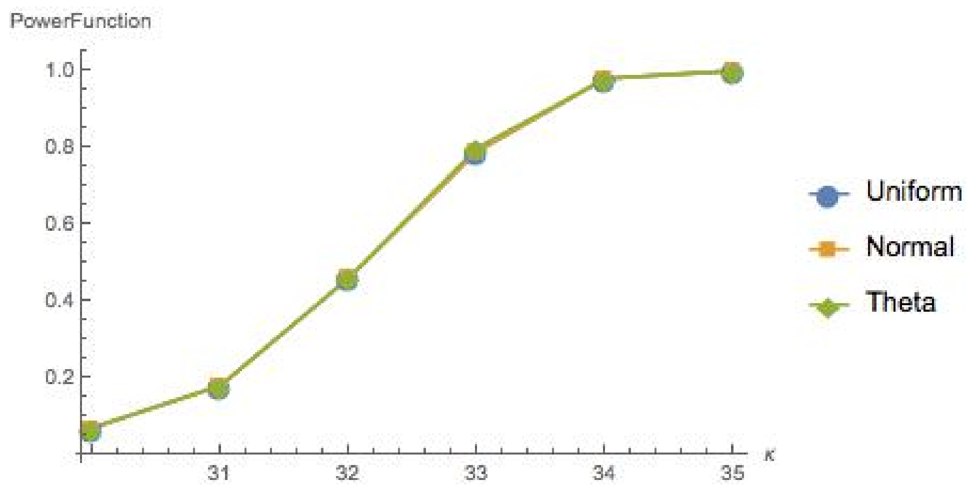

FIGURE 5.2: Power function of hypothesis test (5.11) using three methods to generate uniform data on the sphere, $n = 1000$ and $r = 100$

5.4 Plots

Figures 5.3 to 5.6 show simulated von Mises Fisher distributions on the sphere with different choices of parameters. The method used for the simulation is the method presented in Section 5.2.2 and for the 2 dimensional case, we use the method described in Section 5.2.3 with normal method of simulation on the sphere for both cases.
Figures 5.4 and 5.6 can be compared favourably to the contour plots of 2.1 and 2.2 in Chapter 2.

5.5 Hypothesis Testing through Simulation

In this section, we consider some hypothesis tests for the parameters of the von Mises Fisher distribution and illustrate Wilk's theorem on page 7 of the thesis by the simulation methods introduced in this chapter.
From now on, we use the normal method to simulate uniform data on the sphere. Consider a 3 dimensional von Mises Fisher distribution with parameters κ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T$ and the hypotheses

$$\begin{cases} H_0 & : \kappa = 30, \mu_1 = 0.5, \mu_2 = 0.5 \\ H_A & : \kappa \neq 30, \mu_1 \neq 0.5, \mu_2 \neq 0.5, \end{cases} \tag{5.10}$$

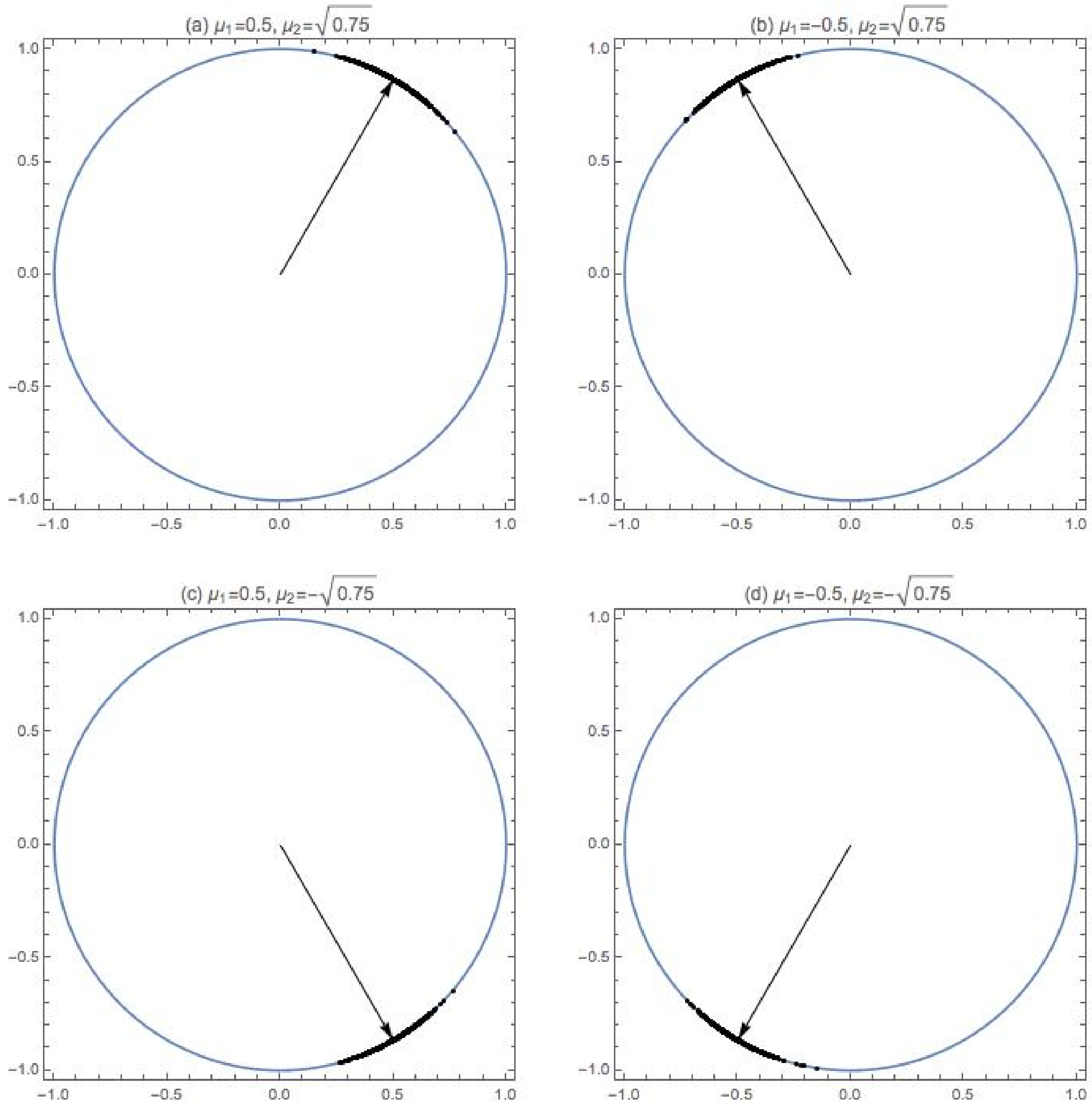

FIGURE 5.3: Comparing different values of the modal vectors in a 2 dimensional von Mises Fisher distribution using 1000 simulated observations from the distribution with parameter $\kappa = 100$

$$\begin{cases} H_0 & : \kappa = 30 \\ H_A & : \kappa \neq 30 \end{cases} \tag{5.11}$$

and

$$\begin{cases} H_0 & : \mu_1 = 0.5, \mu_2 = 0.5 \\ H_A & : \mu_1 \neq 0.5, \mu_2 \neq 0.5. \end{cases} \tag{5.12}$$

To simulate the asymptotic distribution of d_n for these hypotheses by the simulation method, we considered a von Mises Fisher distribution with the concentration parameter

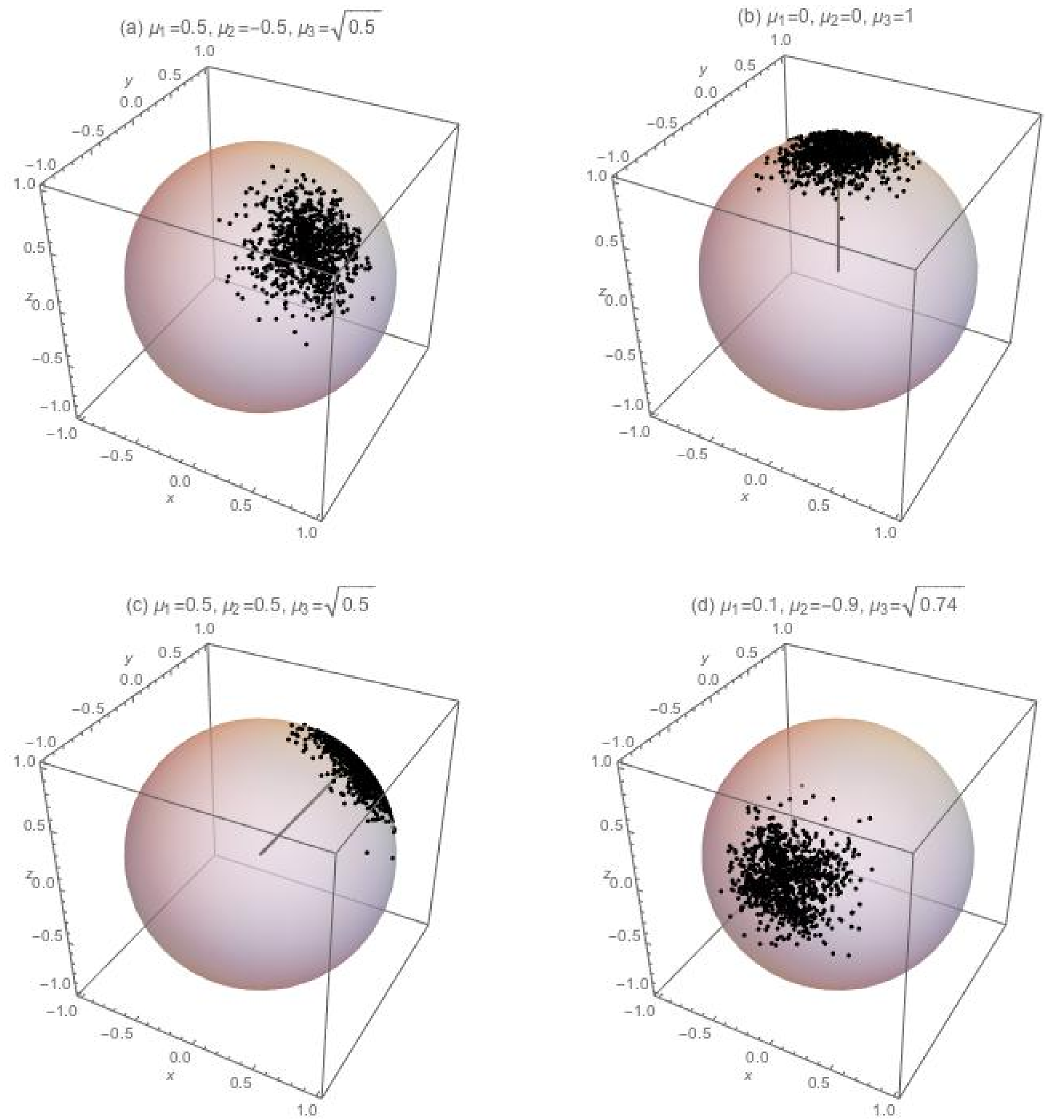

FIGURE 5.4: Comparing different values of the modal vectors in a 3 dimensional von Mises Fisher distribution using 1000 simulated observations from the distribution with parameter $\kappa = 30$

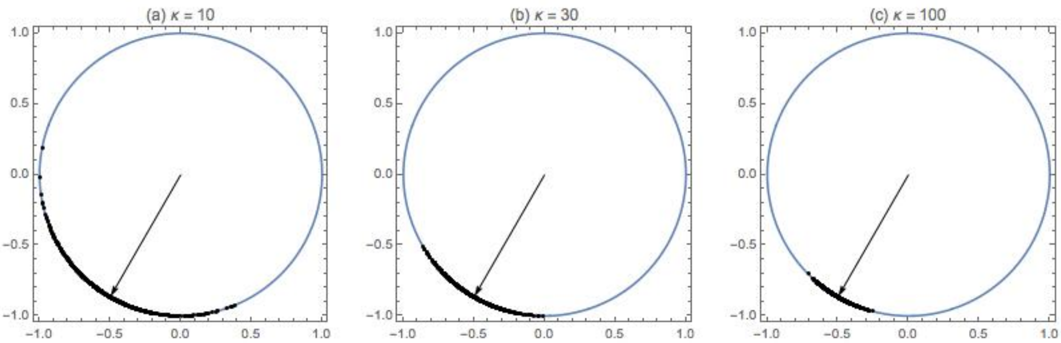

FIGURE 5.5: Comparing different values of κ in a 2 dimensional von Mises Fisher distribution using 1000 simulated observations from the distribution with modal vector $\mu_1 = -0.5, \mu_2 = -\sqrt{1 - 0.5^2}$

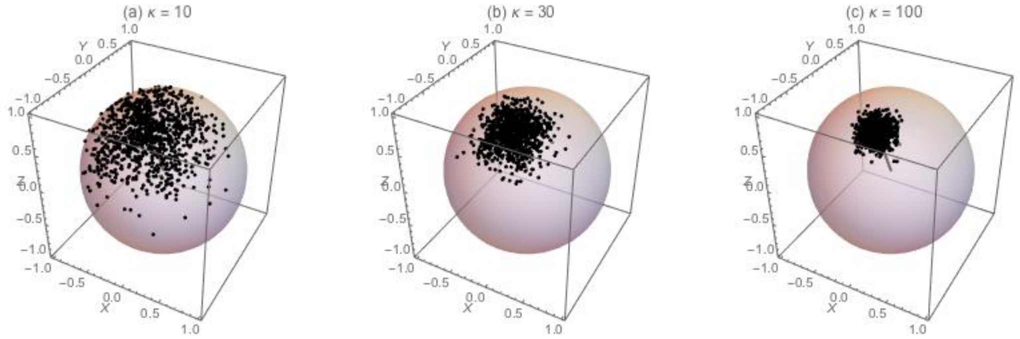

FIGURE 5.6: Comparing different values of κ in a 3 dimensional von Mises Fisher distribution using 1000 simulated observations from the distribution with modal vector $\mu_1 = 0.1, \mu_2 = -0.5, \mu_3 = \sqrt{1 - 0.1^2 - 0.5^2}$

30 and the modal vector $(0.5, 0.5, \sqrt{1 - 2 \times 0.5^2})^T$, and generated n data points from this distribution. Then we calculated the value of d_n using the formula (1.6), and repeated the process r times to simulate the distribution of d_n. Tables 5.1 and 5.2 and Figures 5.8 to 5.10 show the results for different values of n and r. These tables show that with increasing n and r, the simulated distributions coincide closely with the ones from the theory. Tables 5.3 and 5.4 are related to the same process conducted for 2 and 10 dimensions. As these tables show the results are not as good as for 3 dimensions. We suggest this is because for 3 dimensions, we have an explicit formula for the distribution function, so the simulation method is exact and functions very efficiently.

To create Table 5.3 for the 2 dimensional cases, we considered a von Mises Fisher distribution with concentration parameter $\kappa = 100$ and modal vector $(0.5, \sqrt{1 - 0.5^2})^T$ and tested the following hypotheses:

$$\begin{cases} H_0 & : \kappa = 100, \mu_1 = 0.5 \\ H_A & : \kappa \neq 100, \mu_1 \neq 0.5, \end{cases} \tag{5.13}$$

$$\begin{cases} H_0 & : \kappa = 100 \\ H_A & : \kappa \neq 100 \end{cases} \tag{5.14}$$

and

$$\begin{cases} H_0 & : \mu_1 = 0.5 \\ H_A & : \mu_1 \neq 0.5. \end{cases} \tag{5.15}$$

The results are shown in Table 5.3 and Figures 5.11 and 5.12.

For the 10 dimensional cases, the distribution is a 10 dimensional von Mises Fisher distribution with concentration parameter $\kappa = 100$ and modal vector $(0, 0, \ldots, 0, 1)^T$ and the hypotheses are

$$\begin{cases} H_0 & : \kappa = 100, \mu_1 = 0, \ldots, \mu_9 = 0, \mu_{10} = 1 \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{5.16}$$

$$\begin{cases} H_0 & : \kappa = 100 \\ H_A & : \kappa \neq 100 \end{cases} \tag{5.17}$$

and

$$\begin{cases} H_0 & : \mu_1 = 0, \ldots, \mu_9 = 0, \mu_{10} = 1 \\ H_A & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.18}$$

Table 5.4 and Figures 5.13 and 5.14 show the results for these hypothesis tests.

In Figures 5.8 to 5.14, the plots (a), (d) and (g) are demonstrations when the number of repetitions is 100. Plots (b), (e) and (h) are for r=1000 and plots (c), (f) and (i) are for r=10,000. As we see when the number of repetitions becomes larger, the quantile plots show a better fit of the model. Also, it can be seen from Tables 5.1 to 5.4 that when we consider the estimated degrees of freedom, Watson's U^2 p-value and Kolmogorov-Smirnov's p-value, the theory is confirmed by the simulation. Our methodology is well confirmed in Tables 5.1 and 5.2 when the simulation is from the 3 dimensional von Mises Fisher distribution. For Tables 5.3 and 5.4, we used a trimmed mean to calculate the estimated degree of freedom so as to reduce the bias incurred in the simulation methods.

The program for doing the simulations for the hypotheses (5.10) to (5.12) is as follows:

```
n = size; r = replication;
vn1 = {}; vn = {}; sn1 = {}; sn = {}; vonMises3n = {};
k = 30;
mu1 = 0.5;
mu2 = 0.5;
mu3 = Sqrt[1 - mu1^2 - mu2^2];
Q = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
   mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
   Sqrt[1 - mu2^2], mu3}};
dn313 = {}; dn323 = {}; dn333 = {};
Do[
```

```
  Do[
   vn1 = RandomVariate[NormalDistribution[0, 1], 2];
   vn = vn1/Norm[vn1];
   u = RandomVariate[UniformDistribution[{0, 1}]];
   w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*k)])/k;
   sn1 = Sqrt[1 - w^2]*vn;
   sn = Q.Flatten[{sn1, w}];
   AppendTo[vonMises3n, sn];
   vn1 = {}; vn = {}; sn1 = {}; sn = {}
   , {n}];
  LA = n*(-Log[c3[kk3]] + kk3*(Norm[Mean[vonMises3n]]));
  L03 = n*(-Log[c3[kk31]] +
      kk31*({mu1, mu2, mu3}.Mean[vonMises3n]));
  L02 = n*(-Log[c3[30]] + 30*Norm[Mean[vonMises3n]]);
  L01 = n*(-Log[c3[30]] +
      30*({mu1, mu2, mu3}.Mean[vonMises3n]));
  dLA = FindMaxValue[{LA, kk3 > 0}, {kk3, 30}];
  dL03 = FindMaxValue[{L03, kk31 > 0}, {kk31, 30}];
  AppendTo[dn313, 2*(dLA - L01)];
  AppendTo[dn323, 2*(dLA - L02)];
  AppendTo[dn333, 2*(dLA - dL03)];
  vonMises3n = {}
  , {r}];
DistributionFitTest[Flatten[dn313],
 ChiSquareDistribution[3], {"TestDataTable", All}]
EstimatedDistribution[dn313, ChiSquareDistribution[beta]]
DistributionFitTest[Flatten[dn323],
 ChiSquareDistribution[1], {"TestDataTable", All}]
EstimatedDistribution[dn323, ChiSquareDistribution[beta]]
DistributionFitTest[Flatten[dn333],
 ChiSquareDistribution[2], {"TestDataTable", All}]
EstimatedDistribution[dn333, ChiSquareDistribution[beta]]
```

5.6 Hypothesis Testing in the Spherical Subcomponent Models

Consider the spherical subcomponent model introduced in Section 3.4.1 and the estimating function in (3.46). In this section, we illustrate Theorems 3.7 and 3.8 presented in

Chapter 3 by simulation methods. Suppose we are interested in testing the hypotheses

$$\begin{cases} H_0 & : \kappa_{21} = \kappa_{22} = \kappa_{23} \\ H_A & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.19}$$

Theorem 3.10 states that the asymptotic distribution of d_n for this hypothesis test is chi square with 2 degrees of freedom. In this section, we verify this result through simulation. Consider three 2 dimensional von Mises Fisher distributions with parameters

$$\kappa_{21} = 100, \quad \boldsymbol{\mu}_{21} = (0.52, \sqrt{1-0.52^2})^T,$$
$$\kappa_{22} = 100, \quad \boldsymbol{\mu}_{22} = (-0.4, \sqrt{1-0.4^2}),$$
$$\kappa_{23} = 100, \quad \boldsymbol{\mu}_{23} = (0.02, \sqrt{1-0.02^2})^T$$

and one 3 dimensional von Mises Fisher distribution with the parameters

$$\kappa = 30, \quad \boldsymbol{\mu}_{31} = (0.5, 0.5, \sqrt{1-2\times 0.5^2})^T.$$

We simulated n data points from each distribution via the methods discussed in the previous section with two different values of $n = 100$ and $n = 1000$, and calculated d_n 1000 times. Under the null hypothesis, we maximized the estimating function (3.46) when the first three κ_{21}, κ_{22}, and κ_{23} related to the 2 dimensional distributions were equal to an unknown value κ to find the maximum value of the function (3.46) under the constraints $\kappa > 0, ||\boldsymbol{\mu}_{21}|| = 1, ||\boldsymbol{\mu}_{22}|| = 1, ||\boldsymbol{\mu}_{23}|| = 1, \kappa_{31} > 0$ and $||\boldsymbol{\mu}_{31}|| = 1$. For the alternative hypothesis, we found the maximum value of the function (3.46) under the constraints $\kappa_{21} > 0, ||\boldsymbol{\mu}_{21}|| = 1, \kappa_{22} > 0, ||\boldsymbol{\mu}_{22}|| = 1, \kappa_{23} > 0, ||\boldsymbol{\mu}_{23}|| = 1, \kappa_{31} > 0$, and $||\boldsymbol{\mu}_{31}|| = 1$. We repeated the whole procedure r times to calculate d_n and thus find the empirical distribution of the d_n. Row 2 in Tables 5.5 and 5.7 and Figure 5.15 show the results. The seventh column in Tables 5.5 and 5.7 gives the estimated degrees of freedom of the chi square distribution, and column eight gives the p-values of the goodness of fit statistic based on the Watson U^2 and the Kolmogorov-Smirnov tests. Tables 5.5 to 5.8 and Figure 5.15 show the results of the simulations for the hypothesis tests described below.

The third row of Tables 5.5 and 5.7 relates to the following hypothesis. Consider the estimating function

$$\mathcal{L}_n(\boldsymbol{\theta}) = -n_{21}\log c^{(2)}(\kappa^{(2)}) - n_{31}\log c^{(3)}(\kappa^{(3)})$$
$$+ n_{21}\kappa^{(2)}\boldsymbol{\mu}_2^T\overline{\mathbf{X}}_2 + n_{31}\kappa^{(3)}\boldsymbol{\mu}_3^T\overline{\mathbf{X}}_3.$$

The asymptotic distribution of d_n for the hypothesis test

$$\begin{cases} H_0 & : \kappa^{(2)} = \kappa^{(3)} \\ H_A & : \kappa^{(2)} \neq \kappa^{(3)} \end{cases} \tag{5.20}$$

is chi square with one degree of freedom according to Theorem 3.8. To examine this result by simulation, we considered one 2 dimensional and one 3 dimensional von Mises Fisher distribution with parameters

$$\begin{aligned} \kappa^{(2)} &= 100, \quad \boldsymbol{\mu}_2 = (0.52, \sqrt{1-0.52^2})^T \\ \kappa^{(3)} &= 30, \quad\ \ \boldsymbol{\mu}_3 = (-0.2, 0.5, \sqrt{1-0.2^2-0.5^2})^T. \end{aligned}$$

The third rows of Tables 5.5 and 5.7 show the results of the simulations for this test.

As previously discussed the method of simulation for the 3 dimensional von Mises Fisher distribution is more accurate than for higher dimensions. Therefore, we considered a spherical subcomponent model combining only 3 dimensional distributions to conduct the hypothesis tests. That is, we considered the estimating function

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{j=1}^{3} n_{3j} \left(-\log c^{(3)}(\kappa_{3j}) + \boldsymbol{\mu}_{3j}^T \overline{\mathbf{x}}_{3j} \right)$$

and simulated data from this distribution with parameters

$$\begin{aligned} \kappa_{31} &= 30, \quad \boldsymbol{\mu}_{31} = (0.5, 0.5, \sqrt{1-2\times 0.5^2})^T, \\ \kappa_{32} &= 30, \quad \boldsymbol{\mu}_{32} = (0.5, -0.5, \sqrt{1-2\times 0.5^2})^T, \\ \kappa_{33} &= 30, \quad \boldsymbol{\mu}_{33} = (-0.2, 0, \sqrt{1-0.2^2})^T \end{aligned}$$

to test

$$\begin{cases} H_0 & : \kappa_{31} = \kappa_{32} = \kappa_{33} \\ H_a & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.21}$$

The first rows of Tables 5.5 and 5.7 show the results.

The estimating function for the fourth row in Tables 5.5 and 5.7 is

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\theta}) = &- \sum_{j=1}^{6} n_{2j} \log c^{(2)}(\kappa_{2j}) - \sum_{j=1}^{4} n_{3j} \log c^{(3)}(\kappa_{3j}) \\ &- n_{41} \log c^{(4)}(\kappa_{41}) + \sum_{i=2}^{4} \sum_{j=1}^{m_i} n_{ij} \kappa_{ij} \boldsymbol{\mu}_{ij}^T \overline{\mathbf{X}}_{ij}, \end{aligned} \tag{5.22}$$

where $m_2 = 6$, $m_3 = 4$ and $m_4 = 1$, and $\boldsymbol{\mu}_{ij}$ and $\overline{\mathbf{X}}_{ij}$ are i dimensional vectors in $\mathbb{S}^i$. There are in total 11 distributions contained in this estimating function; 6 of them are 2 dimensional, 4 are 3 dimensional and one is a 4-dimensional von Mises Fisher distribution. We simulated 100 observations from each distribution and tested whether the kappas related to the 2 dimensional von Mises Fisher distributions are equal or not. Therefore, the hypothesis is

$$\begin{cases} H_0 & : \kappa_{21} = \kappa_{22} = \ldots = \kappa_{26} \\ H_A & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.23}$$

For the simulation, we choose the parameters to be

$$\begin{aligned}
\kappa_{21} &= 100, & \boldsymbol{\mu}_{21} &= (0.52, \sqrt{1-0.52^2})^T, \\
\kappa_{22} &= 100, & \boldsymbol{\mu}_{22} &= (-0.4, \sqrt{1-0.4^2})^T, \\
\kappa_{23} &= 100, & \boldsymbol{\mu}_{23} &= (0.02, \sqrt{1-0.02^2})^T, \\
\kappa_{24} &= 100, & \boldsymbol{\mu}_{24} &= (0.35, \sqrt{1-0.35^2})^T, \\
\kappa_{25} &= 100, & \boldsymbol{\mu}_{25} &= (0.7, -\sqrt{1-0.7^2})^T, \\
\kappa_{26} &= 100, & \boldsymbol{\mu}_{26} &= (-0.1, -\sqrt{1-0.1^2})^T, \\
\kappa_{31} &= 30, & \boldsymbol{\mu}_{31} &= (0.1, 0.5, \sqrt{1-0.1^2-0.5^2})^T, \\
\kappa_{32} &= 33, & \boldsymbol{\mu}_{32} &= (0.2, -0.6, \sqrt{1-0.2^2-0.6^2})^T, \\
\kappa_{33} &= 37, & \boldsymbol{\mu}_{33} &= (-0.2, 0.3, \sqrt{1-0.2^2-0.3^2})^T, \\
\kappa_{34} &= 40, & \boldsymbol{\mu}_{34} &= (0.5, 0.01, -\sqrt{1-0.5^2-0.01^2})^T, \\
\kappa_{41} &= 150, & \boldsymbol{\mu}_{41} &= (0, 0, 0, 1)^T
\end{aligned}$$

and performed the procedure described for row 2 above in order to calculate d_n.

For the fifth row, the hypothesis is

$$\begin{cases} H_0 & : \kappa_{31} = \kappa_{32} = \kappa_{33} = \kappa_{34} \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{5.24}$$

with the parameters

$$\begin{aligned}
\kappa_{21} &= 110, \quad \boldsymbol{\mu}_{21} = (0.52, \sqrt{1-0.52^2})^T, \\
\kappa_{22} &= 95, \quad \boldsymbol{\mu}_{22} = (-0.4, \sqrt{1-0.4^2})^T, \\
\kappa_{23} &= 150, \quad \boldsymbol{\mu}_{23} = (0.02, \sqrt{1-0.02^2})^T, \\
\kappa_{24} &= 130, \quad \boldsymbol{\mu}_{24} = (0.35, \sqrt{1-0.35^2})^T, \\
\kappa_{25} &= 136, \quad \boldsymbol{\mu}_{25} = (0.7, -\sqrt{1-0.7^2})^T, \\
\kappa_{26} &= 120, \quad \boldsymbol{\mu}_{26} = (-0.1, -\sqrt{1-0.1^2})^T, \\
\kappa_{31} &= 35, \quad \boldsymbol{\mu}_{31} = (0.1, 0.5, \sqrt{1-0.1^2-0.5^2})^T, \\
\kappa_{32} &= 35, \quad \boldsymbol{\mu}_{32} = (0.2, -0.6, \sqrt{1-0.2^2-0.6^2})^T, \\
\kappa_{33} &= 35, \quad \boldsymbol{\mu}_{33} = (-0.2, 0.3, \sqrt{1-0.2^2-0.3^2})^T, \\
\kappa_{34} &= 35, \quad \boldsymbol{\mu}_{34} = (0.5, 0.01, -\sqrt{1-0.5^2-0.01^2})^T, \\
\kappa_{41} &= 170, \quad \boldsymbol{\mu}_{41} = (0, 0, 0, 1)^T.
\end{aligned}$$

For the first row in Tables 5.6 and 5.8, the estimating function is that we considered in (5.22) with the selected parameters

$$\begin{aligned}
\kappa_{21} &= 100, \quad \boldsymbol{\mu}_{21} = (0.52, \sqrt{1-0.52^2})^T, \\
\kappa_{22} &= 100, \quad \boldsymbol{\mu}_{22} = (-0.4, \sqrt{1-0.4^2})^T, \\
\kappa_{23} &= 100, \quad \boldsymbol{\mu}_{23} = (0.02, \sqrt{1-0.02^2})^T, \\
\kappa_{24} &= 100, \quad \boldsymbol{\mu}_{24} = (0.35, \sqrt{1-0.35^2})^T, \\
\kappa_{25} &= 100, \quad \boldsymbol{\mu}_{25} = (0.7, -\sqrt{1-0.7^2})^T, \\
\kappa_{26} &= 100, \quad \boldsymbol{\mu}_{26} = (-0.1, -\sqrt{1-0.1^2})^T, \\
\kappa_{31} &= 100, \quad \boldsymbol{\mu}_{31} = (0.1, 0.5, \sqrt{1-0.1^2-0.5^2})^T, \\
\kappa_{32} &= 100, \quad \boldsymbol{\mu}_{32} = (0.2, -0.6, \sqrt{1-0.2^2-0.6^2})^T, \\
\kappa_{33} &= 100, \quad \boldsymbol{\mu}_{33} = (-0.2, 0.3, \sqrt{1-0.2^2-0.3^2})^T, \\
\kappa_{34} &= 100, \quad \boldsymbol{\mu}_{34} = (0.5, 0.01, -\sqrt{1-0.5^2-0.01^2})^T, \\
\kappa_{41} &= 150, \quad \boldsymbol{\mu}_{41} = (0, 0, 0, 1)^T,
\end{aligned}$$

where the hypothesis is

$$\begin{cases} H_0 & : \kappa_{21} = \ldots = \kappa_{26} = \kappa_{31} = \ldots = \kappa_{34} \\ H_A & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.25}$$

For the second row in Tables 5.6 and 5.8, the estimating function is

$$\begin{aligned}
\mathcal{L}_n(\boldsymbol{\theta}) = &-n_{21} \log c^{(2)}(\kappa^{(2)}) - n_{31} \log c^{(3)}(\kappa^{(3)}) - n_{41} \log c^{(4)}(\kappa^{(4)}) \\
&+ n_{21}\kappa^{(2)} \boldsymbol{\mu}_2^T \overline{\mathbf{X}}_2 + n_{31}\kappa^{(3)} \boldsymbol{\mu}_3^T \overline{\mathbf{X}}_3 + n_{41}\kappa^{(4)} \boldsymbol{\mu}_4^T \overline{\mathbf{X}}_4,
\end{aligned}$$

and the hypothesis is

$$\begin{cases} H_0 & : \kappa^{(2)} = \kappa^{(3)} = \kappa^{(4)} \\ H_A & : \text{at least one equality is not satisfied.} \end{cases} \tag{5.26}$$

The selected parameters for this hypothesis are

$$\begin{aligned} \kappa^{(2)} &= 100, \quad \boldsymbol{\mu}_2 = (0.52, \sqrt{1-0.52^2})^T, \\ \kappa^{(3)} &= 100, \quad \boldsymbol{\mu}_3 = (-0.2, 0.5, \sqrt{1-0.2^2-0.5^2})^T, \\ \kappa^{(4)} &= 100, \quad \boldsymbol{\mu}_4 = (0,0,0,1)^T. \end{aligned}$$

The last row in Tables 5.6 and 5.8 is related to the hypotheses

$$\begin{cases} H_0 & : \kappa^{(2)} = \ldots = \kappa^{(10)}, \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{5.27}$$

with the estimating function

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\sum_{i=1}^{10} n_{i1} \log c^{(i)}(\kappa^{(i)}) + \sum_{i=1}^{10} n_{i1} \kappa^{(i)} \boldsymbol{\mu}_i^T \overline{\mathbf{X}}_i.$$

We selected the parameters for conducting the hypothesis (5.27) to be

$$\begin{aligned} \kappa^{(2)} &= 100, \quad \boldsymbol{\mu}_2 = (0,1)^T, \\ \kappa^{(3)} &= 100, \quad \boldsymbol{\mu}_3 = (0.5, 0.5, \sqrt{1-0.5^2-0.5^2})^T, \\ \kappa^{(4)} &= 100, \quad \boldsymbol{\mu}_4 = (0,0,0,1)^T, \\ \kappa^{(5)} &= 100, \quad \boldsymbol{\mu}_5 = (0,0,0,0,1)^T, \\ \kappa^{(6)} &= 100, \quad \boldsymbol{\mu}_6 = (0,0,0,0,0,1)^T, \\ \kappa^{(7)} &= 100, \quad \boldsymbol{\mu}_7 = (0,0,0,0,0,0,1)^T, \\ \kappa^{(8)} &= 100, \quad \boldsymbol{\mu}_8 = (0,0,0,0,0,0,0,1)^T, \\ \kappa^{(9)} &= 100, \quad \boldsymbol{\mu}_9 = (0,0,0,0,0,0,0,0,1)^T, \\ \kappa^{(10)} &= 100, \quad \boldsymbol{\mu}_{10} = (0,0,0,0,0,0,0,0,0,1)^T. \end{aligned}$$

As Figures in 5.15 show that, when the number of repetitions (r) becomes larger, the histograms and quantile plots represent a better fit of the model. The plots on the right hand side are conducted when r=1000 and the left hand side are for r=100 when we choose to have n=1000 spherical observations simulated from the distributions. The estimated degrees of freedom and the two different p-values calculated in Tables 5.5 to 5.8 are other information that show how well the theory is supported by the simulations.

As these tables show, our model is well confirmed when the simulations are solely from 3 dimensional von Mises Fisher distributions.

5.7 Power Function for Testing some Hypotheses in the Spherical Subcomponent Model

We simulated 100 observations from the distribution (3.46) containing three 2 dimensional von Mises Fisher distributions with parameters

$$\begin{aligned} \kappa_{21} &= 110, & \boldsymbol{\mu}_{21} &= (0.52, \sqrt{1-0.52^2})^T, \\ \kappa_{22} &= 110 + a, & \boldsymbol{\mu}_{22} &= (0.1, \sqrt{1-0.1^2})^T, \\ \kappa_{23} &= 110 + 2a, & \boldsymbol{\mu}_{23} &= (-0.4, \sqrt{1-0.4^2})^T \end{aligned}$$

and one 3 dimensional von Mises Fisher distribution with parameters

$$\kappa_{30} = 1, \quad \boldsymbol{\mu}_{31} = (0.5, -0.5, \sqrt{1 - 2 \times 0.5^2})^T.$$

The purpose is to calculate the power function for the hypothesis test (5.19).

We maximized the estimating function (3.46) assuming that the three kappas related to the 2 dimensions are equal under the null hypothesis and different under the alternative hypothesis to calculate $\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_n^{\Omega})$ and $\mathcal{L}_n(\widehat{\boldsymbol{\theta}}_n^{\Omega\cup\tau})$ respectively. Then we calculated d_n from (1.6). We repeated the procedure until we had 1000 observations and then counted the number of d_ns greater than $\chi^2_{0.95}(2) = 5.9914$ and divided by the total number, 1000. The result is the empirical power of the test at the 0.05 level and for a specific a in the structure of κ.

The first value of the power was calculated for the first choice of κ, namely $\kappa = 110$, for $a = 0$ and it occurs when all the κs in the 2 dimensions are equal. Then we repeated the process for increments of $a = 3$ until the power reached one; approximately after a is greater than 120. Figure 5.7, part (a) shows the different values of the power plotted against a at the 0.05 level.

Figure 5.7, part (b) is the power function for testing the hypothesis (5.21) when we generated 100 observations from three 3 dimensional von Mises Fisher distributions with parameters

$$\begin{aligned} \kappa_{31} &= 50, & \boldsymbol{\mu}_{31} &= (0.1, 0.5, \sqrt{1-0.1^2-0.5^2})^T, \\ \kappa_{32} &= 50 + a, & \boldsymbol{\mu}_{32} &= (0.5, -0.5, \sqrt{1-2\times 0.5^2})^T, \\ \kappa_{33} &= 50 + 2a, & \boldsymbol{\mu}_{33} &= (-0.5, -0.1, \sqrt{1-0.5^2-0.1^2})^T \end{aligned}$$

and an increment of 1 for the value of a in each step.

Part (c) in Figure 5.7 shows the two power functions for the the hypotheses (5.19) and (5.21) in one graph. By presenting this graph, we show that the simulation method for the 3 dimensional von Mises Fisher distribution is more accurate than for the other dimensions.

The code for running this program is as follows:

```
vonMises31 = {}; vonMises32 = {}; vonMises33 = {};
a = 0; ek = 1000; num = 100;
pomix = 0; powermix33 = {};
For[pomix = 0, pomix < ek, a = a + 1,
  pomix = 0;
  Do[
    v1 = {}; v = {}; s1 = {}; s = {}; vonMises31 = {};
    k3 = 50;
    mu1 = 0.1;
    mu2 = 0.5;
    mu3 = Sqrt[1 - mu1^2 - mu2^2];
    Q31 = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
   mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
   Sqrt[1 - mu2^2], mu3}};
    Do[
     v1 = RandomVariate[NormalDistribution[0, 1], 2];
     v = v1/Norm[v1];
     u = RandomVariate[UniformDistribution[{0, 1}]];
     w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*k3)])/k3;
     s1 = Sqrt[1 - w^2]*v;
     s = Flatten[{s1, w}];
     AppendTo[vonMises31, s];
     v1 = {}; v = {}; s1 = {}; s = {}
     , {num}];
    v1 = {}; v = {}; s1 = {}; s = {}; vonMises32 = {};
    k3 = 50 + a;
    mu1 = 0.5;
    mu2 = -0.5;
    mu3 = Sqrt[1 - mu1^2 - mu2^2];
    Q32 = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
   mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
```

```
Sqrt[1 - mu2^2], mu3}};
    Do[
  v1 = RandomVariate[NormalDistribution[0, 1], 2];
  v = v1/Norm[v1];
  u = RandomVariate[UniformDistribution[{0, 1}]];
  w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*k3)])/k3;
  s1 = Sqrt[1 - w^2]*v;
  s = Q32.Flatten[{s1, w}];
  AppendTo[vonMises32, s];
  v1 = {}; v = {}; s1 = {}; s = {}
  , {num}];
 v1 = {}; v = {}; s1 = {}; s = {}; vonMises33 = {};
 k3 = 50 + 2*a;
 mu1 = -0.5;
 mu2 = -0.1;
 mu3 = Sqrt[1 - mu1^2 - mu2^2];
 Q33 = {{-(mu3/Sqrt[1 - mu2^2]), (mu1 mu2)/Sqrt[1 - mu2^2],
mu1}, {0, -Sqrt[1 - mu2^2], mu2}, {mu1/Sqrt[1 - mu2^2], (mu2 mu3)/
Sqrt[1 - mu2^2], mu3}};
 Do[
  v1 = RandomVariate[NormalDistribution[0, 1], 2];
  v = v1/Norm[v1];
  u = RandomVariate[UniformDistribution[{0, 1}]];
  w = 1 + (Log[u] + Log[1 - ((u - 1)/u)*E^(-2*k3)])/k3;
  s1 = Sqrt[1 - w^2]*v;
  s = Q33.Flatten[{s1, w}];
  AppendTo[vonMises33, s];
  v1 = {}; v = {}; s1 = {}; s = {}
  , {num}];
 LA1 =
  num*(-Log[c3[k31]] + k31*Norm[Mean[vonMises31]] - Log[c3[k32]] +
     k32*Norm[Mean[vonMises32]] - Log[c3[k33]] +
     k33*Norm[Mean[vonMises33]]);
 L01 =
  num*(-Log[c3[kk]] + kk*Norm[Mean[vonMises31]] - Log[c3[kk]] +
     kk*Norm[Mean[vonMises32]] - Log[c3[kk]] +
     kk*Norm[Mean[vonMises33]]);
 dLA1 =
  FindMaxValue[{LA1,
```

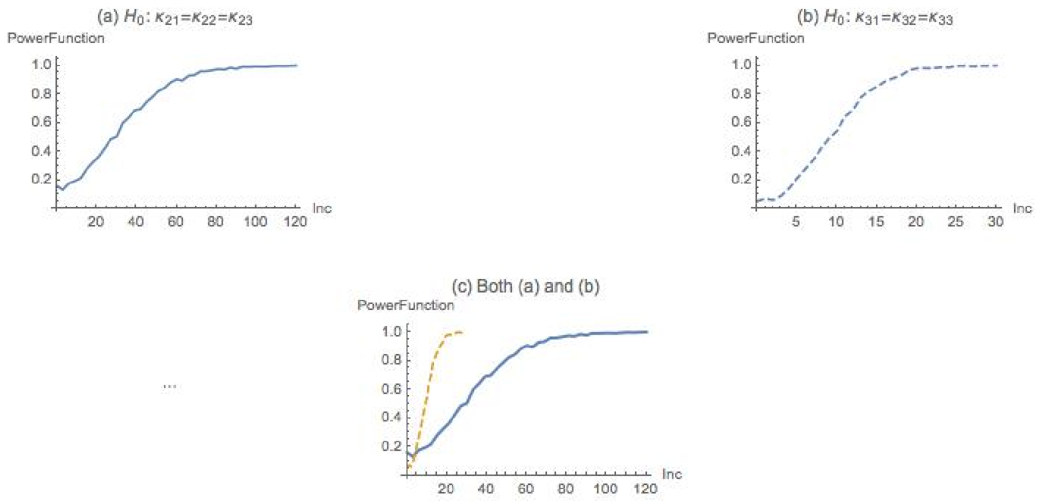

FIGURE 5.7: A comparison between two simulated power functions for the hypothesis tests (5.19) (in Theorem 3.10) and (5.21)

```
      k31 > 0 && k32 > 0 && k33 > 0}, {{k31, 50}, {k32,
       50 + a}, {k33, 50 + 2*a}}];
    dL01 = FindMaxValue[{L01, kk > 0}, {kk, 50 + a}];
    dn = 2*(dLA1 - dL01);
    If[dn > Quantile[ChiSquareDistribution[2], 0.95],
     pomix = pomix + 1];
    vonMises31 = {}; vonMises32 = {}; vonMises33 = {}
    , {ek}]
   AppendTo[powermix33, {a, pomix/ek}];
  ];
ListLinePlot[{powermix33}, AxesLabel -> {Inc, PowerFunction}]
```

Table 5.1: Simulation results for hypothesis tests (5.10), (5.11) and (5.12) in 3 dimensions (Estimated degree of freedom is the mean of the data)

n	r	H_0 vs H_A	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
100	100	5.10	$\chi^2(3)$	2.99529	0.7663 0.8821	5.8: (a)
100	1000	5.10	$\chi^2(3)$	3.1027	0.5678 0.2894	5.8: (b)
100	10,000	5.10	$\chi^2(3)$	2.988	0.0198 0.1381	5.8: (c)
100	100	5.11	$\chi^2(1)$	1.07655	0.2761 0.1851	5.8: (d)
100	1000	5.11	$\chi^2(1)$	1.02524	0.5408 0.3472	5.8: (e)
100	10,000	5.11	$\chi^2(1)$	1.00578	0.0454 0.0178	5.8: (f)
100	100	5.12	$\chi^2(2)$	2.08681	0.0569 0.2650	5.8: (g)
100	1000	5.12	$\chi^2(2)$	2.0819	0.1885 0.1473	5.8: (h)
100	10,000	5.12	$\chi^2(2)$	1.98061	0.2269 0.2253	5.8: (i)
1000	100	5.10	$\chi^2(3)$	2.86446	0.1172 0.1832	5.9: (a)
1000	1000	5.10	$\chi^2(3)$	2.96763	0.8382 0.6667	5.9: (b)
1000	10,000	5.10	$\chi^2(3)$	2.99782	0.8155 0.8299	5.9: (c)
1000	100	5.11	$\chi^2(1)$	1.12204	0.1024 0.1808	5.9: (d)
1000	1000	5.11	$\chi^2(1)$	1.01188	0.3671 0.4461	5.9: (e)
1000	10,000	5.11	$\chi^2(1)$	0.998095	0.7568 0.8312	5.9: (f)
1000	100	5.12	$\chi^2(2)$	2.10698	0.1527 0.5505	5.9: (g)
1000	1000	5.12	$\chi^2(2)$	1.99368	0.3234 0.2004	5.9: (h)
1000	10,000	5.12	$\chi^2(2)$	1.98171	0.5068 0.5678	5.9: (i)

Table 5.2: Continuation of Table 5.1

n	r	H_0 vs H_A	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
10,000	100	5.10	$\chi^2(3)$	3.03235	0.9745 0.9963	5.10: (a)
10,000	1000	5.10	$\chi^2(3)$	2.96816	0.7401 0.8844	5.10: (b)
10,000	10,000	5.10	$\chi^2(3)$	2.98665	0.6104 0.8418	5.10: (c)
10,000	100	5.11	$\chi^2(1)$	1.04509	0.7297 0.3549	5.10: (d)
10,000	1000	5.11	$\chi^2(1)$	0.992251	0.9087 0.9782	5.10: (e)
10,000	10,000	5.11	$\chi^2(1)$	0.996367	0.8203 0.9819	5.10: (f)
10,000	100	5.12	$\chi^2(2)$	1.95931	0.9742 0.9492	5.10: (g)
10,000	1000	5.12	$\chi^2(2)$	2.00737	0.8501 0.9219	5.10: (h)
10,000	10,000	5.12	$\chi^2(2)$	1.98817	0.1386 0.1899	5.10: (i)

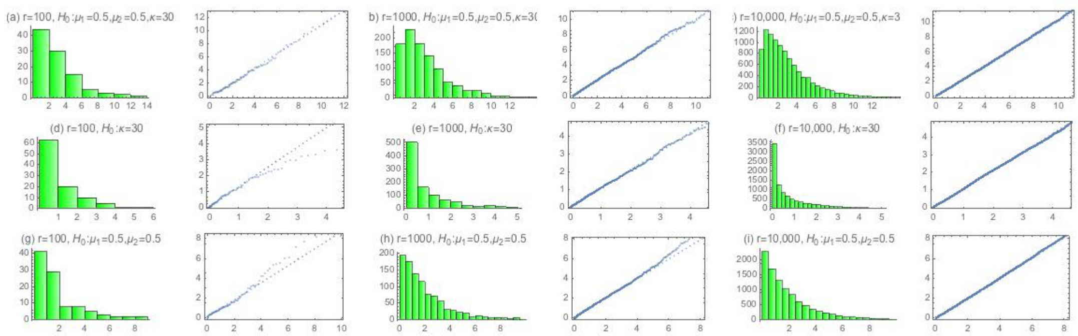

Figure 5.8: Histograms and quantile plots for the hypothesis tests in Table 5.1 for $n = 100$ samples on the sphere in 3 dimensions

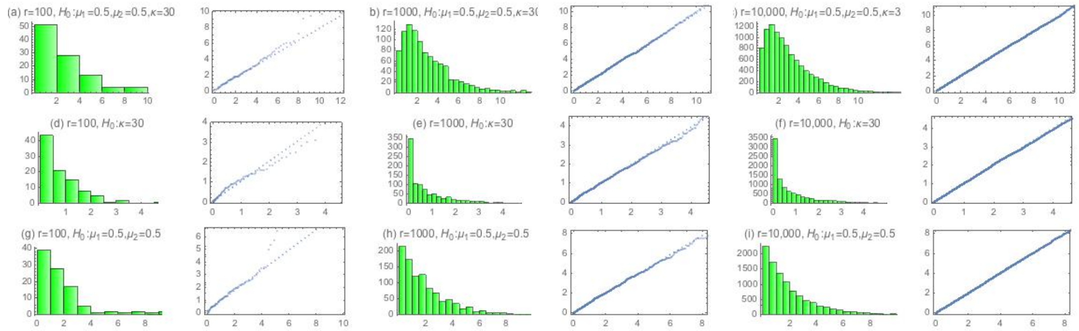

FIGURE 5.9: Histograms and quantile plots for the hypothesis tests in Table 5.1 for $n = 1000$ samples on the sphere in 3 dimensions

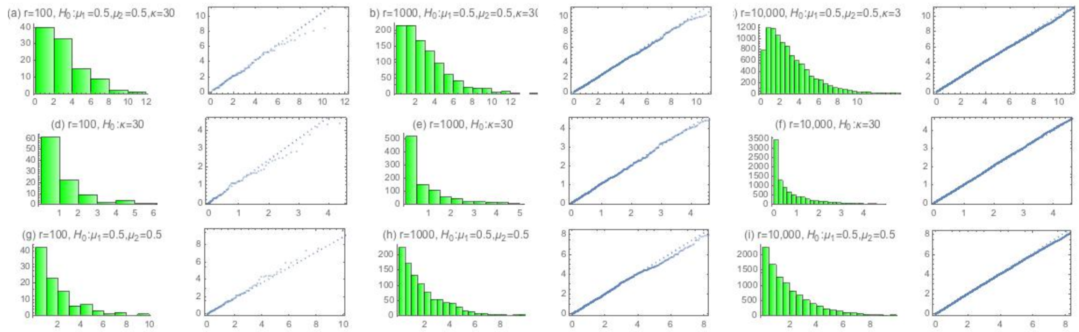

FIGURE 5.10: Histograms and quantile plots for the hypothesis tests in Table 5.2 for $n = 10,000$ samples on the sphere in 3 dimensions

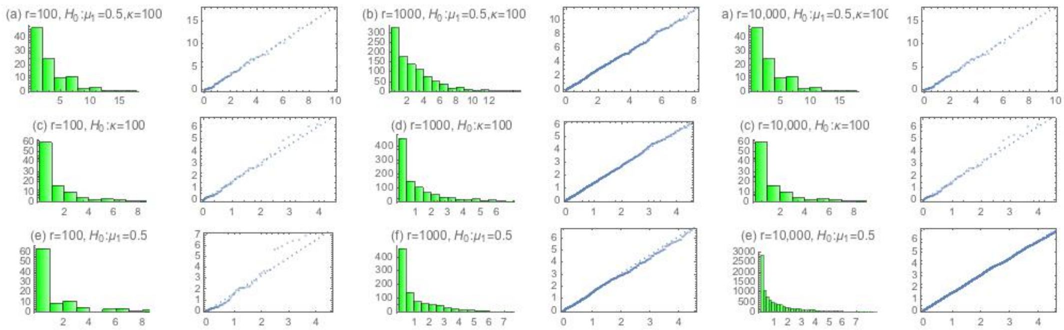

FIGURE 5.11: Histograms and quantile plots for the hypothesis tests in Table 5.3 for $n = 1000$ samples on the sphere in 2 dimensions

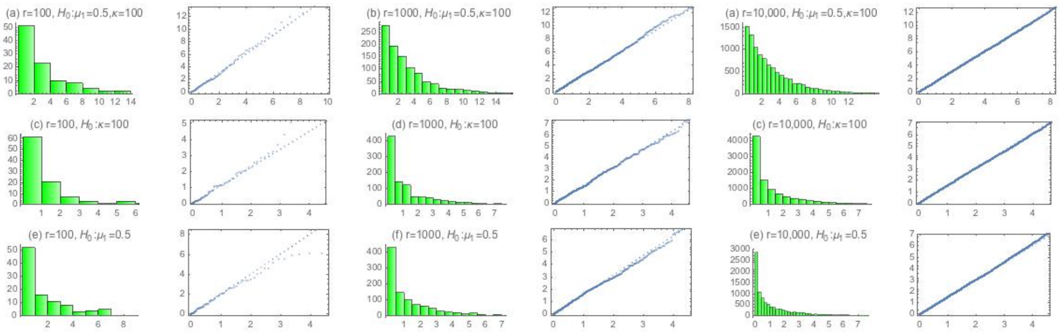

FIGURE 5.12: Histograms and quantile plots for the hypothesis tests in Table 5.3 for $n = 10,000$ samples on the sphere in 2 dimensions

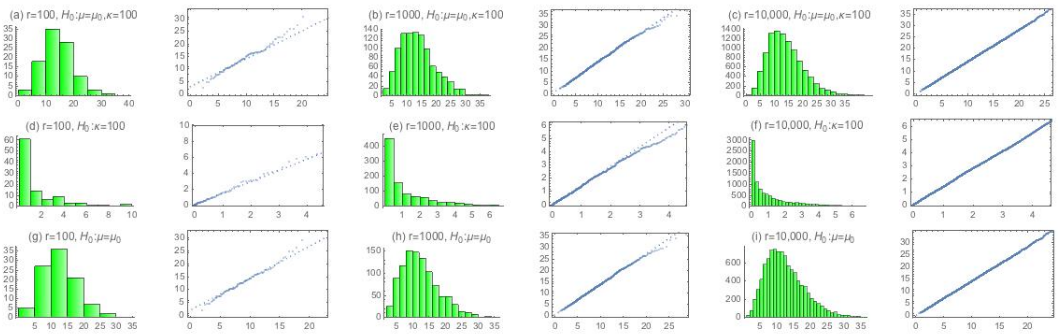

FIGURE 5.13: Histograms and quantile plots for the hypothesis tests in Table 5.4 for $n = 1000$ samples on the sphere in 10 dimensions

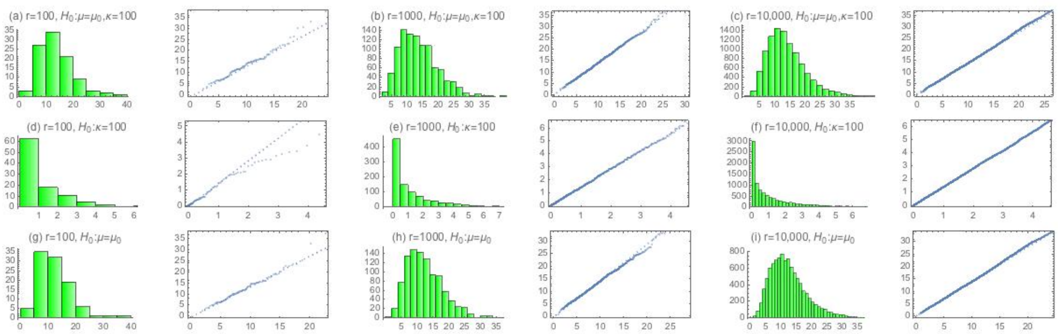

FIGURE 5.14: Histograms and quantile plots for the hypothesis tests in Table 5.4 for $n = 10,000$ samples on the sphere in 10 dimensions

TABLE 5.3: Simulation results for hypothesis tests (5.13), (5.14), and (5.15) in 2 dimensions, using the sample trimmed mean of 0.25 to estimate the degrees of freedom

n	r	H_0 vs H_A	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
1000	100	5.13	$\chi^2(2)$	1.42841	0 0	5.11: (a)
1000	1000	5.13	$\chi^2(2)$	1.66825	0 0	5.11: (b)
1000	10,000	5.13	$\chi^2(2)$	1.62989	0 0	5.11: (c)
1000	100	5.14	$\chi^2(1)$	0.83244	0 0	5.11: (d)
1000	1000	5.14	$\chi^2(1)$	0.991795	0 0	5.11: (e)
1000	10,000	5.14	$\chi^2(1)$	0.940759	0 0	5.11: (f)
1000	100	5.15	$\chi^2(1)$	0.793779	0 0	5.11: (g)
1000	1000	5.15	$\chi^2(1)$	0.921688	0 0	5.11: (h)
1000	10,000	5.15	$\chi^2(1)$	0.944921	0 0	5.11: (i)
10,000	100	5.13	$\chi^2(2)$	1.7458	0.0498 0.0185	5.12: (a)
10,000	1000	5.13	$\chi^2(2)$	1.76804	0 0	5.12: (b)
10,000	10,000	5.13	$\chi^2(2)$	1.67718	0 0	5.12: (c)
10,000	100	5.14	$\chi^2(1)$	1.2946	0.6055 0.4657	5.12: (d)
10,000	1000	5.14	$\chi^2(1)$	1.02477	0 0	5.12: (e)
10,000	10,000	5.14	$\chi^2(1)$	0.988486	0 0	5.12: (f)
10,000	100	5.15	$\chi^2(1)$	0.715621	0.0040 0.0023	5.12: (g)
10,000	1000	5.15	$\chi^2(1)$	1.03058	0 0	5.12: (h)
10,000	10,000	5.15	$\chi^2(1)$	0.978061	0 0	5.12: (i)

Table 5.4: Simulation results for hypothesis tests (5.16), (5.17) and (5.18) in 10 dimensions, using the sample trimmed mean of 0.25 to estimate the degrees of freedom

n	r	H_0 vs H_A	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
1000	100	5.16	$\chi^2(10)$	10.5886	0 0	5.13: (a)
1000	1000	5.16	$\chi^2(10)$	11.0025	0 0	5.13: (b)
1000	10,000	5.16	$\chi^2(10)$	11.2523	0 0	5.13: (c)
1000	100	5.17	$\chi^2(1)$	1.12281	0.2011 0.1473	5.13: (d)
1000	1000	5.17	$\chi^2(1)$	0.926188	0 0	5.13: (e)
1000	10,000	5.17	$\chi^2(1)$	0.887687	0 0	5.13: (f)
1000	100	5.18	$\chi^2(9)$	9.37858	0 0	5.13: (g)
1000	1000	5.18	$\chi^2(9)$	9.66806	0 0	5.13: (h)
1000	10,000	5.18	$\chi^2(9)$	9.97305	0 0	5.13: (i)
10,000	100	5.16	$\chi^2(10)$	12.2731	0 0	5.14: (a)
10,000	1000	5.16	$\chi^2(10)$	11.3179	0 0	5.14: (b)
10,000	10,000	5.16	$\chi^2(10)$	11.2056	0 0	5.14: (c)
10,000	100	5.17	$\chi^2(1)$	0.798673	0.3638 0.4796	5.14: (d)
10,000	1000	5.17	$\chi^2(1)$	0.872307	0 0	5.14: (e)
10,000	10,000	5.17	$\chi^2(1)$	0.872739	0 0	5.14: (f)
10,000	100	5.18	$\chi^2(9)$	11.0776	0 0	5.14: (g)
10,000	1000	5.18	$\chi^2(9)$	10.0918	0 0	5.14: (h)
10,000	10,000	5.18	$\chi^2(9)$	9.9154	0 0	5.14: (i)

TABLE 5.5: Simulation results for the distribution of d_n in hypothesis tests for the spherical subcomponent model and for $r = 100$; here i is the dimension and m_i is the number of replications of the von Mises Fisher distribution in the spherical subcomponent model

i	m_i	n	H_0 vs H_A	r	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
		n_{31} =1000						
3	3	n_{32} =1000	(5.21)	100	$\chi^2(2)$	2.02027	0.9919	5.15: (a)
		n_{33} =1000					0.9944	
		n_{21} =1000						
2	3	n_{22} =1000						
		n_{23} =1000	(5.19)	100	$\chi^2(2)$	2.69724	0.0181	5.15: (c)
3	1	n_{31} =1000					0	
2	1	n_{21} =1000	(5.20)	100	$\chi^2(1)$	0.970214	0.1005	5.15: (e)
3	1	n_{31} =1000					0.5596	
		n_{21} =1000						
		n_{22} =1000						
2	6	n_{23} =1000						
		n_{24} =1000						
		n_{25} =1000						
		n_{26} =1000	(5.23)	100	$\chi^2(5)$	6.18541	0.0048	5.15: (g)
		n_{31} =1000					0	
3	4	n_{32} =1000						
		n_{33} =1000						
		n_{34} =1000						
4	1	n_{41} =1000						
		n_{21} =1000						
		n_{22} =1000						
2	6	n_{23} =1000						
		n_{24} =1000						
		n_{25} =1000						
		n_{26} =1000	(5.24)	100	$\chi^2(3)$	2.92367	0.8864	5.15: (i)
		n_{31} =1000					0.8193	
3	4	n_{32} =1000						
		n_{33} =1000						
		n_{34} =1000						
4	1	n_{41} =1000						

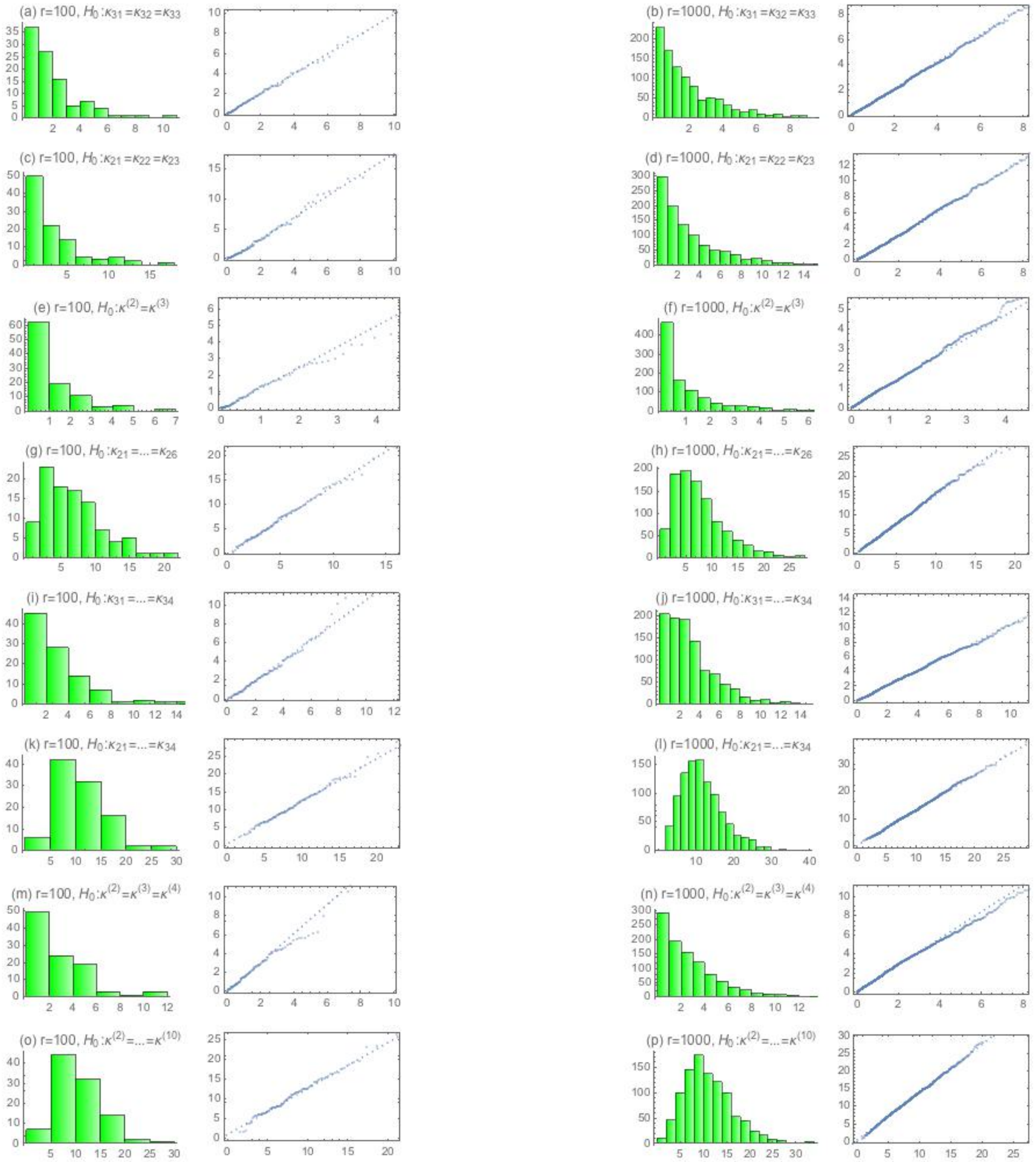

FIGURE 5.15: Histograms and quantile plots for the hypothesis tests in the spherical subcomponent model represented in Tables 5.5, 5.6, 5.7, and 5.8

Table 5.6: Continuation of Table 5.5

i	m_i	n	H_0 vs H_A	r	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
2	6	n_{21} =1000 n_{22} =1000 n_{23} =1000 n_{24} =1000 n_{25} =1000 n_{26} =1000	(5.25)	100	$\chi^2(9)$	11.1108	0.0167 0	5.15: (k)
3	4	n_{31} =1000 n_{32} =1000 n_{33} =1000 n_{34} =1000						
4	1	n_{41} =1000						
2 3 4	1 1 1	n_{21} =1000 n_{31} =1000 n_{41} =1000	(5.26)	100	$\chi^2(2)$	2.63738	0.0708 0.0079	5.15: (m)
2 3 4 5 6 7 8 9 10	1 1 1 1 1 1 1 1 1	n_{21} =1000 n_{31} =1000 n_{41} =1000 n_{51} =1000 n_{61} =1000 n_{71} =1000 n_{81} =1000 n_{91} =1000 n_{101}=1000	(5.27)	100	$\chi^2(8)$	10.2972	0 0	5.15: (o)

TABLE 5.7: Simulation results for the distribution of d_n in hypothesis tests for the spherical subcomponent model and for $r = 1000$; here i is the dimension and m_i is the number of replications of the von Mises Fisher distribution in the spherical subcomponent model

i	m_i	n	H_0 vs H_A	r	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
3	3	n_{31} =1000 n_{32} =1000 n_{33} =1000	(5.21)	1000	$\chi^2(2)$	1.97211	0.8289 0.7939	5.15: (b)
2 3	3 1	n_{21} =1000 n_{22} =1000 n_{23} =1000 n_{31} =1000	(5.19)	1000	$\chi^2(2)$	2.55994	0 0	5.15: (d)
2 3	1 1	n_{21} =1000 n_{31} =1000	(5.20)	1000	$\chi^2(1)$	1.08671	0.0127 0.0011	5.15: (f)
2 3 4	6 4 1	n_{21} =1000 n_{22} =1000 n_{23} =1000 n_{24} =1000 n_{25} =1000 n_{26} =1000 n_{31} =1000 n_{32} =1000 n_{33} =1000 n_{34} =1000 n_{41} =1000	(5.23)	1000	$\chi^2(5)$	7.06293	0 0	5.15: (h)
2 3 4	6 4 1	n_{21} =1000 n_{22} =1000 n_{23} =1000 n_{24} =1000 n_{25} =1000 n_{26} =1000 n_{31} =1000 n_{32} =1000 n_{33} =1000 n_{34} =1000 n_{41} =1000	(5.24)	1000	$\chi^2(3)$	3.05996	0.0755 0.1621	5.15: (j)

Table 5.8: Continuation of Table 5.7

i	m_i	n	H_0 vs H_A	r	Asymptotic Distribution of d_n	Estimated degree of freedom	P-Values: Watson-U^2 Kolm.-Smir.	Figure
		n_{21} =1000						
		n_{22} =1000						
2	6	n_{23} =1000						
		n_{24} =1000						
		n_{25} =1000						
		n_{26} =1000	(5.25)	1000	$\chi^2(9)$	11.3965	0	5.15: (l)
		n_{31} =1000					0	
3	4	n_{32} =1000						
		n_{33} =1000						
		n_{34} =1000						
4	1	n_{41} =1000						
2	1	n_{21} =1000						
3	1	n_{31} =1000	(5.26)	1000	$\chi^2(2)$	2.5324	0	5.15: (n)
4	1	n_{41} =1000					0	
2	1	n_{21} =1000						
3	1	n_{31} =1000						
4	1	n_{41} =1000						
5	1	n_{51} =1000						
6	1	n_{61} =1000						
7	1	n_{71} =1000	(5.27)	1000	$\chi^2(8)$	10.8239	0	5.15: (p)
8	1	n_{81} =1000					0	
9	1	n_{91} =1000						
10	1	n_{101}=1000						

Chapter 6

Discussion

6.1 Overview

In this thesis asset allocations in financial portfolios are used to motivate data analysis using spherical subcomponent models. After transformation, the data can be regarded as spherical data of different dimensions and a composite likelihood approach is used to model this.

In our application, the composite likelihood is composed of von Mises Fisher distributions from different dimensions. As is shown in Lindsay (1988), page 224, the maximum estimator leads to a consistent method of estimation under appropriate assumptions. Construction of the composite likelihood requires the combination of marginal or conditional density functions as described in Example 2A, page 224 of Lindsay.

Standard statistical methods require the expected value of the score function vector to be zero. However, in the von Mises Fisher distribution the expected value of the score function vector is not zero and the variance matrix of the score function vector is not equal to the expectation of the square of the score function vector. Furthermore, the expectation of the negative second derivative matrix of the estimating function is a singular matrix. Consequently, the "information matrix" concept is unhelpful for the von Mises Fisher distribution. Also, one of the main assumptions in Vu and Zhou (1997) about the positive definiteness of the negative expectation of the second derivative of the log likelihood function is not satisfied in this distribution. It might be emphasized here that eliminating one parameter and substituting it with other parameters in the von Mises Fisher distribution removes these concerns and standard assumptions are satisfied. However, such an approach in the spherical subcomponent model would lead to very unwieldy expressions. Therefore, introducing a methodology that encompasses the singularity of the second derivative matrix is useful. The required methodology was developed by Maller (2015) and is used throughout this thesis.

In summary, the focus of this thesis is to extend Maller's methodology and use composite likelihood ideas to conduct asymptotic analysis for spherical subcomponent models. As a motivating example we analyse asset allocation data, when a von Mises Fisher distribution is assumed to fit the data. The result of the hypothesis tests for the equality of the concentration parameters illustrates the possibilities of a general approach. The analysis of dimension 2 in the spherical subcomponent model consisting of three 2 dimensional and one 3 dimensional von Mises Fisher distributions shows that there is no significant difference among them for S&P500, NIKKEI225 and AllOrdinaries indices. However, when using an equicorrelation model to estimate the covariance matrix in building the portfolios, we find a significant difference between the concentration parameters of the dimension 2 components at the 0.05 level. For the ten indices introduced in Table 4.2, the hypothesis test for the equality of the concentration parameters in the 9 different von Mises Fisher distributions from dimensions 2 to 10 shows that there is a significant difference between these concentration parameters at the 0.05 level.

In initiating this investigation, the work of a number of other researchers including Chernoff (1954), Self and Liang (1987), Silvapulle and Sen (2005), Silvey (1959), Watson (1983), and Vu and Zhou (1997), etc. was reviewed. Chernoff's result on the asymptotic distribution of the deviance statistic was generalised later by Silvapulle and Sen and others in more complex situations, as in Self and Liang's approach to boundary problems. We found Silvey's solution to the problem of the singularity of the second derivative of the log likelihood function very practical. In Silvey's approach the negative expectation of the second derivative of the log likelihood function is equal to the variance matrix of the score function vector, therefore an "information matrix" is meaningful in this setup. Watson's geometrical approach to spherical distributions, especially the wide range of formulae he introduced for the statistical analysis on the sphere, provided basic foundations. Some other useful properties of the von Mises Fisher distribution via its distribution function, which I worked out, were presented in Chapter 2 of the thesis.

Checking the methodology of the theory using simulation is logical and common in the statistical framework. The simulation method introduced by Wood (1994) has been adopted by many statistical researchers who work in this field. I used his method in my investigations. Fortunately, the 3 dimensional von Mises distribution has an explicit distribution function that makes the simulation method reliable. Our theory was well confirmed by the results of simulations of 3 dimensional distributions.

To conclude the thesis we use the data on 10 financial indices to suggest how our results might be used in practical portfolio analysis.

6.2 Portfolio Analysis

Chapter 3 of the thesis was totally devoted to introduce a methodology for the data that were in the form of the subcomponent model; that is having different dimensions in one set of data. Also, in Chapter 4 we saw that spherical data produced by asset allocations of different indices are categorised in the spherical subcomponent model. We considered SP500, NIKKEI225 and AllOrdinaries as 3 indices, obtained their daily returns from Yahoo Finance website and did data analysis based on the methodology of Chapter 3 of the thesis. In Section 6.2.1, we choose other indices and conduct certain hypothesis tests to facilitate the process of making decisions to select the most profitable portfolio in 3 dimensions given certain risks.

We chose the von Mises Fisher distribution as a model for the data in order to illustrate the methodology. As it turned out (see Section 4.4) this distribution does not fit the data too well and for future analysis the use of some other spherical distributions should be considered. But such investigations are not within the scope of this thesis. For the purposes of this discussion we continue to consider the von Mises Fisher to illustrate the way the analysis could be used.

Tables 6.3 to 6.10 show that there is not a substantial difference among different selections in 3 dimensions. This judgment is based on the values of W_n and d_n. As we see, for most but not all values of W_n, the null hypothesis of fitting a von Mises Fisher distribution to the data is not rejected and the values of d_n accept the null hypothesis of equality of κs. In the 10 dimensional analysis, where we choose 10 different indices and have data in 9 different dimensions from 2 to 10, the comparison is more complex. We first suggest to conduct a hypothesis test for the equality of the κs and when there is a significant difference amongst the κs, we choose the dimension of the portfolio which is related to the smallest κ. After choosing the dimension with the smallest κ, it is suggested to have another test to select the ultimate indices and their asset allocations in order to finalise the process of decision making. Table 6.2 can be considered for the case when we do not apply a spherical subcomponent model and we assume a 10 dimensional von Mises Fisher distribution to the data.

6.2.1 3 Dimensional Portfolio Analysis

Tables 6.3 to 6.10 show different indices in 3 dimensions in which the total number of monthly square roots of the asset allocations for each 3 dimensions is given and a spherical subcomponent model is applied to them. The daily historical prices of each index was obtained from Yahoo Finance website for their maximum available date untill

28 April 2017. After the calculation of asset allocations and subsequently the spherical data, Mardia et al's goodness of fit test for the von Mises Fisher distribution against Fisher Bingham distribution for each subcomponent was done and W_n was calculated. When the dimension is 3, W_n is distributed as a chi square distribution with 5 degrees of freedom and for the 2 dimensions, the degrees of freedom is 2. The asterisk next to each W_n in Tables 6.3 to 6.10 shows that the null hypothesis assuming that the related data are from a von Mises Fisher distribution is not rejected at 0.05 level; two asterisks show that the null hypothesis is not rejected at 0.01 level.

The d_n in the last column of Tables 6.3 to 6.10 is the deviance statistic for the hypotheses

$$\begin{cases} H_0 & : \kappa_1 = \kappa_2 = \kappa_3 \\ H_A & : \text{at least one equality is not satisfied.} \end{cases}$$

The parameters κ_1, κ_2 and κ_3 are the concentration parameters in the 2 dimensional subcomponent model where a 2 dimensional von Mises Fisher distribution is assumed to fit the data. A single asterisk next to the value of d_n in these tables shows that the null hypothesis of the equality of κs is not rejected at 0.05 level. As we see, all of the d_n values, except the first one, have a single asterisk which shows that the null hypothesis of the equality of the κs is not rejected at 0.05 level. For the first portfolio, which is a combination of SP500, Bond 20^+ and NASDAQ, the null hypothesis of the equality of κs is rejected, so we suggest that the 2 dimensional portfolio combined of SP500 and Bond 20^+ has the lowest risk for the reason of having the smallest κ. Therefore, it is suggested to invest the ratio of $0.6717^2 = 0.4512$ of our capital on SP500 and $0.7407^2 = 0.5488$ on Bond 20^+ in order to have the most profit with a certain amount of risk.

6.2.2 10 Dimensional Portfolio Analysis

Table 6.1 shows 10 different indices selected from Yahoo Finance website. From this site we obtained the daily historical prices from maximum available date of each index untill 28 April 2017. Monthly asset allocations for these indices are calculated and the square roots of them result in the spherical data. The total number of data are 294 and the MLEs for the parameters in the 10 dimensional von Mises Fisher distribution are

$$\begin{aligned} \widehat{\kappa} &= 7.58833 \\ \widehat{\boldsymbol{\mu}}^T &= (0.253804, 0.474701, 0.287676, 0.381364, 0.24742, 0.179443, 0.14135, \\ &\quad 0.484017, 0.320999, 0.177025). \end{aligned}$$

TABLE 6.1: Ten different indices to form a portfolio based on daily returns in Yahoo Finance website before 28 April 2017

Orders In Portfolios	**Name**	**Country**	**Abb.**	**asteriskt Date**	**Number of Data**
1	DAX	Germany	^GDAXI	26/Nov/1990	6678
2	NASDAQ	U.S.	^IXIC	5/Feb/1971	11658
3	HSI	China	^HSI	31/Dec/1986	7492
4	DOW	U.S.	^DJI	29/Jan/1985	8135
5	NIKKEI225	Japan	$^N225$	4/Jun/1984	8192
6	SP/ASX 200	Australia	^AXJO	23/Nov/1992	6179
7	SP500	U.S.	^GSPC	3/Jan/1950	16939
8	Volatility SP500	China	^VIX	2/Jan/1990	6890
9	AllOrdinaries	Australia	^AORD	3/Aug/1984	8274
10	CAC40	France	^FCHI	1/Mar/1990	6883

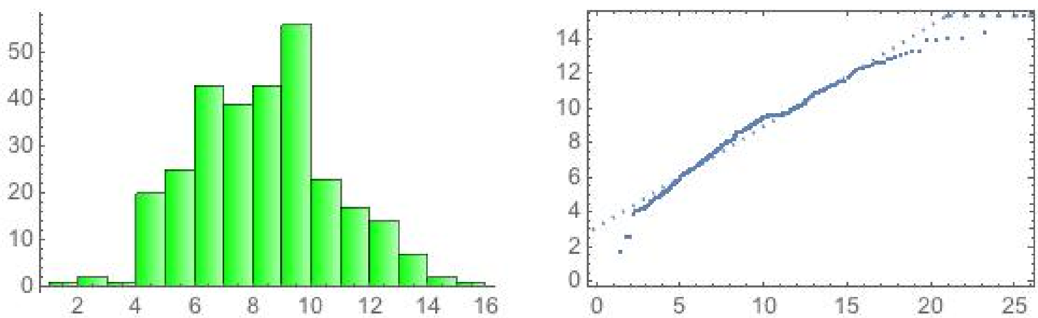

FIGURE 6.1: Watson's goodness of fit test for the 10 dimensional indices represented in Table 6.1; $r_q = 0.9857$

The goodness of fit test for the 10 dimensional spherical data, calculated for the 10 indices in Table 6.1, where the null hypothesis considers the data having a 10 dimensional von Mises Fisher distribution, was applied via Watson's method explained in Section 4.4.1.3. The results are shown in Figure 6.1.

In Table 6.2 a set of hypothesis tests for the parameters of a 10 dimensional von Mises Fisher distribution is considered and is conducted for the 10 dimensional spherical data calculated for the indices presented in Table 6.1. The single asterisk next to the value of d_ns show that the related null hypothesis is not rejected at 0.05 level. According to these results, the null hypothesis in

$$\begin{cases} H_0 & : \mu_1 = 0.25, \mu_2 = 0.5, \mu_3 = 0.25, \mu_4 = 0.35, \mu_5 = 0.25, \mu_6 = 0.15, \\ & \mu_7 = 0.15, \mu_8 = 0.5, \mu_9 = 0.35, \mu_{10} = 0.15 \\ H_A & : \text{at least one equality is not satisfied,} \end{cases}$$

is not rejected at 0.05 level.

Consider the hypothesis test

$$\begin{cases} H_0 & : \kappa^{(2)} = \kappa^{(3)} = \ldots = \kappa^{(10)} \\ H_A & : \text{at least one equality is not satisfied,} \end{cases} \tag{6.1}$$

where $\kappa^{(i)}$, for $i = 2, 3, \ldots, 10$, is the concentration parameter of the i dimensional von Mises Fisher distribution in the spherical subcomponent model combined of 10 indices. For the 10 indices in Table 6.1, the value of the deviance statistic d_n for this hypothesis test is 53.2406. The distribution of d_n is a chi square with 8 degrees of freedom, so the null hypothesis in (6.1) is rejected at 0.05 level.

In each subcomponent, the value of $\widehat{\kappa}$ is

$$\begin{aligned} &\widehat{\kappa}^{(2)} = 4.10167, && \widehat{\kappa}^{(3)} = 7.3811, && \widehat{\kappa}^{(4)} = 6.38826, \\ &\widehat{\kappa}^{(5)} = 8.76313, && \widehat{\kappa}^{(6)} = 10.2545, && \widehat{\kappa}^{(7)} = 10.5057, \\ &\widehat{\kappa}^{(8)} = 7.39954, && \widehat{\kappa}^{(9)} = 26.554, && \widehat{\kappa}^{(10)} = 33.771 \end{aligned}$$

where the number of data for each dimension is

$$\begin{aligned} &n_2 = 9, && n_3 = 11, && n_4 = 22, \\ &n_5 = 22, && n_6 = 35, && n_7 = 23, \\ &n_8 = 10, && n_9 = 6, && n_{10} = 6. \end{aligned}$$

As we see the value of κ for the 2 dimensional data is the smallest one amongst the other dimensions and it is suggested to have a 2 dimensional portfolio in order to have the highest profit given a certain risk.

The 9 observations on the 2 dimensional square roots of the asset allocations are

$$\begin{array}{llllllllll}
\{0.0003174, & 0.9554, & 0.0002174, & 0.0007675, & 0.000328, & 0.0007119, & 0.000366, & 0.2952, & 0.000374, & 0.000386\}, \\
\{0.9381, & 0.000795, & 0.0005039, & 0.0009388, & 0.0006857, & 0.0009483, & 0.000701, & 0.3462, & 0.000725, & 0.000635\}, \\
\{0.0003820, & 0.901864, & 0.0004846, & 0.0009381, & 0.0005030, & 0.0008131, & 0.000644, & 0.43201, & 0.000655, & 0.000354\}, \\
\{0.0003923, & 0.001422, & 0.0001962, & 0.0005054, & 0.0003742, & 0.999998, & 0.000405, & 0.00048, & 0.000379, & 0.000255\}, \\
\{0.0002966, & 0.921164, & 0.0002446, & 0.0003665, & 0.0002687, & 0.0004207, & 0.000390, & 0.38917, & 0.000412, & 0.000314\}, \\
\{0.0004530, & 0.893623, & 0.0005528, & 0.0008918, & 0.0008489, & 0.0009938, & 0.000673, & 0.44881, & 0.000686, & 0.000541\}, \\
\{0.0004224, & 0.999997, & 0.0008667, & 0.0009343, & 0.0009998, & 0.0009804, & 0.000657, & 0.00116, & 0.000655, & 0.000414\}, \\
\{0.0005311, & 0.000331, & 0.0006572, & 0.0007335, & 0.0004424, & 0.0005689, & 0.001408, & 0.00034, & 0.9999, & 0.000696\}, \\
\{0.0005063, & 0.000264, & 0.0002477, & 0.000339, & 0.0001537, & 0.0004971, & 0.000982, & 0.00019, & 0.9492, & 0.3143\}.
\end{array} \tag{6.2}$$

We suggest taking an allocation to be zero if it is less than 0.001. In (6.2), in each row, there are only two cases with values greater than zero or precisely greater than 0.001. We proceed by conducting the hypothesis test of equality of κs for the 2 dimensional case and then, if there is a significant difference among κs, we suggest selecting the one with the smallest κ. In the 10 dimensional case, there are in total 45 possible 2 dimensional portfolios, so the equality of 45 κs should be checked. In our specific example with the indices in Table 6.1, the number of observations is too small to do this. However, we can find an interesting pattern in the outcome. As we see from (6.2), we have only 9 observations with 5 of them indicating to select the second and eighth indices as the target portfolios. From Table 6.1, we gather that the proposed indices by this methodology are NASDAQ from U.S. and Volatility SP500 from China, with approximately $0.9^2 = 0.81$ portion of the capital to be invested in NASDAQ and the rest in Volatility SP500.

TABLE 6.2: Testing some hypotheses for the 10 dimensional spherical data calculated for the 10 indices in Table 6.1 when assuming the data to come from a von Mises Fisher distribution

H_0	$\sup_{\theta\in\Omega}\mathcal{L}_n(\theta)$	d_n	dis. of d_n
$\mu_1 = 0$	-470.157	78.3379	$\chi^2(1)$
$\mu_2 = 0$	-563.916	265.855	$\chi^2(1)$
$\mu_3 = 0$	-481.131	100.286	$\chi^2(1)$
$\mu_4 = 0$	-518.062	174.149	$\chi^2(1)$
$\mu_5 = 0$	-468.234	74.4926	$\chi^2(1)$
$\mu_6 = 0$	-450.691	39.4063	$\chi^2(1)$
$\mu_7 = 0$	-443.243	24.5106	$\chi^2(1)$
$\mu_8 = 0$	-568.959	275.942	$\chi^2(1)$
$\mu_9 = 0$	-493.177	124.378	$\chi^2(1)$
$\mu_{10} = 0$	-450.167	38.3581	$\chi^2(1)$
$\mu_6 = \mu_7 = \mu_{10} = 0.15$	-432.06	2.1431*	$\chi^2(2)$
$\mu_6 = \mu_7 = \mu_{10} = 0.2$	-433.914	5.8509*	$\chi^2(2)$
$\mu_1 = \mu_3 = \mu_5$	-431.565	1.1528*	$\chi^2(3)$
$\mu_1 = \mu_3 = \mu_5 = 0.25$	-431.958	1.9405*	$\chi^2(2)$
$\mu_1 = \mu_3 = \mu_5 = 0.30$	-434.856	7.7355	$\chi^2(2)$
$\mu_2 = \mu_4 = \mu_8$	-434.952	7.9282	$\chi^2(3)$
$\mu_2 = \mu_8$	-431.015	0.05344*	$\chi^2(2)$
$\mu_2 = \mu_4 =$ $\mu_8 = \mu_9$	-442.214	22.4508	$\chi^2(4)$
$\mu_2 = 0.50$	-431.505	1.0335*	$\chi^2(1)$
$\mu_8 = 0.50$	-431.196	0.4151*	$\chi^2(1)$
$\mu_2 = \mu_8 = 0.5$	-432.022	2.0672*	$\chi^2(1)$
$\mu_2 = 0.55$	-435.71	9.4428	$\chi^2(1)$
$\mu_4 = \mu_9 = 0.35$	-432.113	2.2505*	$\chi^2(1)$
$\mu_1 = \mu_3 = \mu_5 = 0.25$ $\mu_2 = \mu_8 = 0.5$ $\mu_4 = \mu_9 = 0.35$ $\mu_6 = \mu_7 = \mu_{10} = 0.15$	-434.571	7.1653*	$\chi^2(9)$

Table 6.3: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
SP500 Bond20$^+$ NASDAQ	Total	178	56.56	3.31826	$\widehat{\mu}_1 = 0.5178$ $\widehat{\mu}_2 = 0.6621$ $\widehat{\mu}_3 = 0.5416$	12.4083
	Middle	76	19.36	4.04482	$\widehat{\mu}_1 = 0.5123$ $\widehat{\mu}_2 = 0.7107$ $\widehat{\mu}_3 = 0.4819$	
	S_1	31	11.1455	2.28742	$\widehat{\mu}_1 = 0.6717$ $\widehat{\mu}_2 = 0.7407$	
	S_2	25	0.5199*	6.7073	$\widehat{\mu}_1 = 0.6016$ $\widehat{\mu}_2 = 0.7987$	
	S_3	25	0.7268*	5.3189	$\widehat{\mu}_1 = 0.6004$ $\widehat{\mu}_2 = 0.7996$	
SP500 NIKKEI225 AllOrdinaries	Total	393	82.2962	3.59361	$\widehat{\mu}_1 = 0.6395$ $\widehat{\mu}_2 = 0.5023$ $\widehat{\mu}_3 = 0.5818$	1.6256*
	Middle	131	6.4889*	6.25714	$\widehat{\mu}_1 = 0.6009$ $\widehat{\mu}_2 = 0.5854$ $\widehat{\mu}_3 = 0.5441$	
	S_1	50	3.0876*	4.34964	$\widehat{\mu}_1 = 0.6356$ $\widehat{\mu}_2 = 0.7719$	
	S_2	103	7.0062**	3.94805	$\widehat{\mu}_1 = 0.6959$ $\widehat{\mu}_2 = 0.7180$	
	S_3	29	6.7334**	3.03703	$\widehat{\mu}_1 = 0.8675$ $\widehat{\mu}_2 = 0.4974$	

TABLE 6.4: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
SP500 NIKKEI225 DAX	Total	318	77.3485	3.39169	$\widehat{\mu}_1 = 0.6759$ $\widehat{\mu}_2 = 0.5104$ $\widehat{\mu}_3 = 0.5315$	
	Middle	110	11.1911**	5.14067	$\widehat{\mu}_1 = 0.6098$ $\widehat{\mu}_2 = 0.5646$ $\widehat{\mu}_3 = 0.5561$	
	S_1	49	2.9107*	4.3286	$\widehat{\mu}_1 = 0.7275$ $\widehat{\mu}_2 = 0.6860$	
	S_2	54	5.7769*	3.36106	$\widehat{\mu}_1 = 0.6837$ $\widehat{\mu}_2 = 0.7296$	1.4417*
	S_3	20	1.7963*	3.45741	$\widehat{\mu}_1 = 0.6742$ $\widehat{\mu}_2 = 0.7385$	
SP500 NIKKEI225 HSI	Total	364	88.1747	3.43032	$\widehat{\mu}_1 = 0.6606$ $\widehat{\mu}_2 = 0.5061$ $\widehat{\mu}_3 = 0.5543$	
	Middle	99	16.9434	5.74063	$\widehat{\mu}_1 = 0.5941$ $\widehat{\mu}_2 = 0.5510$ $\widehat{\mu}_3 = 0.5860$	
	S_1	67	3.9499*	4.20977	$\widehat{\mu}_1 = 0.6520$ $\widehat{\mu}_2 = 0.7581$	
	S_2	87	3.8807*	4.24914	$\widehat{\mu}_1 = 0.6482$ $\widehat{\mu}_2 = 0.7614$	1.2756*
	S_3	20	2.6944*	3.15006	$\widehat{\mu}_1 = 0.7640$ $\widehat{\mu}_2 = 0.6451$	

Table 6.5: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
SP500 NIKKEI225 CAC40	Total	326	88.5903	3.3868	$\widehat{\mu}_1 = 0.7094$ $\widehat{\mu}_2 = 0.5069$ $\widehat{\mu}_3 = 0.4896$	
	Middle	103	12.0879**	5.38275	$\widehat{\mu}_1 = 0.6007$ $\widehat{\mu}_2 = 0.6222$ $\widehat{\mu}_3 = 0.5018$	
	S_1	54	5.8815*	3.8135	$\widehat{\mu}_1 = 0.7677$ $\widehat{\mu}_2 = 0.6407$	
	S_2	60	6.8034**	3.38836	$\widehat{\mu}_1 = 0.7501$ $\widehat{\mu}_2 = 0.6612$	1.3747*
	S_3	22	2.5940*	3.15109	$\widehat{\mu}_1 = 0.7177$ $\widehat{\mu}_2 = 0.6962$	

Table 6.6: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
NASDAQ NIKKEI225 AllOrdinaries	Total	393	85.9166	3.57637	$\widehat{\mu}_1 = 0.6792$ $\widehat{\mu}_2 = 0.4593$ $\widehat{\mu}_3 = 0.5724$	
	Middle	123	6.7947*	6.31174	$\widehat{\mu}_1 = 0.6790$ $\widehat{\mu}_2 = 0.5009$ $\widehat{\mu}_3 = 0.5366$	
	S_1	43	1.0864*	5.1208	$\widehat{\mu}_1 = 0.7445$ $\widehat{\mu}_2 = 0.6676$	
	S_2	95	6.7622**	3.84307	$\widehat{\mu}_1 = 0.7573$ $\widehat{\mu}_2 = 0.6530$	1.9843*
	S_3	42	6.4542**	3.5134	$\widehat{\mu}_1 = 0.8667$ $\widehat{\mu}_2 = 0.4986$	
NASDAQ NIKKEI225 DAX	Total	318	89.3749	3.37098	$\widehat{\mu}_1 = 0.6923$ $\widehat{\mu}_2 = 0.4814$ $\widehat{\mu}_3 = 0.5374$	
	Middle	96	10.7742*	5.07745	$\widehat{\mu}_1 = 0.6521$ $\widehat{\mu}_2 = 0.5229$ $\widehat{\mu}_3 = 0.5488$	
	S_1	55	3.1595*	5.30484	$\widehat{\mu}_1 = 0.7690$ $\widehat{\mu}_2 = 0.6391$	
	S_2	54	5.7812*	3.35053	$\widehat{\mu}_1 = 0.7551$ $\widehat{\mu}_2 = 0.6555$	4.2037*
	S_3	29	3.4791*	3.27337	$\widehat{\mu}_1 = 0.6252$ $\widehat{\mu}_2 = 0.7804$	

Table 6.7: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
NASDAQ NIKKEI225 HSI	Total	364	93.6527	3.3501	$\widehat{\mu}_1 = 0.6847$ $\widehat{\mu}_2 = 0.4883$ $\widehat{\mu}_3 = 0.5409$	
	Middle	100	11.6169**	5.27293	$\widehat{\mu}_1 = 0.6185$ $\widehat{\mu}_2 = 0.5379$ $\widehat{\mu}_3 = 0.5727$	
	S_1	64	3.8352*	4.10432	$\widehat{\mu}_1 = 0.7183$ $\widehat{\mu}_2 = 0.6956$	
	S_2	83	5.4670*	3.8882	$\widehat{\mu}_1 = 0.7350$ $\widehat{\mu}_2 = 0.6780$	1.0808*
	S_3	19	3.2673*	2.96613	$\widehat{\mu}_1 = 0.7708$ $\widehat{\mu}_2 = 0.6370$	
NASDAQ NIKKEI225 CAC40	Total	326	101.593	3.31441	$\widehat{\mu}_1 = 0.7097$ $\widehat{\mu}_2 = 0.4912$ $\widehat{\mu}_3 = 0.5049$	
	Middle	96	12.8649**	5.07947	$\widehat{\mu}_1 = 0.6529$ $\widehat{\mu}_2 = 0.5824$ $\widehat{\mu}_3 = 0.4842$	
	S_1	58	5.6190*	4.34954	$\widehat{\mu}_1 = 0.8018$ $\widehat{\mu}_2 = 0.5974$	
	S_2	49	5.2737*	3.41655	$\widehat{\mu}_1 = 0.7575$ $\widehat{\mu}_2 = 0.6528$	1.5246*
	S_3	31	3.6623*	3.2	$\widehat{\mu}_1 = 0.7016$ $\widehat{\mu}_2 = 0.7124$	

TABLE 6.8: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
NIKKEI225 AllOrdinaries DAX	Total	318	85.0197	3.30875]	$\widehat{\mu}_1 = 0.5007$ $\widehat{\mu}_2 = 0.6272$ $\widehat{\mu}_3 = 0.5965$	
	Middle	87	5.9944**	5.3569	$\widehat{\mu}_1 = 0.5349$ $\widehat{\mu}_2 = 0.5382$ $\widehat{\mu}_3 = 0.6512$	
	S_1	42	9.0458**	3.75787	$\widehat{\mu}_1 = 0.9148$ $\widehat{\mu}_2 = 0.4037$	
	S_2	30	1.1124*	4.29465	$\widehat{\mu}_1 = 0.7551$ $\widehat{\mu}_2 = 0.6555$	0.3459*
	S_3	79	4.4845*	4.30217	$\widehat{\mu}_1 = 0.7254$ $\widehat{\mu}_2 = 0.6882$	
NIKKEI225 AllOrdinaries HSI	Total	364	113.817	2.955231	$\widehat{\mu}_1 = 0.5196$ $\widehat{\mu}_2 = 0.6125$ $\widehat{\mu}_3 = 0.5955$	
	Middle	92	19.0363	4.32738	$\widehat{\mu}_1 = 0.5213$ $\widehat{\mu}_2 = 0.5420$ $\widehat{\mu}_3 = 0.6591$	
	S_1	60	9.1440**	3.23623	$\widehat{\mu}_1 = 0.8267$ $\widehat{\mu}_2 = 0.5625$	
	S_2	30	2.4688*	3.52365	$\widehat{\mu}_1 = 0.7211$ $\widehat{\mu}_2 = 0.6927$	0.5276*
	S_3	81	14.0939	2.93703	$\widehat{\mu}_1 = 0.6350$ $\widehat{\mu}_2 = 0.7724$	

Table 6.9: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
NIKKEI225 AllOrdinaries CAC40	Total	326	90.3661	3.2196	$\widehat{\mu}_1 = 0.5076$ $\widehat{\mu}_2 = 0.6281$ $\widehat{\mu}_3 = 0.5896$	
	Middle	93	15.4918	5.05423	$\widehat{\mu}_1 = 0.6329$ $\widehat{\mu}_2 = 0.5132$ $\widehat{\mu}_3 = 0.5795$	
	S_1	38	6.2704**	3.07276	$\widehat{\mu}_1 = 0.8101$ $\widehat{\mu}_2 = 0.5862$	
	S_2	24	1.7895*	3.7916	$\widehat{\mu}_1 = 0.7415$ $\widehat{\mu}_2 = 0.6709$	0.9168*
	S_3	82	6.0744**	3.84648	$\widehat{\mu}_1 = 0.6873$ $\widehat{\mu}_2 = 0.7262$	
AllOrdinaries DAX HSI	Total	318	81.0191	3.27618	$\widehat{\mu}_1 = 0.5729$ $\widehat{\mu}_2 = 0.6126$ $\widehat{\mu}_3 = 0.5443$	
	Middle	100	13.949**	5.17624	$\widehat{\mu}_1 = 0.4950$ $\widehat{\mu}_2 = 0.6517$ $\widehat{\mu}_3 = 0.5745$	
	S_1	65	5.0389*	3.76555	$\widehat{\mu}_1 = 0.6797$ $\widehat{\mu}_2 = 0.7334$	
	S_2	42	6.5697**	3.13185	$\widehat{\mu}_1 = 0.5922$ $\widehat{\mu}_2 = 0.8057$	0.9178*
	S_3	34	2.3327*	4.04882	$\widehat{\mu}_1 = 0.5967$ $\widehat{\mu}_2 = 0.8024$	

TABLE 6.10: A 3 dimensional spherical subcomponent model for different indices

Indices	Category	Number	W_n	$\widehat{\kappa}$	$\widehat{\mu}$	d_n
DAX HSI CAC40	Total	318	103.775	2.99674	$\widehat{\mu}_1 = 0.5774$ $\widehat{\mu}_2 = 0.6634$ $\widehat{\mu}_3 = 0.4757$	
	Middle	105	27.5565	4.12695	$\widehat{\mu}_1 = 0.6121$ $\widehat{\mu}_2 = 0.6939$ $\widehat{\mu}_3 = 0.3791$	
	S_1	49	4.3058*	3.9146	$\widehat{\mu}_1 = 0.5590$ $\widehat{\mu}_2 = 0.8291$	
	S_2	63	19.0588	2.42082	$\widehat{\mu}_1 = 0.6491$ $\widehat{\mu}_2 = 0.7606$	5.2642*
	S_3	25	1.6257*	3.97362	$\widehat{\mu}_1 = 0.6665$ $\widehat{\mu}_2 = 0.7454$	
SP500 DAX HSI	Total	318	77.6297	3.38419	$\widehat{\mu}_1 = 0.6243$ $\widehat{\mu}_2 = 0.5409$ $\widehat{\mu}_3 = 0.5636$	
	Middle	118	14.1241**	4.90115	$\widehat{\mu}_1 = 0.5111$ $\widehat{\mu}_2 = 0.5998$ $\widehat{\mu}_3 = 0.6155$	
	S_1	53	7.7558*	3.02504	$\widehat{\mu}_1 = 0.6609$ $\widehat{\mu}_2 = 0.7503$	
	S_2	55	1.1060*	5.0067	$\widehat{\mu}_1 = 0.6881$ $\widehat{\mu}_2 = 0.7255$	4.2104*
	S_3	20	2.3292*	3.65815	$\widehat{\mu}_1 = 0.6991$ $\widehat{\mu}_2 = 0.7150$	

Appendix A

Maller's notation and theorems

Maller (2015) setup is as follows. Suppose given a sample of n observations on random variables whose distribution depends on a parameter $\boldsymbol{\theta} \in \mathbb{R}^d$. The "true" distribution of the random variables corresponds to the value $\boldsymbol{\theta}_0 \in \mathbb{R}^d$. From the observations we construct an estimating function $\mathcal{L}_n(\boldsymbol{\theta}) : \mathbb{R}^d \mapsto \mathbb{R}$, which is to be maximised, subject to restrictions, to obtain parameter estimates. The parameter $\boldsymbol{\theta}$ is assumed to lie in a subset $\Theta \subset \mathbb{R}^d$ which coincides locally with a cone of the form

$$\Theta := \{\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \widetilde{\theta}_1 \mathbf{u}_1 + \cdots + +\widetilde{\theta}_d \mathbf{u}_d : \widetilde{\theta}_j \in I_j, j = 1, \ldots, d\}, \tag{A.1}$$

having vertex at $\boldsymbol{\theta}_0$, where the $\mathbf{u}_j$ are d linearly independent unit vectors in $\mathbb{R}^d$, and the I_j are closed, half-open or open intervals (subsets of $\mathbb{R}$) containing 0, $j = 1, \ldots, d$.

Throughout, we assume:

(A1) for a neighbourhood $\mathcal{N}$ of $\boldsymbol{\theta}_0$, the function $\mathcal{L}_n(\boldsymbol{\theta})$ has finite first and second directional derivatives on $\Theta \cap \mathcal{N}$. (See Vu-Zhou (1997), pp.900-901, for the definition of directional derivatives, possibly on the boundary of an I_j.)

Under (A1) we can define the first derivative vectors

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{A.2}$$

and the negative second derivative matrices

$$\mathbf{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}, \; n = 1, 2, \ldots, \boldsymbol{\theta} \in \Theta. \tag{A.3}$$

Consider the subset Θ^h of the parameter space Θ in $\mathbb{R}^d$ which is determined by placing s, $1 \leq s \leq d$, restrictions on $\boldsymbol{\theta}$ of the form $h_1(\boldsymbol{\theta}) = 0, h_2(\boldsymbol{\theta}) = 0, \cdots, h_s(\boldsymbol{\theta}) = 0$, where each $h_\ell(\boldsymbol{\theta}) : \mathbb{R}^d \mapsto \mathbb{R}$ is a twice differentiable function, $1 \le \ell \le s$. Write these restrictions as $\mathbf{h}(\boldsymbol{\theta}) = (0, \cdots, 0)$, so that $\Theta^h = \{\boldsymbol{\theta} \in \Theta : h(\boldsymbol{\theta}) = \mathbf{0}\}$, where

$$\mathbf{h}(\boldsymbol{\theta}) = \Big(h_1(\boldsymbol{\theta}) \cdots h_s(\boldsymbol{\theta})\Big)^T_{s\times 1}$$

is a twice differentiable function from $\mathbb{R}^d$ to $\mathbb{R}^s$. The true parameter $\boldsymbol{\theta}_0$ is assumed to lie in Θ^h, so $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$. The first derivative matrix of $\mathbf{H}(\boldsymbol{\theta})$ is denoted by

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial h^T}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial h_1}{\partial \boldsymbol{\theta}} & \cdots & \frac{\partial h_s}{\partial \boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} & \cdots & \frac{\partial h_s}{\partial \theta_1} \\ \vdots & & \vdots \\ \frac{\partial h_1}{\partial \theta_d} & \cdots & \frac{\partial h_s}{\partial \theta_d} \end{pmatrix}_{d\times s}, \tag{A.4}$$

for $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\theta} = (\theta_1 \cdots \theta_d)^T$. By removing redundant restrictions if necessary, without loss of generality we can assume that $\mathbf{H}(\boldsymbol{\theta}_0)$ has rank s.

Following in essence the approach of Aitchison and Silvey (1958) and Silvey (1959), we set out to maximise the Lagrangian function

$$\mathcal{L}_n^\lambda(\boldsymbol{\theta}) := \mathcal{L}_n(\boldsymbol{\theta}) + \boldsymbol{\lambda}_n^T \mathbf{H}(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) + \sum_{\ell=1}^{s} \lambda_\ell(n) h_\ell(\boldsymbol{\theta}), \tag{A.5}$$

where $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\lambda}_n := [\lambda_1(n) \cdots \lambda_s(n)]^T \in \mathbb{R}^s$ are Lagrange multipliers, which may depend on n. Maximisation is to be done subject to

$$h_\ell(\boldsymbol{\theta}) = 0,\ \ell = 1, \ldots, s,\ \boldsymbol{\theta} \in \Theta. \tag{A.6}$$

Let $\mathbf{S}_n^\lambda(\theta)$ be the first derivative vector of $\mathcal{L}_n$:

$$\mathbf{S}_n^\lambda(\boldsymbol{\theta}) := \frac{\partial \mathcal{L}_n^\lambda}{\partial \boldsymbol{\theta}} = \mathbf{S}_n(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\lambda}_n, \tag{A.7}$$

and let $\mathbf{F}_n^\lambda(\boldsymbol{\theta})$ denote the $d \times d$ negative second derivative matrix

$$\mathbf{F}_n^\lambda(\boldsymbol{\theta}) = -\frac{\partial^2 \mathcal{L}_n^\lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \mathbf{F}_n(\boldsymbol{\theta}) - a_n \sum_{\ell=1}^{s} \lambda_\ell \frac{\partial^2 h_\ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \tag{A.8}$$

In some problems, $\mathbf{F}_n^\lambda(\boldsymbol{\theta})$ is not positive definite or even semidefinite, even for $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}_0$. To cater for this, let $\{g_1(n), \cdots, g_{s_n}(n)\}$, $n = 1, 2, \cdots$ be a sequence of (possibly random) positive numbers, whose choice is at our discretion, and set $\mathbf{G}_n := diag(g_1(n), \cdots, g_{s_n}(n))$. Let $\mathbf{H}_n(\boldsymbol{\theta})$ be $d \times s_n$ submatrices of $\mathbf{H}(\boldsymbol{\theta})$, where $1 \le s_n \le s$ and the choice of $\mathbf{H}_n(\boldsymbol{\theta})$ may be random, and augment $\mathbf{F}_n^\lambda(\boldsymbol{\theta})$ to the $d \times d$ symmetric matrices

$$\mathbf{F}_n^{\lambda *}(\boldsymbol{\theta}) = \mathbf{F}_n^\lambda(\boldsymbol{\theta}) + \mathbf{H}_n(\boldsymbol{\theta}) \mathbf{G}_n \mathbf{G}_n^T \mathbf{H}_n^T(\boldsymbol{\theta}). \tag{A.9}$$

The idea is that the augmented matrices will be positive definite when $\boldsymbol{\theta}$ is near Θ^h if the elements of $\mathbf{G}_n$ are chosen sufficiently large. Here we follow Silvey (1959) who treats a simplified version of our setup.

A.1 Existence and Consistency of MEs

Assume given a sequence of $d \times d$ a.s. nonsingular matrices $\mathbf{D}_n$ (possibly random). For each $n = 1, 2, \dots$ and $A > 0$, let $N_n(A)$ and $N_n^h(A)$ be the (in general, random) neighbourhoods

$$N_n(A) = \{\theta \in \Theta : (\theta - \theta_0)^T \mathbf{D}_n \mathbf{D}_n^T (\theta - \theta_0) \le A^2\} \quad \text{and} \quad N_n^h(A) = N_n(A) \cap \Theta^h. \tag{A.10}$$

Theorem A.1. *Assume (A1) and in addition the* $\mathbf{D}_n$ *in* (A.10) *satisfy*

$$\lambda_{\min}(\mathbf{D}_n \mathbf{D}_n^T) \xrightarrow{\text{P}} \infty \text{ as } n \to \infty. \tag{A.11}$$

Assume further that, for some $\boldsymbol{\lambda}_0 \in \mathbb{R}^s$,

$$\mathbf{D}_n^{-1} \mathbf{S}_n^{\lambda_0}(\theta_0) = O_P(1) \text{ as } n \to \infty, \tag{A.12}$$

and that it is possible to choose a_n *in* (A.8) *and* b_n *in* (A.9) *so that*

$$\lim_{c \to 0} \lim_{A \to \infty} \limsup_{n \to \infty} P\left(\inf_{\theta \in N_n(A)} \lambda_{\min}\left(\mathbf{D}_n^{-1} \mathbf{F}_n^{\lambda_0 *}(\theta) \mathbf{D}_n^{-T} \right) \le c \right) = 0. \tag{A.13}$$

Then there is an estimator $\widehat{\theta}_n \in \Theta^h$ *which, with probability approaching 1 as* $n \to \infty$ *then* $A \to \infty$, *satisfies* (A.6), *and maximises* $\mathcal{L}_n(\theta)$ *uniquely on* $N_n^h(A)$. *The estimator* $\widehat{\theta}_n$ *is consistent for* θ_0 *and is such that*

$$\lim_{A \to \infty} \liminf_{n \to \infty} P\{\widehat{\theta}_n \in N_n^h(A)\} = 1. \tag{A.14}$$

Moreover, $\widehat{\theta}_n$ *does not depend on* λ_0 *or on the choice of* a_n *in* (A.8), b_n *in* (A.9), *or* $\mathbf{D}_n$ *in* (A.11)–(A.13).

Once $\widehat{\boldsymbol{\theta}}_n$ is found in Theorem A.1, an estimator $\widehat{\boldsymbol{\lambda}}_n$ can be found by setting the RHS of (A.7) to 0. Note that no upper bound on the growth of $\mathbf{F}_n^{\lambda *}(\theta)$ is required for the existence and consistency of $\widehat{\boldsymbol{\theta}}_n$ in Theorem A.1. Some extra restrictions are needed however to ensure consistency of $\widehat{\boldsymbol{\lambda}}_n$. In (A.7), $\boldsymbol{\lambda}$ is allowed to depend on n. At this stage we specify the relation to be of the form

$$\boldsymbol{\lambda}_n = \mathbf{C}_n \boldsymbol{\lambda}, \tag{A.15}$$

where the $\mathbf{C}_n$ are non-singular $s \times s$ matrices and $\boldsymbol{\lambda} \in \mathbb{R}^s$ is a Lagrange multiplier.

Theorem A.2. *Assume (A1). Suppose $\boldsymbol{\lambda}_n$ in (A.7) has the form (A.15) and there is a consistent estimator $\widehat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_0$ satisfying (A.6). Assume the system of equations*

$$\mathbf{S}_n^{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\theta}}_n) = \mathbf{S}_n(\widehat{\boldsymbol{\theta}}_n) + \mathbf{H}(\widehat{\theta}_n)\mathbf{C}_n\boldsymbol{\lambda} = \mathbf{0} \tag{A.16}$$

has a unique solution $\widehat{\boldsymbol{\lambda}}_n \in \mathbb{R}^s$. Assume there is a sequence $a_n > 0$ such that

$$\frac{1}{a_n}\mathbf{C}_n \xrightarrow{\text{P}} \mathbf{C}, \text{ as } n \to \infty. \tag{A.17}$$

where $\mathbf{C}$ is a finite nonsingular $s \times s$ matrix. Assume further that

$$a_n^{-1}\mathbf{S}_n(\theta_0) \xrightarrow{\text{P}} L_0 \tag{A.18}$$

for a finite (possibly random) vector L_0, and the system of equations

$$L_0 + \mathbf{H}(\theta_0)\lambda = 0 \tag{A.19}$$

has an a.s. unique solution $\boldsymbol{\lambda}_0 := [\lambda_{10} \cdots \lambda_{s0}]^T$, which we designate as the "true" value of $\boldsymbol{\lambda}$, and that

$$\frac{1}{a_n}\Big(\mathbf{S}_n(\widehat{\theta}_n) - \mathbf{S}_n(\theta_0)\Big) \xrightarrow{\text{P}} 0, \text{ as } n \to \infty. \tag{A.20}$$

Then the estimator $\widehat{\lambda}_n$ obtained from (A.16) is consistent for λ_0: we have $\widehat{\boldsymbol{\lambda}}_n \xrightarrow{\text{P}} \boldsymbol{\lambda}_0$ as $n \to \infty$.

A.2 Asymptotic Distribution of $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$, Interior Case

The positioning of the components of $\widehat{\boldsymbol{\theta}}_n$ with respect to the boundaries of the I_j may be complicated in the sample, though tending in probability to the conformation of the components of $\boldsymbol{\theta}_0$ as $n \to \infty$. So writing down the asymptotic distribution of $\widehat{\boldsymbol{\theta}}_n$ itself, rather than that of the deviance, as in Theorem A.4, is complicated in general. We can do this when $\boldsymbol{\theta}_0$ is in the interior of Θ^h, because then the derivative $\mathbf{S}_n^{\lambda}(\boldsymbol{\theta})$ of $\mathcal{L}_n^{\lambda}(\boldsymbol{\theta})$ takes value $\mathbf{0}$ at $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$. In that case $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$ is a solution to the simultaneous equations obtained from (A.6) and (A.7):

$$\mathbf{S}_n^{\widehat{\lambda}_n}(\widehat{\boldsymbol{\theta}}_n) = \mathbf{S}_n(\widehat{\boldsymbol{\theta}}_n) + \mathbf{H}(\widehat{\boldsymbol{\theta}}_n)\mathbf{C}_n\widehat{\boldsymbol{\lambda}}_n = \mathbf{0}, \quad \text{and} \quad h_\ell(\widehat{\boldsymbol{\theta}}_n) = 0,\ \ell = 1, \ldots, s, \tag{A.21}$$

and we can derive the asymptotic distribution of $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$. To formulate this properly we need some more notation. For $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\lambda} \in \mathbb{R}$, and with $\mathbf{C}_n$ as in (A.15), define

$$\mathbf{U}_n^{\lambda}(\theta) := \begin{pmatrix} \mathbf{F}_n^{\lambda}(\boldsymbol{\theta}) & -\mathbf{H}(\boldsymbol{\theta})\mathbf{C}_n \\ -\mathbf{C}_n^T\mathbf{H}^T(\boldsymbol{\theta}) & \mathbf{0}_{s\times s} \end{pmatrix}_{(d+s)\times(d+s)} \tag{A.22}$$

where $\mathbf{0}_{s\times s}$ is the $s \times s$ zero matrix.

Theorem A.3. *Assume that there exist $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\lambda}}_n)$ consistent for $(\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)$ and satisfying (A.14) and (A.21) for some choice of nonsingular $d \times d$ matrices $\mathbf{D}_n$ in (A.10). Suppose also that*

$$\mathbf{D}_n^{-1}\mathbf{S}_n^{\lambda_0}(\boldsymbol{\theta}_0) \xrightarrow{\text{D}} \mathbf{Z}, \ \text{as } n \to \infty, \tag{A.23}$$

for an a.s. finite random vector $\mathbf{Z}$, and there exist nonsingular $s \times s$ matrices $\mathbf{E}_n$ and a nonsingular $(d+s)\times(d+s)$ matrix $\mathbf{U}_0$ such that, with

$$\mathbf{J}_n := \begin{bmatrix} \mathbf{D}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_n \end{bmatrix}, \tag{A.24}$$

we have

$$\mathbf{J}_n^{-1}\mathbf{U}_n^{\lambda_n}(\boldsymbol{\theta}_n)\mathbf{J}_n^T \xrightarrow{\text{P}} \mathbf{U}_0, \tag{A.25}$$

for any random sequence $(\boldsymbol{\theta}_n, \boldsymbol{\lambda}_n) \xrightarrow{\text{P}} (\boldsymbol{\theta}_0, \boldsymbol{\lambda}_0)$, as $n \to \infty$. Then as $n \to \infty$

$$\mathbf{J}_n \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{\text{D}} -\mathbf{U}_0^{-1} \begin{pmatrix} \mathbf{Z} \\ \mathbf{0} \end{pmatrix}. \tag{A.26}$$

A.3 Hypothesis Tests

We wish to test the hypothesis $H:\ \theta \in \Omega$ versus $\theta \in \tau$, where Ω and τ are two fixed, disjoint, subsets of Θ. Given consistent estimators $\widehat{\theta}_n^{\Omega}$ and $\widehat{\theta}_n^{\tau}$ for θ_0 which maximise $\mathcal{L}_n(\theta)$ over Ω and τ respectively, we use the deviance statistic (an analogue of the minus twice log-likelihood ratio statistic)

$$d_n := 2\left(\mathcal{L}_n(\widehat{\theta}_n^{\tau}) - \mathcal{L}_n(\widehat{\theta}_n^{\Omega})\right) \tag{A.27}$$

to compare Ω with τ.

As in Chernoff (1954) and Vu-Zhou we assume Ω and τ can be approximated locally by cones. So we make some further assumptions.

(A2) A subset $\Omega \subseteq \Theta$ is said to satisfy (A2) if there is a closed cone C_Ω with vertex at θ_0 such that

$$C_\Omega \cap \mathcal{N} = \Omega \cap \mathcal{N},$$

where $\mathcal{N}$ is a closed neighbourhood in $\mathbb{R}^d$ of θ_0.

Suppose Ω and τ satisfy (A2) with corresponding disjoint cones C_Ω and C_τ. For $\mathbf{T}_n$ nonstochastic orthogonal $d \times d$ matrices and $\mathbf{D}_n$ possibly random a.s. nonsingular $d \times d$ matrices define

$$\widetilde{C}_{\Omega_n} := \left\{ \widetilde{\theta} : \widetilde{\theta} = \mathbf{T}_n \mathbf{D}_n (\theta - \theta_0), \theta \in C_\Omega \right\},$$

and similarly for $\widetilde{C}_{\tau_n}$.

(A3): A subset $\Omega \subseteq \Theta$ is said to satisfy (A3) if there is a closed cone $\widetilde{C}_\Omega$ with vertex at 0, not depending on n, such that the sets $\widetilde{C}_{\Omega_n}$ asymptotically coincide with $\widetilde{C}_\Omega$ in the sense that

$$\lim_{n\to\infty} \sup_{|\beta|=1} \left| \inf_{\theta \in \widetilde{C}_{\Omega_n}} |\beta - \theta|^2 - \inf_{\theta \in \widetilde{C}_\Omega} |\beta - \theta|^2 \right| = 0. \tag{A.28}$$

Recall the definition of the neighbourhoods $N_n^h(A)$ in (A.10), depending on $\mathbf{D}_n$ and $A > 0$. Assume n is so large for fixed A that $N_n^h(A)$ is contained in the neighbourhood $\mathcal{N}$ in (A2).

Theorem A.4. *Let Ω and τ be two disjoint subsets of Θ^h satisfying conditions (A1)–(A3) for some orthogonal $d \times d$ matrices $\mathbf{T}_n$ and nonsingular $d \times d$ matrices $\mathbf{D}_n$. Suppose there is a $\lambda_0 \in \mathbb{R}^s$ such that*

$$\mathbf{D}_n^{-1} \mathbf{S}_n^{\lambda_0}(\theta_0) \xrightarrow{\text{D}} \mathbf{Z} \tag{A.29}$$

for an a.s. finite random vector $\mathbf{Z} \in \mathbb{R}^d$, and suppose there is a finite non-singular symmetric matrix $\mathbf{U}_0$ such that, with $\mathbf{U}_n^\lambda(\boldsymbol{\theta})$ as defined in (A.22) and $\mathbf{J}_n$ as in (A.24),

$$\mathbf{J}_n^{-1} \mathbf{U}_n^{\lambda_0}(\boldsymbol{\theta}_0) \mathbf{J}_n^{-T} \xrightarrow{\text{P}} \mathbf{U}_0, \quad n \to \infty. \tag{A.30}$$

Write

$$\mathbf{U}_0^{-1} = \begin{pmatrix} \mathbf{P}_{d\times d} & \mathbf{Q}_{d\times s} \\ \mathbf{Q}_{s\times d}^T & \mathbf{R}_{s\times s} \end{pmatrix}. \tag{A.31}$$

Assume (A.15) *and also that*

$$\sup_{\boldsymbol{\theta}\in N_n(A)} ||\mathbf{J}_n^{-1}(\mathbf{U}_n^{\lambda_0}(\boldsymbol{\theta}) - \mathbf{U}_n^{\lambda_0}(\boldsymbol{\theta}_0))\mathbf{J}_n^{-T}||_1 \xrightarrow{\mathrm{P}} 0, \tag{A.32}$$

as $n \to \infty$, *for each* A *sufficiently large.*

Let $\widehat{\theta}_n^{\Omega}$ *and* $\widehat{\theta}_n^{\tau}$ *be MEs of* $\mathcal{L}_n(\boldsymbol{\theta})$ *over* Ω *and* τ *respectively, each consistent for* $\boldsymbol{\theta}_0$ *and each satisfying the constraints* (A.6), *as well as* (A.14). *Then the deviance statistic* d_n *in* (A.27) *converges to a limiting random variable as follows:*

$$d_n \xrightarrow{\mathrm{D}} \inf_{\boldsymbol{\theta}\in\dot{C}_\Omega}\left(\mathbf{P}^{T/2}\mathbf{Z} - \boldsymbol{\theta}\right) - \inf_{\boldsymbol{\theta}\in\dot{C}_\tau}\left(\mathbf{P}^{T/2}\mathbf{Z} - \boldsymbol{\theta}\right), \text{ as } n \to \infty. \tag{A.33}$$

where $\dot{C}_\Omega = \{\mathbf{P}^{T/2}\mathbf{F}_0\boldsymbol{\theta} : \boldsymbol{\theta} \in \widetilde{C}_\Omega\}$, $\dot{C}_\tau = \{\mathbf{P}^{T/2}\mathbf{F}_0\boldsymbol{\theta} : \boldsymbol{\theta} \in \widetilde{C}_\tau\}$, *and* $\mathbf{F}_0$ *is the limit of* $\mathbf{D}_n^{-1}\mathbf{F}_n^{\lambda_0}(\boldsymbol{\theta})\mathbf{D}_n^{-T}$ *in probability, which exists under* (A.30).

Notes: (i) The limit random variable $\mathbf{Z}$ in (A.29) need not be full, that is, its distribution may be concentrated on a lower dimensional subspace.

(ii) Theorem A.1 gives sufficient conditions for the application of Theorem A.4 namely, (A.11)–(A.13). When these hold, MEs $\widehat{\theta}_n^{\Omega}$ and $\widehat{\theta}_n^{\tau}$ exist and are unique in $N_n^h(A)$, in the sense of (A.14). The estimators $\widehat{\theta}_n^{\Omega}$ and $\widehat{\theta}_n^{\tau}$ are assumed to satisfy the constraints (A.6) in both Theorems A.1 and A.4.

(iii) When (A.30) holds we can write

$$\mathbf{U}_0 = \begin{pmatrix} \mathbf{F}_0 & -\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{C} \\ -\mathbf{C}^T\mathbf{H}^T(\boldsymbol{\theta}_0) & \mathbf{0}_{s\times s} \end{pmatrix}_{(d+s)\times(d+s)}, \tag{A.34}$$

where $-\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{C}$ is the limit of $-\mathbf{D}_n^{-1}\mathbf{H}(\boldsymbol{\theta}_0)\mathbf{C}_n\mathbf{D}_n^{-T}$ in probability, which exists under (A.30).

Appendix B

Spherical Data Calculated from Monthly Portfolios

B.1 The Data in 3 dimensions

After calculating the monthly portfolios from 3 indices in Table 4.1, we transform them by taking a square root to be able to locate them on the sphere. The raw are listed in Tables B.1 to B.9.

The mean vector of data is

$$\bar{\mathbf{x}} = (0.461926, 0.359668, 0.429174)^T,$$

and the covariance matrix is

$$\mathbf{S} = \begin{pmatrix} 0.149039 & -0.0642449 & -0.0579053 \\ -0.0642449 & 0.158363 & -0.0848397 \\ -0.0579053 & -0.0848397 & 0.166997 \end{pmatrix}.$$

If we model the data a von Mises Fisher distribution, the MLEs of the parameters are

$$\widehat{\kappa} = 3.62962$$

and

$$\widehat{\boldsymbol{\mu}} = (0.636351, 0.495481, 0.591233)^T.$$

B.2 The Data in 10 dimensions

Tables B.10 to B.16 show the spherical data in 10 dimensions after transformation from asset allocation data compiled from the 10 indices in Table 4.2. The estimated $\widehat{\kappa}$ and $\widehat{\boldsymbol{\mu}}$ are given in (4.22). We print out the whole data here. There are totally 144 data with the mean vector

$$\overline{\mathbf{x}} = (0.130603, 0.0806646, 0.156781, 0.434826, 0.115253, 0.0574893, 0.342915, 0.166109, 0.141441, 0.0598201)^T,$$

and the covariance matrix

$$\mathbf{S} = \{\{0.0559012, -0.00721278, -0.00737622, -0.0197796, 0.00692629, -0.00318147, -0.0163106, -0.00283348, 0.00750688, -0.00636056\}, \{-0.00721278, 0.0321734, -0.00102886, 0.0106829, -0.0036412, 0.00126389, -0.0102383, -0.0088255, 0.00370223, -0.00432247\}, \{-0.00737622, -0.00102886, 0.0640193, -0.0114939, -0.011325, -0.000993111, -0.00599782, -0.00415643, -0.0114629, -0.00132637\}, \{-0.0197796, 0.0106829, -0.0114939, 0.113205, -0.01917, -0.00178454, -0.0201325, -0.0263122, -0.021801, -0.000450928\}, \{0.00692629, -0.0036412, -0.011325, -0.01917, 0.0478827, -0.00121497, 0.00103254, -0.00443293, -0.00390525, -0.000734449\}, \{-0.00318147, 0.00126389, -0.0009931, -0.00178454, -0.00121497, 0.0278576, -0.00472032, -0.000519653, -0.00423302, -0.00128753\}, \{-0.0163106, -0.0102383, -0.00599782, -0.0201325, 0.00103254, -0.00472032, 0.103298, 0.0000180312, -0.0126301, -0.00202383\}, \{-0.00283348, -0.0088255, -0.00415643, -0.0263122, -0.00443293, -0.000519653, 0.0000180312, 0.0423305, -0.00874948, -0.00454568\}, \{0.00750688, 0.00370223, -0.0114629, -0.021801, -0.00390525, -0.00423302, -0.0126301, -0.00874948, 0.0711155, 0.00161919\}, \{-0.006360, -0.00432247, -0.00132637, -0.00045093, -0.00073445, -0.00128753, -0.00202383, -0.00454568, 0.00161919, 0.0236817\}\},$$

which is a 10×10 matrix.

Table B.1: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19)

Region	Percentage	Data
middle	$\frac{185}{358}$	{0.352014 , 0.935993 , 0.00172995}, { 0.719622 , 0.694363 , 0.00187 }, { 0.453054 , 0.550747 , 0.701014 },{ 0.505663 , 0.788367 , 0.350402 }, { 0.156803 , 0.577273 , 0.801354 },{ 0.816509 , 0.577328 , 0.00225134 }, { 0.00116328 , 0.999982 , 0.00584933 },{ 0.655729 , 0.590812 , 0.470064 }, { 0.40531 , 0.707142 , 0.579374 },{ 0.319246 , 0.00104044 , 0.947671 }, { 0.476678 , 0.622477 , 0.620726 },{ 0.00277283 , 0.0040553 , 0.999988 }, { 0.00122355 , 0.999998 , 0.00128296 },{ 0.630502 , 0.00101314 , 0.776187 }, { 0.429435 , 0.28218 , 0.857881 },{ 0.56406 , 0.706627 , 0.427217 }, { 0.450746 , 0.477628 , 0.754122 },{ 0.452173 , 0.510234 , 0.731574 }, { 0.00133412 , 0.733519 , 0.679668 },{ 0.395236 , 0.597519 , 0.697682 }, { 0.0944287 , 0.63109 , 0.769941 },{ 0.99995 , 0.00940451 , 0.00347539 }, { 0.00145393 , 0.729909 , 0.683542 },{ 0.291491 , 0.872815 , 0.391442 }, { 0.00419149 , 0.00101627 , 0.999991 },{ 0.837169 , 0.546937 , 0.00287808 }, { 0.00528753 , 0.245804 , 0.969305 },{ 0.528812 , 0.76922 , 0.358691 }, { 0.792323 , 0.363692 , 0.48985 },{ 0.650989 , 0.304896 , 0.695163 }, { 0.808757 , 0.262453 , 0.526337 },{ 0.571447 , 0.765914 , 0.294659 }, { 0.5393 , 0.597425 , 0.593497 },{ 0.00153638 , 0.999998 , 0.00105736 }, { 0.543899 , 0.769562 , 0.334588 },{ 0.00144256 , 0.00159231 , 0.999998 }, { 0.00249333 , 0.999995 , 0.00201288 },{ 0.740521 , 0.672032 , 0.0016219 }, { 0.99999 , 0.00276406 , 0.00354366 },{ 0.00130902 , 0.999998 , 0.00137527 }, { 0.854824 , 0.518917 , 0.00101987 },{ 0.621896 , 0.00196542 , 0.783098 }, { 0.0014591 , 0.743203 , 0.669064 },{ 0.999992 , 0.00155435 , 0.00369187 }, { 0.00110687 , 0.999999 , 0.00113883 },{ 0.00187082 , 0.31901 , 0.94775 }, { 0.999994 , 0.0012376 , 0.00336893 },{ 0.00198499 , 0.999998 , 0.00102413 }, { 0.546521 , 0.071753 , 0.834366 },{ 0.138399 , 0.691625 , 0.708872 }, { 0.489315 , 0.00487074 , 0.872094 },{ 0.756828 , 0.348398 , 0.55302 }, { 0.0032686 , 0.00197617 , 0.999993 },{ 0.584688 , 0.365474 , 0.724271 }, { 0.807016 , 0.103627 , 0.581367 },{ 0.430948 , 0.788546 , 0.438724 }, { 0.614291 , 0.666601 , 0.422244 },{ 0.794941 , 0.143177 , 0.58955 }, { 0.462945 , 0.760027 , 0.456115 },{ 0.890808 , 0.116432 , 0.439209 }, { 0.680151 , 0.422776 , 0.598878 },{ 0.00326561 , 0.77411 , 0.633043 }, { 0.00966262 , 0.999947 , 0.00341721 },{ 0.70968 , 0.289638 , 0.642234 }, { 0.634314 , 0.77144 , 0.0502561 },{ 0.448544 , 0.600277 , 0.662175 }, { 0.364704 , 0.481503 , 0.796961 },{ 0.396215 , 0.918156 , 0.00176941 }, { 0.810265 , 0.586059 , 0.00233471 },{ 0.00352423 , 0.00101407 , 0.999993 }, { 0.925717 , 0.00129069 , 0.378215 },{ 0.313122 , 0.540661 , 0.780795 },

Table B.2: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
middle		{0.0020832 , 0.999995 , 0.00223426 },{ 0.650597 , 0.715236 , 0.255266 }, { 0.425096 , 0.26748 , 0.864724 },{ 0.45468 , 0.274157 , 0.84741 }, { 0.781736 , 0.00117752 , 0.623608 },{ 0.00274746 , 0.999984 , 0.00485581 }, { 0.57901 , 0.00141995 , 0.815319 },{ 0.553503 , 0.829866 , 0.0704103 }, { 0.840593 , 0.00115424 , 0.541666 },{ 0.0012668 , 0.999999 , 0.00107994 }, { 0.982865 , 0.184306 , 0.00288599 },{ 0.00534752 , 0.570057 , 0.821588 }, { 0.999993 , 0.00205115 , 0.00307036 },{ 0.64037 , 0.407008 , 0.651361 }, { 0.00666318 , 0.614025 , 0.789259 },{ 0.00146178 , 0.00299114 , 0.999994 }, { 0.456126 , 0.360506 , 0.813625 },{ 0.460438 , 0.240944 , 0.854367 }, { 0.615892 , 0.693357 , 0.374077 },{ 0.00264466 , 0.999995 , 0.00198088 }, { 0.939059 , 0.343753 , 0.00155842 },{ 0.464958 , 0.461429 , 0.755577 }, { 0.701062 , 0.125289 , 0.702007 },{ 0.00205227 , 0.999996 , 0.00218503 }, { 0.291863 , 0.767281 , 0.571048 },{ 0.856728 , 0.515764 , 0.00225367 }, { 0.340253 , 0.343103 , 0.875505 },{ 0.999991 , 0.00162885 , 0.00397511 }, { 0.769166 , 0.387951 , 0.507817 },{ 0.00236017 , 0.00179758 , 0.999996 }, { 0.999994 , 0.00138217 , 0.00328602 },{ 0.514616 , 0.00110037 , 0.85742 }, { 0.00195254 , 0.00116118 , 0.999997 },{ 0.999998 , 0.00109702 , 0.00154891 }, { 0.215416 , 0.00140023 , 0.976521 },{ 0.675369 , 0.353087 , 0.647462 }, { 0.440629 , 0.0141629 , 0.897577 },{ 0.999993 , 0.00254548 , 0.0026769 }, { 0.00258383 , 0.999987 , 0.00437612 },{ 0.00161248 , 0.999997 , 0.00180137 }, { 0.00467245 , 0.999988 , 0.00175159 },{ 0.641586 , 0.767047 , 0.00245486 }, { 0.415965 , 0.0386967 , 0.908557 },{ 0.36811 , 0.288618 , 0.883852 }, { 0.00191664 , 0.999996 , 0.0022545 },{ 0.672558 , 0.00238063 , 0.74004 }, { 0.00209776 , 0.00806007 , 0.999965 },{ 0.536242 , 0.419139 , 0.732644 }, { 0.594695 , 0.00203825 , 0.803949 },{ 0.659548 , 0.308268 , 0.685542 }, { 0.332122 , 0.943193 , 0.00901089 },{ 0.00411263 , 0.99998 , 0.00477705 }, { 0.497295 , 0.00125381 , 0.867581 },{ 0.642056 , 0.562828 , 0.520566 }, { 0.257266 , 0.700344 , 0.665832 },{ 0.57758 , 0.718779 , 0.386985 }, { 0.611144 , 0.43646 , 0.660307 },{ 0.00313655 , 0.0017256 , 0.999994 }, { 0.668025 , 0.263346 , 0.695983 },{ 0.802325 , 0.00134369 , 0.596886 }, { 0.999996 , 0.0022126 , 0.00169603 },{ 0.798085 , 0.411303 , 0.440329 }, { 0.517011 , 0.855971 , 0.00358713 },{ 0.679391 , 0.00302605 , 0.733771 }, { 0.00148002 , 0.999968 , 0.00787203 },{ 0.814276 , 0.562126 , 0.144808 }, { 0.811235 , 0.00185845 , 0.584717 },{ 0.683472 , 0.00111453 , 0.729976 }, { 0.00100199 , 0.00109114 , 0.999999 },{ 0.999997 , 0.00147852 , 0.00170078 }, { 0.00173969 , 0.999996 , 0.00201011 },{ 0.00316617 , 0.00451845 , 0.999985 },

TABLE B.3: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
middle		{0.00369626 , 0.00255144 , 0.99999 },{ 0.700728 , 0.713427 , 0.00157152 }, { 0.00104474 , 0.999999 , 0.00135609 },{ 0.524648 , 0.0012504 , 0.851319 }, { 0.00118336 , 0.999999 , 0.00120177 },{ 0.323697 , 0.946159 , 0.00185815 }, { 0.405551 , 0.00237317 , 0.91407 },{ 0.592651 , 0.278844 , 0.755652 }, {0.00590924 , 0.676373 , 0.736536 },{ 0.604346 , 0.00266217 , 0.796717 }, { 0.655814 , 0.00171312 , 0.754921 },{ 0.686843 , 0.721924 , 0.0840957 }, { 0.87692 , 0.00283066 , 0.480628 },{ 0.607859 , 0.461429 , 0.646212 }, { 0.00307043 , 0.00227016 , 0.999993 },{ 0.898236 , 0.350289 , 0.265463 }, { 0.800788 , 0.00126549 , 0.598946 },{ 0.737032 , 0.253456 , 0.626533 }, { 0.358402 , 0.0029963 , 0.933562 },{ 0.916671 , 0.399632 , 0.00303627 }, { 0.710646 , 0.0121922 , 0.703445 },{ 0.999993 , 0.00168474 , 0.00329647 }, { 0.78165 , 0.473189 , 0.406344 },{ 0.800826 , 0.598892 , 0.0025671 }, { 0.0017053 , 0.00112822 , 0.999998 },{ 0.785623 , 0.618703 , 0.00190052 }, { 0.780627 , 0.145255 , 0.607884 },{ 0.883262 , 0.001569 , 0.468877 }, { 0.0010447 , 0.0015468 , 0.999998 },{ 0.00117585 , 0.999999 , 0.001219 }, { 0.00282398 , 0.492396 , 0.870367 },{ 0.756939 , 0.320292 , 0.56961 }, { 0.0063221 , 0.862596 , 0.505854 },{ 0.00280627 , 0.999992 , 0.0027815 }, { 0.940758 , 0.00109455 , 0.339078 },{ 0.719586 , 0.694402 , 0.00130426 }, { 0.848328 , 0.441867 , 0.291709 },{ 0.00157944 , 0.999995 , 0.00272289 }, { 0.90712 , 0.420798 , 0.00795884 },{ 0.867588 , 0.49728 , 0.00170719 }, { 0.516103 , 0.00382542 , 0.856518 },

Table B.4: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
S_1	$\frac{39}{358}$	{0.692551,0.721368,0.000662363},{0.924267,0.381746,0.000131011}, {0.578301,0.815824,0.0000634708},{0.601562,0.798826,0.000622569}, {0.52333,0.85213,0.000162075},{0.0013159,0.999999,0.000329351}, {0.00147416,0.999998,0.000959268},{0.274977,0.961451,0.000363493}, {0.294123,0.955767,0.000509399},{0.999784,0.0207562,0.000415547}, {0.00341641,0.999994,0.000828222},{0.00285444,0.999996,0.000847743}, {0.370688,0.928757,0.000711938},{0.391372,0.920232,0.000528523}, {0.568233,0.822868,0.000461149},{0.00121168,0.999999,0.00091224}, {0.909791,0.415067,0.000477528},{0.00112507,0.999999,0.000868788}, {0.488726,0.872437,0.000650375},{0.574966,0.818177,0.00016392}, {0.971086,0.23873,0.000258139},{0.987442,0.157983,0.000140499}, {0.951011,0.309156,0.000726525},{0.00111912,0.999999,0.00089369}, {0.929526,0.368756,0.000857375},{0.636562,0.771225,0.000670257}, {0.903072,0.42949,0.000176267},{0.0026456,0.999996,0.00092242}, {0.999996,0.00265885,0.000396292},{0.00174086,0.999998,0.000273311}, {0.564003,0.825772,0.00071705},{0.00107175,0.999999,0.000945818}, {0.00135338,0.999999,0.000703461},{0.999998,0.00184758,0.000578143}, {0.72147,0.692445,0.000669136},{0.891276,0.45346,0.00026598}, {0.901589,0.432593,0.000250701},{0.74232,0.670045,0.000614302}, {0.999998,0.00201108,0.000432169}

Table B.5: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
S_2	$\frac{71}{358}$	{0.00983382 , 0.000930043 , 0.999951 } , { 0.315145 , 0.000372726 , 0.949043 } , { 0.705468 , 0.0000533349 , 0.708742 } , { 0.999999 , 0.000553946 , 0.00113985 } , { 0.34087 , 0.000427287 , 0.94011 } , { 0.999996 , 0.000980637 , 0.00253416 } , { 0.419335 , 0.000791702 , 0.907831 } , { 0.00185117 , 0.00058955 , 0.999998 } , { 0.816053 , 0.000176969 , 0.577977 } , { 0.889821 , 0.000305323 , 0.45631 } , { 0.707562 , 0.000864791 , 0.70665 } ,{ 0.00149118 , 0.000915772 , 0.999998 } , { 0.667661 , 0.000782409 , 0.744465 } , { 0.00147649 , 0.000695025 , 0.999999 } , { 0.999999 , 0.00070683 , 0.00104652 } , { 0.99999 , 0.00060179 , 0.00436907 } , { 0.787894 , 0.000190273 , 0.615811 } , { 0.999997 , 0.000986421 , 0.00215789 } , { 0.558642 , 0.00027366 , 0.829409 } ,{ 0.00211912 , 0.000507069 , 0.999998 } , { 0.525418 , 0.000337187 , 0.850844 } , { 0.964968 , 0.000943768 , 0.262367 } , { 0.861322 , 0.000149072 , 0.50806 } , { 0.999999 , 0.000554499 , 0.00103681 } , { 0.999999 , 0.000965597 , 0.00113552 } , { 0.999983 , 0.0008334 , 0.00583628 } , { 0.999998 , 0.000486069 , 0.00200949 } , { 0.0937899 , 0.0000607236 , 0.995592 } , { 0.475748 , 0.000649723 , 0.879581 } , { 0.914229 , 0.000217842 , 0.405197 } , { 0.999998 , 0.000841697 , 0.00195421 } , { 0.627513 , 0.000206781 , 0.778606 } , { 0.999999 , 0.000149862 , 0.00104882 } , { 0.00548942 , 0.000683536 , 0.999985 } , { 0.999995 , 0.000836362 , 0.0030401 } , { 0.773556 , 0.000431861 , 0.633728 } , { 0.999993 , 0.000500732 , 0.00378522 } , { 0.999998 , 0.000562973 , 0.00181362 } , { 0.999992 , 0.000849438 , 0.00388277 } , { 0.277071 , 0.000947365 , 0.960849 } , { 0.247888 , 0.0000510817 , 0.968789 } , { 0.00812399 , 0.000581734 , 0.999967 } , { 0.699413 , 0.000614444 , 0.714718 } , { 0.00245308 , 0.000976751 , 0.999997 } , { 0.834484 , 0.000619478 , 0.551032 } , { 0.00269469 , 0.000635312 , 0.999996 } , { 0.702514 , 0.000960816 , 0.711669 } , { 0.999999 , 0.000368756 , 0.00104145 } , { 0.6018 , 0.000439399 , 0.798647 } ,{ 0.180177 , 0.0000675976 , 0.983634 } , { 0.277594 , 0.000539052 , 0.960698 } , { 0.0599314 , 0.0000479495 , 0.998202 } , { 0.758261 , 0.000884991 , 0.651951 } , { 0.999901 , 0.000205284 , 0.0140472 } , { 0.72795 , 0.000818611 , 0.68563 } , { 0.894966 , 0.000825934 , 0.446134 } , { 0.626585 , 0.00040761 , 0.779353 } , { 0.999999 , 0.000768054 , 0.00104189 } , { 0.729394 , 0.00090221 , 0.684094 } , { 0.999996 , 0.000473962 , 0.00270149 } , { 0.999993 , 0.000957018 , 0.00358158 } , { 0.602469 , 0.000266577 , 0.798142 } , { 0.769519 , 0.000407282 , 0.638624 } , { 0.00203299 , 0.000611137 , 0.999998 } , { 0.999999 , 0.000687586 , 0.0012133 } , { 0.00356908 , 0.000984778 , 0.999993 } , { 0.468379 , 0.000454021 , 0.883528 } , { 0.374546 , 0.000664263 , 0.927208 } , { 0.605579 , 0.000296383 , 0.795785 } , { 0.672291 , 0.000910217 , 0.740287 } , { 0.518679 , 0.000591915 , 0.854969 }

Table B.6: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
S_3	$\frac{17}{358}$	{ 0.000265227 , 0.90102 , 0.433778 } , { 0.000302373 , 0.0124206 , 0.999923 } , { 0.000704761 , 0.934563 , 0.355797 } , { 0.000980538 , 0.00267897 , 0.999996 } , { 0.000714549 , 0.690603 , 0.723234 } , { 0.0005711 , 0.35875 , 0.933434 } , { 0.00050039 , 0.999995 , 0.00318889 } , { 0.000813993 , 0.00105878 , 0.999999 } , { 0.000663268 , 0.999999 , 0.0011022 } , { 0.000845078 , 0.999999 , 0.00148456 } , { 0.000745387 , 0.999999 , 0.00131026 } , { 0.000969013 , 0.999993 , 0.00371893 } , { 0.000456491 , 0.261907 , 0.965093 } , { 0.000617822 , 0.999997 , 0.00240552 } , { 0.000649666 , 0.755317 , 0.65536 } , { 0.000766754 , 0.999999 , 0.00120093 } , { 0.000557947 , 0.999998 , 0.00202205 } ,

Table B.7: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
Corner3	$\frac{20}{358}$	{0.000212722 , 0.000281871 , 1. }, { 0.000022675 , 0.0000110575 , 1.}, { 0.000435162 , 0.000357203 , 1. }, 0.000637495 , 0.000464282 , 1.}, { 0.000583694 , 0.00053032 , 1. }, 0.0000385556 , 0.0000267907 , 1.}, { 0.000235976 , 0.000705921 , 1. }, 0.000749169 , 0.00020275 , 1.} , { 0.000681898 , 0.0000932437 , 1. }, 0.000235522 , 0.000161494 , 1.}, { 0.000712615 , 0.0000553099 , 1. }, 0.000522713 , 0.000568132 , 1.}, { 0.000466775 , 0.000270535 , 1. }, 0.000556154 , 0.000646453 , 1.}, { 0.000600956 , 0.000298167 , 1. }, 0.000852049 , 0.000436768 , 1.}, { 0.000253668 , 0.000388621 , 1. }, 0.0000665706 , 0.000166509 , 1.}, { 0.00072244 , 0.000691143 , 1. }, 0.000565177 , 0.000343496 , 1. }

Table B.8: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
Corner2	$\frac{11}{358}$	{0.000724882, 1., 0.00067779}, {0.00038898, 1., 0.000327751}, {0.000835572, 0.999999, 0.000772552}, {0.000444709, 1., 0.000402904}, {0.000476156, 1., 0.000487885}, {0.000291821, 1., 0.000162756}, {0.000454314, 1., 0.000521894}, {0.000714907, 1., 0.000638554}, {0.000395162, 1., 0.000724607}, {0.000864367, 0.999999, 0.000604575}, {0.000715478, 1., 0.000433823}

TABLE B.9: Spherical data calculated from monthly portfolios based on 3 indices in Table 4.1, categorised in 7 regions of (4.19) (continued)

Region	Percentage	Data
Corner1	$\frac{15}{358}$	{1., 0.000292693, 0.000302654}, {1., 0.000203084, 0.000868925}, {0.999999, 0.000710909, 0.00081809}, {1., 0.00031117, 0.000558795}, {0.999999, 0.000709405, 0.000940272}, {1., 0.000570585, 0.000559562}, {1., 0.000683763, 0.000278765}, {1., 0.00033426, 0.00090931}, {1., 0.000342414, 0.000309186}, {1., 0.000579557, 0.000489138}, {1., 0.000620808, 0.000661229}, {1., 0.000317232, 0.000544397}, {1., 4.39876×10^{-6}, 4.70714×10^{-6}}, {1., 0.0000323364, 0.0000473472}, {1., 6.944×10^{-6}, 0.0000168692}

TABLE B.10: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero

Region	Percentage	Data
S_1	$\frac{2}{144}$	{0.001, 0.001, 0, 0.001, 0.001, 0.002, 0.002, 1.00, 0.003, 0.001}, {0.001, 0.561, 0, 0.291, 0.001, 0.006, 0.537, 0.316, 0.461, 0.006}
S_2	$\frac{3}{144}$	{0.003, 0.001,0 , 0.002, 0.728, 0.003, 0.512, 10^{-4}, 0.003,0.456}, {0.399, 10^{-6}, 0.189, 0.253, 0.235, 10^{-6}, 0.355, 0.182, 0.630, 0.359}, {10^{-4}, 0.272, 0.179, 0.561, 0.265, 0.178, 0.671, 0.165, 10^{-4}, 0.003}
S_3	$\frac{7}{144}$	{0.001,0.337,10^{-4}, 0.412,10^{-4},0.003,0.171,10^{-4},0.819, 0.134} {0.207, 0.306, 0.009, 0.484, 10^{-4}, 0.002, 0.470, 10^{-4}, 0.639, 10^{-4}}, {0.669, 10^{-4}, 10^{-4}, 0.398, 0.610, 0.002, 0.014, 0.021, 0.148, 10^{-4}}, {10^{-4}, 0.002, 10^{-4}, 10^{-4}, 0.002, 0.007, 0.469, 0.002, 0.708, 0.527}, {0.503, 0.001,10^{-3}, 0.411, 10^{-4}, 0.002, 0.007, 0.210, 0.730, 10^{-4}}, {0.191, 0.138, 10^{-5}, 0.330, 0.399, 10^{-4}, 0.615, 0.243, 0.488, 10^{-5}}, {10^{-4}, 0.495, 0.265, 0.667, 10^{-5}, 0.280, 0.400, 0.055, 10^{-4}, 0.003}
S_4	$\frac{15}{144}$	{0.323, 10^{-4}, 10^{-4}, 0.791, 10^{-4}, 0.307, 0.419, 10^{-4}, 0.003, 0.025, {10^{-4}, 0.094, 0.304, 0.349, 0.282, 10^{-4}, 0.203, 10^{-4}, 0.810, 10^{-4}}, {0.386, 10^{-4}, 0.157, 0.215, 10^{-5}, 0.001, 0.214, 10^{-5}, 0.857, 10^{-4}}, {0.327, 10^{-4}, 10^{-4}, 0.468, 10^{-4}, 0.568, 0.577, 0.135, 0.020, 10^{-4}}, {10^{-5}, 10^{-5}, 0.452, 10^{-5}, 0.253, 10^{-5}, 0.739, 0.235, 0.022, 0.360}, {0.001, 0.002, 10^{-4}, 10^{-4}, 0.708, 0.656, 0.263, 10^{-5}, 0.001, 10^{-4}}, {10^{-4}, 0.580, 10^{-4}, 0.604, 0.380, 0.002, 10^{-4}, 0.125, 0.372, 10^{-4}}, {0.336, 10^{-5}, 0.016, 0.816, 10^{-5}, 10^{-5}, 0.420, 0.147, 10^{-5}, 0.152}, {10^{-4}, 0.292, 0.571, 0.462, 10^{-5}, 0.349, 0.504, 10^{-5}, 0.001, 10^{-4}}, {0.636, 10^{-4}, 0.001, 10^{-4}, 10^{-4}, 0.003, 0.676, 0.340, 0.149, 10^{-4}}, {10^{-7}, 10^{-6}, 0.078, 0.063, 0.395, 10^{-6}, 0.866, 0.285, 10^{-6}, 0.062}, {10^{-4}, 10^{-4}, 10^{-4}, 0.454, 10^{-4}, 0.464, 0.136, 0.190, 0.338, 0.640}, {10^{-4}, 0.003, 0.206, 0.727, 0.002, 10^{-4}, 0.655, 10^{-4}, 10^{-4}, 0.002}, {0.180, 10^{-5}, 0.259, 0.677, 10^{-5}, 0.109, 0.637, 0.159, 10^{-4}, 10^{-5}}, {10^{-5}, 0.120, 10^{-5}, 0.543, 0.158, 10^{-4}, 0.638, 0.187, 0.473, 10^{-5}}

Table B.11: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_5	$\frac{28}{144}$	$\{10^{-5}, 10^{-5}, 0.033, 0.766, 10^{-5}, 10^{-6}, 0.452, 0.070, 10^{-6}, 0.450\}$,
		$\{10^{-5}, 0.386, 10^{-5}, 0.299, 10^{-5}, 0.413, 0.041, 10^{-5}, 0.768, 10^{-5}\}$,
		$\{0.165, 0.758, 10^{-4}, 10^{-4}, 10^{-4}, 0.003, 10^{-4}, 0.191, 0.601, 10^{-4}\}$,
		$\{0.481, 10^{-4}, 10^{-4}, 0.427, 0.227, 10^{-4}, 10^{-4}, 0.171, 0.711, 10^{-4}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-5}, 0.600, 10^{-5}, 0.001, 0.440, 0.193, 0.640, 10^{-4}\}$,
		$\{10^{-4}, 10^{-4}, 0.563, 10^{-4}, 0.003, 0.526, 0.619, 0.153, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.001, 10^{-4}, 0.678, 10^{-4}, 0.001, 10^{-4}, 10^{-4}, 0.718, 0.157\}$,
		$\{10^{-4}, 10^{-4}, 0.533, 0.440, 10^{-4}, 0.002, 10^{-5}, 0.130, 0.710, 10^{-5}\}$,
		$\{10^{-4}, 0.379, 0.011, 0.770, 0.485, 10^{-4}, 10^{-4}, 0.168, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 10^{-3}, 0.931, 10^{-4}, 10^{-4}, 0.001, 0.252, 0.262, 0.004, 10^{-4}\}$,
		$\{10^{-4}, 10^{-4}, 0.284, 0.862, 0.166, 10^{-4}, 10^{-4}, 10^{-5}, 0.350, 0.161\}$,
		$\{0.238, 10^{-5}, 10^{-5}, 0.537, 10^{-5}, 10^{-4}, 0.758, 0.242, 0.148, 10^{-5}\}$,
		$\{10^{-4}, 0.086, 10^{-5}, 0.953, 0.139, 10^{-4}, 0.002, 10^{-5}, 10^{-4}, 0.257\}$,
		$\{0.221, 10^{-4}, 0.221, 0.762, 10^{-5}, 10^{-4}, 0.533, 0.192, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.536, 0.342, 0.713, 10^{-4}, 10^{-4}, 0.294, 0.039, 10^{-4}, 10^{-4}\}$,
		$\{0.346, 10^{-4}, 10^{-4}, 0.627, 10^{-4}, 10^{-4}, 0.392, 0.268, 0.511, 10^{-4}\}$,
		$\{0.485, 0.003, 0.755, 10^{-4}, 0.441, 0.001, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.118, 0.286, 0.403, 10^{-4}, 10^{-4}, 0.793, 0.336, 10^{-4}, 10^{-4}\}$,
		$\{0.571, 10^{-5}, 10^{-4}, 0.197, 0.712, 10^{-5}, 0.178, 0.310, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 0.253, 0.378, 0.341, 10^{-5}, 0.755, 0.326, 10^{-5}, 10^{-5}\}$,
		$\{0.439, 10^{-4}, 10^{-5}, 0.630, 0.131, 10^{-4}, 0.553, 0.296, 10^{-4}, 10^{-4}\}$,
		$\{10^{-6}, 0.174, 10^{-6}, 0.206, 0.219, 10^{-6}, 0.756, 10^{-6}, 0.555, 10^{-6}\}$,
		$\{0.003, 10^{-4}, 0.452, 0.334, 10^{-4}, 10^{-4}, 0.795, 0.228, 10^{-4}, 10^{-4}\}$,
		$\{0.624, 10^{-5}, 10^{-5}, 0.421, 0.435, 10^{-5}, 0.422, 0.256, 10^{-5}, 10^{-5}\}$,
		$\{0.469, 10^{-5}, 0.116, 0.596, 10^{-6}, 0.634, 10^{-5}, 0.089, 10^{-5}, 10^{-5}\}$,
		$\{0.208, 10^{-6}, 0.299, 0.681, 10^{-6}, 10^{-6}, 0.609, 0.181, 10^{-6}, 10^{-6}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-4}, 0.290, 0.037, 10^{-4}, 0.702, 0.203, 10^{-4}, 0.618\}$,
		$\{10^{-5}, 10^{-5}, 0.136, 0.252, 0.203, 10^{-5}, 0.910, 0.220, 10^{-5}, 10^{-6}\}$

Table B.12: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_6	$\frac{31}{144}$	$\{10^{-4}, 10^{-5}, 10^{-4}, 0.741, 10^{-5}, 10^{-5}, 0.627, 0.020, 10^{-5}, 0.240\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 0.369, 10^{-5}, 0.764, 0.398, 0.348, 10^{-4}, 10^{-5}\}$,
		$\{10^{-5}, 10^{-4}, 0.497, 0.581, 10^{-5}, 10^{-5}, 0.586, 0.267, 10^{-5}, 10^{-5}\}$,
		$\{10^{-7}, 10^{-7}, 0.382, 0.542, 10^{-7}, 0.727, 10^{-7}, 0.178, 10^{-7}, 10^{-7}\}$,
		$\{10^{-4}, 0.362, 0.393, 0.713, 10^{-4}, 10^{-4}, 0.454, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{0.229, 10^{-6}, 0.622, 10^{-6}, 10^{-6}, 10^{-6}, 0.472, 10^{-6}, 0.581, 10^{-6}\}$,
		$\{10^{-5}, 0.526, 10^{-5}, 0.728, 10^{-5}, 0.441, 10^{-5}, 10^{-5}, 10^{-4}, 0.002\}$,
		$\{10^{-5}, 0.486, 0.275, 0.772, 10^{-5}, 10^{-4}, 0.303, 10^{-5}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-5}, 10^{-4}, 0.507, 0.525, 10^{-5}, 10^{-5}, 0.643, 0.234, 10^{-5}, 10^{-4}\}$,
		$\{0.548, 10^{-4}, 0.717, 0.417, 10^{-5}, 10^{-4}, 10^{-4}, 0.111, 10^{-4}, 10^{-4}\}$,
		$\{0.418, 10^{-4}, 10^{-4}, 10^{-4}, 0.710, 10^{-4}, 0.539, 0.175, 10^{-4}, 10^{-4}\}$,
		$\{0.377, 10^{-4}, 10^{-4}, 0.618, 10^{-4}, 10^{-4}, 0.686, 0.075, 10^{-4}, 10^{-4}\}$,
		$\{0.664, 10^{-5}, 10^{-5}, 10^{-5}, 0.153, 10^{-5}, 0.677, 0.278, 10^{-4}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 0.895, 0.406, 10^{-4}, 10^{-4}, 0.001, 0.186, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.001, 0.765, 0.598, 10^{-4}, 10^{-4}, 10^{-4}, 0.238, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.001, 10^{-4}, 0.998, 10^{-4}, 10^{-4}, 0.001, 10^{-4}, 10^{-4}, 0.062\}$,
		$\{0.315, 10^{-4}, 10^{-4}, 10^{-4}, 0.687, 10^{-4}, 10^{-4}, 0.224, 0.615, 10^{-4}\}$,
		$\{10^{-4}, 0.687, 0.377, 0.620, 10^{-4}, 0.046, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{0.085, 10^{-5}, 0.431, 0.852, 10^{-5}, 10^{-4}, 0.284, 10^{-5}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-5}, 10^{-5}, 0.474, 0.044, 10^{-6}, 10^{-6}, 0.817, 0.325, 10^{-6}, 10^{-6}\}$,
		$\{10^{-5}, 10^{-4}, 10^{-5}, 0.096, 10^{-5}, 10^{-4}, 0.738, 10^{-5}, 0.603, 0.288\}$,
		$\{10^{-5}, 0.332, 0.162, 0.822, 10^{-5}, 10^{-5}, 0.434, 10^{-5}, 10^{-5}, 10^{-4}\}$,
		$\{10^{-4}, 10^{-4}, 0.378, 0.632, 10^{-4}, 10^{-4}, 0.676, 0.023, 10^{-4}, 10^{-4}\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 0.471, 0.445, 10^{-4}, 0.742, 10^{-5}, 0.171, 10^{-5}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 0.047, 0.558, 10^{-6}, 0.802, 0.208, 10^{-6}, 10^{-6}\}$,
		$\{10^{-6}, 0.427, 0.410, 0.715, 10^{-6}, 10^{-5}, 10^{-5}, 10^{-6}, 0.372, 10^{-6}\}$,
		$\{10^{-4}, 0.544, 10^{-4}, 0.535, 0.321, 10^{-4}, 0.561, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 0.448, 10^{-5}, 0.229, 0.838, 0.209, 10^{-5}, 10^{-6}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 0.426, 0.324, 10^{-6}, 0.827, 0.172, 10^{-6}, 10^{-6}\}$,
		$\{0.487, 10^{-5}, 10^{-4}, 10^{-5}, 0.305, 10^{-5}, 0.775, 0.261, 10^{-5}, 10^{-5}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 0.425, 0.203, 10^{-6}, 0.855, 0.217, 10^{-6}, 10^{-6}\}$

Table B.13: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_7	$\frac{26}{144}$	$\{10^{-4}, 10^{-5}, 10^{-5}, 0.930, 0.001, 10^{-5}, 0.367, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 0.640, 10^{-5}, 10^{-5}, 0.708, 0.298, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-5}, 0.559, 0.805, 10^{-4}, 10^{-4}, 0.196, 10^{-4}, 10^{-4}\}$,
		$\{10^{-5}, 10^{-5}, 0.598, 0.800, 10^{-5}, 10^{-5}, 10^{-5}, 0.057, 10^{-5}, 10^{-5}\}$,
		$\{0.515, 10^{-5}, 0.356, 0.779, 10^{-4}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-6}, 10^{-5}, 0.401, 10^{-5}, 10^{-5}, 10^{-5}, 0.858, 0.319, 10^{-5}, 10^{-5}\}$,
		$\{0.497, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-4}, 10^{-5}, 0.225, 0.838, 10^{-5}\}$,
		$\{10^{-5}, 10^{-5}, 0.826, 10^{-5}, 10^{-5}, 10^{-5}, 0.457, 0.330, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 0.695, 0.473, 10^{-4}, 10^{-4}, 0.542, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{0.067, 10^{-6}, 10^{-6}, 0.923, 10^{-6}, 10^{-6}, 0.379, 10^{-6}, 10^{-6}, 10^{-6}\}$,
		$\{0.582, 0.438, 0.685, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-4}\}$,
		$\{10^{-4}, 0.431, 10^{-4}, 0.854, 10^{-4}, 10^{-4}, 0.292, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-5}, 0.890, 10^{-6}, 10^{-6}, 0.409, 0.202, 10^{-6}, 10^{-6}\}$,
		$\{10^{-5}, 10^{-5}, 10^{-4}, 0.890, 0.198, 10^{-5}, 0.411, 10^{-6}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-5}, 10^{-5}, 0.939, 10^{-5}, 10^{-5}, 10^{-5}, 0.005, 0.343, 10^{-5}, 10^{-5}\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 0.631, 10^{-5}, 10^{-5}, 0.750, 0.197, 10^{-5}, 10^{-5}\}$,
		$\{0.915, 10^{-4}, 10^{-4}, 0.006, 10^{-4}, 10^{-4}, 10^{-4}, 0.403, 10^{-4}, 10^{-4}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-7}, 0.681, 10^{-6}, 0.600, 0.420, 10^{-6}\}$,
		$\{0.760, 10^{-5}, 10^{-5}, 0.620, 0.195, 10^{-5}, 10^{-4}, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 0.727, 10^{-6}, 10^{-5}, 0.652, 0.217, 10^{-6}, 10^{-6}$,
		$\{10^{-4}, 10^{-5}, 10^{-5}, 0.863, 10^{-5}, 10^{-5}, 0.168, 10^{-5}, 10^{-5}, 0.476\}$,
		$\{10^{-4}, 10^{-4}, 0.803, 0.398, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 0.444\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 0.692, 0.527, 10^{-5}, 10^{-5}, 10^{-6}, 10^{-5}, 0.493\}$,
		$\{10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 0.015, 10^{-6}, 0.366, 0.931, 10^{-6}\}$,
		$\{10^{-5}, 10^{-5}, 10^{-5}, 10^{-6}, 0.272, 10^{-5}, 0.914, 0.302, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 0.580, 0.435, 10^{-4}, 10^{-4}, 0.689, 10^{-4}, 10^{-4}, 10^{-4}\}$

TABLE B.14: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_8	$\frac{17}{144}$	$\{10^{-5}, 10^{-5}, 10^{-5}, 0.963, 10^{-4}, 10^{-5}, 10^{-5}, 0.271, 10^{-5}, 10^{-5}\}$,
		$\{0.596, 10^{-5}, 10^{-5}, 10^{-4}, 0.803, 10^{-5}, 10^{-4}, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{0.559, 10^{-5}, 10^{-4}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-4}, 10^{-5}, 0.829, 10^{-5}\}$,
		$\{10^{-5}, 10^{-4}, 10^{-5}, 0.916, 10^{-5}, 10^{-5}, 10^{-5}, 0.402, 10^{-5}, 10^{-4}\}$,
		$\{10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 0.925, 0.381, 10^{-7}, 10^{-7}\}$,
		$\{10^{-6}, 10^{-5}, 10^{-6}, 0.960, 10^{-6}, 10^{-6}, 10^{-5}, 0.278, 10^{-6}, 10^{-5}\}$,
		$\{0.866, 10^{-5}, 10^{-5}, 0.500, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-5}, 10^{-5}, 0.428, 10^{-5}, 10^{-4}, 10^{-5}, 10^{-5}, 0.904, 10^{-5}, 10^{-5}\}$,
		$\{10^{-7}, 10^{-7}, 10^{-7}, 0.956, 10^{-7}, 10^{-7}, 10^{-7}, 0.294, 10^{-7}, 10^{-7}\}$,
		$\{0.972, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 0.236, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 0.854, 0.520, 10^{-4}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-4}, 0.976, 10^{-4}, 10^{-4}, 10^{-4}, 0.219, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 0.569, 10^{-5}, 0.822, 10^{-5}, 10^{-4}, 10^{-4}, 10^{-5}, 10^{-5}, 10^{-4}\}$,
		$\{10^{-5}, 10^{-4}, 10^{-5}, 0.815, 10^{-5}, 10^{-5}, 0.580, 10^{-5}, 10^{-5}, 10^{-5}\}$,
		$\{10^{-4}, 10^{-4}, 10^{-4}, 10^{-5}, 0.599, 10^{-4}, 0.801, 10^{-4}, 10^{-4}, 10^{-4}\}$,
		$\{10^{-4}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-4}, 10^{-5}, 0.919, 0.394, 10^{-5}, 10^{-5}\}$,
		$\{10^{-7}, 10^{-7}, 10^{-7}, 0.319, 10^{-7}, 10^{-6}, 10^{-5}, 0.948, 10^{-7}, 10^{-7}\}$

TABLE B.15: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_9	$\frac{2}{144}$	$\{10^{-4}, 10^{-4}, 10^{-4}, 1.00, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}\}$
		$\{10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 1.00, 10^{-5}, 10^{-4}\}$

TABLE B.16: Spherical data calculated from monthly portfolios based on 10 indices in Table 4.2, categorised in different regions; S_i shows i elements are zero or close to zero (continued)

S_0	$\frac{13}{144}$	{0.158, 0.375, 0.332, 0.360, 0.233, 0.308, 0.349, 0.311, 0.309, 0.363},
		{0.004, 0.002, 0.001, 0.894, 0.001, 0.003, 0.346, 0.111, 0.089, 0.248},
		{0.008, 0.006, 0.753, 0.022, 0.002, 0.006, 0.006, 0.003, 0.007, 0.657},
		{0.032, 0.032, 0.032, 0.032, 0.032, 0.032, 0.032, 0.032, 0.032, 0.032},
		{0.002, 0.479, 0.002, 0.877, 0.001, 0.004, 0.002, 0.001, 0.013, 0.002},
		{0.011, 0.005, 0.004, 0.607, 0.009, 0.006, 0.008, 0.324, 0.007, 0.725},
		{0.007, 0.004, 0.004, 0.006, 0.006, 0.018, 0.679, 0.005, 0.733, 0.005},
		{0.003, 0.006, 0.001, 0.004, 0.003, 0.004, 0.003, 1.00, 0.004, 0.003},
		{0.778, 0.004, 0.011, 0.004, 0.013, 0.011, 0.009, 0.003, 0.628, 0.005},
		{0.004, 0.002, 0.287, 0.590, 0.512, 0.489, 0.236, 0.001, 0.116, 0.002},
		{0.008, 0.625, 0.021, 0.758, 0.002, 0.180, 0.029, 0.001, 0.025, 0.005},
		{0.130, 0.004, 0.002, 0.004, 0.979, 0.003, 0.009, 0.155, 0.003, 0.005},
		{0.002, 0.003, 0.002, 0.002, 0.001, 0.005, 0.877, 0.002, 0.347, 0.334}

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* Applied mathematics series. Dover Publications, 1964.

[2] J. Aitchison and S. D. Silvey. Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, 29(3):813–828, 09 1958.

[3] D. W. K. Andrews. Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383, 1999.

[4] D. Chant. On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika*, 61(2):291–298, 1974.

[5] H. Chernoff. On the distribution of the likelihood ratio. *Ann. Math. Statist.*, 25(3):573–578, 09 1954.

[6] W. G. Cochran. The χ^2 test of goodness of fit. *Ann. Math. Statist.*, 23(3):315–345, 09 1952.

[7] K. M. Creer, E. Irving, and A. E. M. Nairn. Palaeomagnetism of the great whin sill. *Geophysical Journal*, 2:306–323, December 1959.

[8] R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20:339–350, 2002.

[9] R. Engle and B. Kelly. Dynamic equicorrelation. *Journal of Business and Economic Statistics*, 30:2:212–228, 2012.

[10] P. I. Feder. On the distribution of the log likelihood ratio test statistic when the true parameter is "near" the boundaries of the hypothesis regions. *The Annals of Mathematical Statistics*, 39(6):2044–2055, 1968.

[11] C. J. Geyer. On the asymptotics of constrained M-estimation. *Ann. Statist.*, 22(4):1993–2010, 12 1994.

[12] M. Kendall and A. Stuart. *The Advanced Theory of Statistics (Volume 2)*. New York: Hafner, 1979.

[13] B. G. Lindsay. Composite likelihood methods. *Contemporary mathematics*, 80(1):221–39, 1988.

[14] K. V. Mardia, D. Holmes, and J. T. Kent. A goodness-of-fit test for the von mises-fisher distribution. *Journal of the Royal Statistical Society, Series B: Methodological*, 46:72–78, 1984.

[15] P. A. P. Moran. Maximum-likelihood estimation in non-standard conditions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 70:441–450, 11 1971.

[16] B. E. Neuenschwander and B. D. Flury. A note on Silvey's (1959) theorem. *Statistics and Probability Letters*, 36(3):307–317, 1997.

[17] C. Radhakrishna Rao. *Linear Statistical Inference and Its Applications*. John Wiley and Sons, Inc., 1973.

[18] R. T. Rockafellar and R. Wets. *Variational Analysis*, volume 317. Springer-Verlag, Berlin, 1998.

[19] A. Rotnitzky, D. R. Cox, M. Bottai, and J. Robins. Likelihood-based inference with singular information matrix. *Bernoulli*, 6(2):243–284, 2000.

[20] J. G. Saw. A family of distributions on the m-sphere and some hypothesis tests. *Biometrika*, 65(1):69–73, 1978.

[21] S. G. Self and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

[22] M. J. Silvapulle and P. K. Sen. *Constrained Statistical Inference: Order, Inequality, and Shape Constraints*. Wiley Series in Probability and Statistics. Wiley, 2011.

[23] R. A. Silverman. *Calculus With Analytic Geometry*. Prentice Hall College Div, 1985.

[24] S. D. Silvey. The Lagrangian Multiplier test. *Ann. Math. Statist.*, 30(2):389–407, 06 1959.

[25] G. Ulrich. Computer generation of distributions on the m-sphere. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(2):158–163, 1984.

[26] A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.

[27] C. Varin, N. M. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, Vol.21(No.1):5–42, 2011.

[28] H. T. V. Vu and S. Zhou. Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist.*, 25(2):897–916, 04 1997.

[29] G. S. Watson. Goodness-of-fit tests on a circle. *Biometrika*, 48(1/2):109–114, 1961.

[30] G. S. Watson. Large sample theory of the Langevin distribution. *Journal of Statistical Planning and Inference*, 8(3):245 – 256, 1983.

[31] G. S. Watson. *Statistics on spheres*. University of Arkansas lecture notes in the mathematical sciences. Wiley, 1983.

[32] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62, 03 1938.

[33] A. T. A. Wood. Simulation of the von Mises Fisher distribution. *Communications in Statistics - Simulation and Computation*, 23(1):157–164, 1994.